Omega 105 (2021) 102505

Contents lists available at ScienceDirect

Omega

journal homepage: www.elsevier.com/locate/omega

# Quality assessment of the Portuguese public hospitals: A multiple criteria approach[☆]

António Rocha[a], Ana Sara Costa[b,*], José Rui Figueira[b], Diogo Cunha Ferreira[c], Rui Cunha Marques[c]

[a] Instituto Superior Técnico, Universidade de Lisboa, Portugal
[b] CEG-IST, Instituto Superior Técnico, Universidade de Lisboa, Portugal
[c] CERIS, Instituto Superior Técnico, Universidade de Lisboa, Portugal

ARTICLE INFO

*Article history:*
Received 18 November 2020
Accepted 20 June 2021
Available online 22 June 2021

*Keywords:*
Quality
Hospitals
National health service
Electre tri-nc
Multiple criteria
Decision analysis
Multidimensional scales

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) and all of them should deliver quality care services. Quality in healthcare is assessed according to several criteria, such as patient safety, care appropriateness or access. However, over the last years political and economic events have had an impact on the SNS. Hence, structural reforms have occurred, and new healthcare policies have been implemented, mostly focused on improving efficiency and reducing costs. It associated to divestment can increment barriers to access, compromise infrastructures and equipment, and, above all, the service's quality. This work aims to assess quality of the Portuguese public hospitals (secondary care providers) in this line. To this aim, we adopt a multiple criteria decision aiding approach, applying the Electre Tri-nC method to build a decision model with intervention of an expert, who acts as the decision maker. Hospitals are assessed and assigned to predefined categories, taking into account the hospitals' performances on various criteria. Each criterion is characterized by different subcriteria, resulting in a complex criteria tree. Thus, to construct a multidimensional scale for each criterion, we propose an innovative approach using an Electre Tri-based method. The results are analyzed and the robustness of the model is tested. This work's findings may have potential application to healthcare policy and hospital funding in the SNS, in which financial sustainability is a permanent challenge.



## 1. Introduction

The health sector assumes a highly influential position in the society's bigger picture, since it is related to both social and economic aspects and has substantial impacts on modern civilizations' lives [1]. This particular combination of elements translated into a susceptible sector, mainly because attaining the highest health standard is one of every world citizen's fundamental rights [2]. However, ensuring access to quality healthcare services has had its toll over the last decades, with a significant health expenditure increase [3]. Indeed, demographic shifts, increasing life expectancy, complex and chronic diseases, expanding coverage of public health services, and technological improvements are some of the reasons for this escalation [4]. Furthermore, the inefficient management of resources by governments and healthcare institutions, and even their workforce are also determinants to be reckoned [5].

In 2015, as part of its Sustainable Development Goals' agenda, the United Nations declared its ambition to achieve universal health coverage by 2030: all people and communities should receive the quality services they need, being protected from health threats, regardless of their capacity to pay [6]. Nevertheless, it is vital to guarantee universal access to healthcare and to ensure that it follows safe and appropriate guidelines to provide quality health care [7], particularly nowadays with the COVID-19 (Sars-Cov-2) pandemic threatening our way of life and emphasizing pre-existing systemic issues [8]. One may think that the lack of quality health services only occurs in emerging countries. Nonetheless, the Organization for Economic Co-operation and Development (OECD) con-

☆ Area: Decision Analysis and Preference-Driven Analytics. This manuscript was processed by Associate Editor Nuria Agell.

* Corresponding author.
*E-mail addresses:* antoniomccrocha@gmail.com (A. Rocha), anasaracosta@tecnico.ulisboa.pt (A.S. Costa), figueira@tecnico.ulisboa.pt (J.R. Figueira), diogo.cunha.ferreira@tecnico.ulisboa.pt (D.C. Ferreira), rui.marques@tecnico.ulisboa.pt (R.C. Marques).

https://doi.org/10.1016/j.omega.2021.102505
0305-0483/

cluded that 15% of all hospital costs in OECD nations resulted from of patient harm from adverse events [9].

The scarcity of resources to allocate to healthcare is a reality, although the needs are virtually unlimited. Thus, increasing health expenditure raises issues related to the system's financial sustainability and equity in accessing healthcare [10]. It can be seen as an optimization problem where it is necessary to reduce costs without compromising quality, *i.e.*, to minimize costs while guaranteeing the health system's financial sustainability and providing quality healthcare to its users [11]. As one of the main healthcare providers, hospitals find it essential that their quality of healthcare is assessed to propose the adoption of the possible best practices. Accordingly, in this paper, we conduct a study concerning the quality assessment of the Portuguese public hospitals – the prime responsible for more than 50% of the Portuguese public health expenditure [12].

Moreover, the Portuguese Constitution, which grounds and governs the Portuguese State's principles and organization, declares: any Portuguese citizen has the right to health protection and the duty to defend and promote it. The right to access health is ensured through a National Health Service (SNS, from the Portuguese abbreviation of *Servio Nacional de Sade*). SNS aims to promote appropriate and equitable care, tendentiously free to its citizens. SNS is one of the oldest national services globally, but it is not meeting the needs of its users [13].

In 2018, 9.4% of Portugal's gross domestic product (GDP) was devoted to health,[1] more than the OECD average (8.8%).[2] Despite the attempts to maintain the investment in the systems, the SNS is struggling to accomplish its goals in a sector full of new and complex challenges, such as aging, population growth, and chronic illnesses and comorbidities [14]. Policymakers and healthcare managers have been applying measures to improve efficiency, mostly via cost containment, which must never compromise the quality of the provided services, and, if possible, maximize them [15]. However, the Portuguese public hospitals are funded based on contracts [16] that do not bear in mind the quality of those services [17]. For this reason, it is essential to assess their quality, making room for uncovering the effects of scarifying it and possibly setting benchmarks later to be inserted on funding.

In the present case study, hospitals' quality assessment is accomplished by using a multiple criteria approach due to the multidimensional nature of quality in health. Hence, to assess quality over its full extent, this study is supported by a Multiple Criteria Decision Aiding (MCDA) method – Electre Tri-nC [18]. This method allows the assignment of each hospital under analysis to a category while considering a set of criteria to assess its performance. It is also necessary to construct at least one reference hospital to define each category, i.e., a hospital representing that category. Preference parameters like weights have to be attributed to the criteria for constructing a decision model according to the decision-maker (DM) preferences. The preference information is gathered through a decision aiding process involving the interaction between the analyst and the DM.

Quality in health is, in general, modeled through the use of various indicators that can be used to form a family of criteria. It is necessary to study indicators, criteria, and all the variables related to health in the interest of conducting an appropriate evaluation. We use the hospital benchmarking database of the Portuguese Central Health System Administration (ACSS), which provides the key performance indicators regarding most hospitals of the SNS. Thus, in this case, as it happens in several multiple criteria decision situations, a criterion can involve a multidimensional description denoted by a logical group of indicators or subcriteria. In fact, at the beginning of this work, a thorny decision tree has been built and two possible approaches were considered: constructing criteria or applying a hierarchical method.

A hierarchical-based method, as in [19,20], could also be an alternative to our problem. However, it is less adequate for our case study since an active intervention of the DM is not wholly possible and her/his expressed willingness to participate in the process of selecting the criteria and constructing their qualitative scales from the subcriteria. The DM also specified that a qualitative scale for each criterion should be used, to consider both qualitative and quantitative aspects conjointly. Besides, imperfect data characteristics and arbitrariness may affect the construction of the criteria' scales, and compensatory effects are not suitable for this analysis. Since in a hierarchical method non-compensatory effects and the imperfect knowledge of the data cannot be taken into account, we have decided to discard the hypothesis of using a method of this kind.

The following approaches have been developed to construct a multidimensional criterion scale that accounts the (qualitative or quantitative) scales of the subcriteria:

1. The *constructed attributes procedure* by [21] for designing *constructed attributes* from a set of *natural attributes* (those having a common interpretation to everyone). The constructed attribute levels are obtained through a value model by considering combinations of levels of the natural attributes.
2. The *systematic procedure* for building multidimensional scales based on the concept of factorial design developed by [22] to build a scale. The possible combinations of all levels of the various dimensions are identified. Implausible combinations must be eliminated and plausible combinations representing a similar impact-level have to be considered at the same scale level. The number of combinations can be reduced by applying a *two-dimensions-at-a-time* fractional design technique [23] successively.
3. The *determinants procedure* proposed by [24], which is based on the sequential application of a set of key rules. These rules can be summed up as follows: *(i)* Define, for each subcriterion, two reference levels, called satisfactory and neutral; *(ii)* Assign, to each characteristic, a label (determinant, important, or secondary); and *(iii)* Define, for each subcriterion, the good level, when all determinant characteristics are satisfactory and the majority of important characteristics are satisfactory; and, the neutral when the majority of determinant and essential characteristics are neutral, without any negative.

The procedures described above can be used for both, qualitative and quantitative attributes. In general, these procedures consider many combinations. The analysis of all those combinations may require a tremendous cognitive effort from the DM or may not even be manageable when a combinatorial explosion occurs. A new more adequate procedure should be developed for our circumstances.

Due to our case's specific characteristics, we propose a new technique/procedure for building a multidimensional scale. It constitutes an innovative aspect of our work and it is mainly due to the following reasons: *(i)* The participation of the DM was a crucial aspect in the construction of the criteria scales; in fact, the DM expressed the intention of actively collaborating during the process of building the criteria scales, providing additional information, and sharing the knowledge and preferences about the dimensions used for assessing the quality of the Portuguese hospitals; *(ii)* During the interaction with the DM, we considered that a qualita-

[1] https://www.pordata.pt/en/Europe/Health+expenditure+as+a+percentage+of+GDP-1962, accessed November 02, 2020

[2] https://data.oecd.org/healthres/health-spending.htm, accessed November 02, 2020

tive scale for each criterion was the most adequate to model the different (quantitative and qualitative) aspects of the problem; and *(iii)* The qualitative scales should take into account the imperfect knowledge of the data; for such a reason, we needed to create a pessimistic and an optimistic scale for each criterion.

In line with the previous innovative aspects, we propose to make use of the Electre Tri-C outranking method to build the scales of each criterion. It is used in a co-constructive interaction process between the analyst and the DM, and can take into account both the quantitative and the qualitative nature of data, the imperfect character and arbitrariness of the data. In addition, it is a powerful tool for avoiding the systematic compensatory effects. We developed this innovative approach to aggregate the subcriteria scales to construct criteria scales by using the Electre Tri-C method [25], which differs from Electre Tri-nC, since it uses only one reference per category.

Therefore, the objectives of this work are as follows: *(i)* Create a criteria tree (with criteria and respective subcriteria) based on key performance indicators; *(ii)* Construct the criteria scales to assess the performances of the hospitals based on the assessment of the hospitals on the subcriteria, using Electre Tri-C as a novel approach; and *(iii)* Conceive a decision model to assess the overall quality of the Portuguese public hospitals, using Electre Tri-nC.

This paper is organized as follows. Section 2 introduces the concept of quality in healthcare services and the approaches for assessing that. Section 3 describes the case study addressed here, which is related to the Portuguese public hospitals' quality assessment. Section 4 is devoted to the presentation and discussion of the main results of the study. Section 5 performs a robustness analysis and discusses its results. Section 6 presents managerial insights. Section 7 provides concluding remarks and suggests future research lines.

## 2. Quality in healthcare services

Hospitals must provide 'services to citizens seeking healthcare services' to improve their quality of life. Healthy citizens are likely more productive, meaning that public investments in health should boost national economies. According to the American Institute of Medicine, those services delivered by hospitals must be [26,27]: *(i)* Safe for both the patient and the practitioner; *(ii)* Centered on the patient, *i.e.*, based on evidence, guideline, and scientific knowledge, and, therefore, appropriate; *(iii)* Effective in the sense that health status' improvements are fully attained; *(iv)* Equitable, *i.e.*, similar health conditions/situations demand similar treatments and resources allocation; *(v)* Accessible and timely; and *(vi)* Efficient, *i.e.*, the most significant health improvements are achieved at the lowest cost.

These six domains are generally used to define the quality of healthcare. Furthermore, in ethical terms, any medical and nursing care act should focus on four main principles [28]: *(i)* Beneficence, *i.e.*, do good; *(ii)* Non-maleficence, *i.e.*, do no harm; *(iii)* Justice, with equitable and fair resources allocation; and *(iv)* Autonomy, *i.e.*, the patient has free-will.

It is interesting to note that these ethical principles still figure in the *Corpus Hippocraticum* that embodies the physician's ideal with the medical profession's ethical and moral values. Based on those four principles and the six domains featuring healthcare quality, we can formulate a framework and construct a set of criteria to assess the hospitals' quality performance.

Safe care is free from errors and preventable adverse effects. Since Dame Florence Nightingale has first introduced the concepts of *preventable harm* and *first, not harm* [29], patient safety has been in the core of quality. Medical and nursing act frequently entails some risks to the patient [30], and such a risk arises with the illness's complexity and severity. Adverse effects like bloodstream infections and postoperative pulmonary embolisms can be either preventable or not [31]. Unless unpreventable and unpredictable, adverse effects result from care errors and, for that reason, can be understood as lack of care safety. These situations have a meaningful impact on society: patients' physical/psychological discomfort, dissatisfaction, and loss of trust in the healthcare system, healthcare staff frustration, loss of productivity and households' income, and wasted financial resources due to the additional care needed by error. Therefore, patients' safety is unquestionably a criterion to figure in any hospital performance assessment [32]. Because of the tremendous costs associated with medical errors ($735-980 billion in US, see [33]), safer care, with fewer patients harmed or injured, is less expensive care, thus better, more efficient, and less wasteful [34].

Besides safety, medical acts must be appropriate, *i.e.*, centered on the patient, non-discriminatory, and based on evidence guidelines and existing scientific knowledge [35]. Although some procedures may not harm the patient, if they are not recognized by the scientific knowledge ruling the traditional medicine, they are not considered appropriate care [36]. Hence, this kind of act should be avoided, and appropriateness should be a performance-based criterion [37]. At this point, attention should be paid to cesarean sections. There is an increasing worldwide need to reduce this kind of deliveries [38]. refer that sections' cesarean rates should be mitigated, especially in the categories of prior cesarean delivery and dystocia (obstructed labor). Meanwhile, [39] mentions that cesarean sections are associated with maternal and neonatal in-hospital complications. For instance, an increased maternal infection rate, a more extended period of healing, and potential complications in subsequent pregnancies, as well as transient tachypnea of the newborn and persistent pulmonary hypertension on infants. According to [40], clinical guidelines are useful to reduce those rates and, for that reason, to deliver appropriate care to the mother and the fetus. In the same vein, [41] conclude that the reduction in (nulliparous term singleton vertex) cesarean deliveries results in quality improvements.

The equity of access to healthcare is another quality dimension. There is great inconsistency in definitions of equity [42]. However, we may define it as the system's ability to deliver healthcare services to any citizen at their/his will and if required [43,44]. The quality of the provided service should not depend on the patients' characteristics. Missing resources and some barriers (either physical, organizational, or psychographic) make services less accessible to the users. An accessible health care service exhibits acceptable levels of resources per (potential) user or per demanded care act; is located near users; charges fair fees per medical/nursing act corresponding to the users' willingness to pay; handles its patients promptly; and presents short waiting lists and short waiting times [45,46].

Efficient services should obtain the best health outcomes at the minimum possible cost [47]. However, in several cases, health care providers are technically efficient because they divest on safety, appropriateness, and access to reduce operating costs and investments. It poses problems to the health care system: while it should remain financially sustainable, it also must deliver the best possible care. Therefore, trade-offs may exist between efficiency and other quality dimensions in healthcare [36], and, for that reason, the efficiency of resource utilization should be a particular criterion.

Avedis Donabedian, author of the remarkable paper *Evaluating the Quality of Medical Care* [48–50], defines effectiveness as *the degree to which attainable health improvements are realized* [47]. Meanwhile, the Agency for Healthcare Research and Quality defines it as the capacity of *providing services based on scientific knowledge to all who could benefit and refraining from providing services to those not likely to benefit (avoiding under-use and misuse,* *respectively)*". It means that effectiveness results from safe, appropriate, timely, accessible, and efficient care.

## 3. Case study: Constructing the decision model

In this section, we start by providing an overview of the case study. Then, we present the methods and describe the data sets used in this study. In the next subsections, we detail the decision model's construction, namely the criteria, categories, and parameters.

### *3.1. Description of the study*

The Portuguese SNS was implemented in 1979 after the transition between political regimes. The system is mainly based on the Beveridge model. The Portuguese Government is responsible for ruling and financing primary and secondary public healthcare providers. These public entities are sustained predominantly through public taxes distributed from an annual and global budget. The SNS is a statutory instrument for ensuring the right to health protection under the Constitution of the Portuguese Republic. The primary goal of SNS is to provide appropriate and equitable care to its beneficiaries (all Portuguese citizens), being tendentiously free.

Over the last decades, there has been an increase of pressure in the system through different factors: demographic shifts, new drugs and technologies due to the emergence of increasingly complex diseases and chronic diseases, an increase of life expectancy, and advances in technology, being thus of limited access. Due to these events, the health expenditure in Portugal is rising. For instance, the health expenditure *per capita* is approximately € 1500, one of the highest in the European Union. It compromises the financial sustainability of the SNS [15]. Thereby, it is essential for the performance evaluation of all entities belonging to the SNS. This case study centers on hospitals as they consume more than half of the national budget allocated to health.

Hospital performance criteria must reflect the quality of the services provided. Thus, the problem of evaluating their performance levels arrives when neither the concept of efficiency is clear, nor the criteria used for that reflect correctly the quality of the provided services [51]. Therefore, it is requested a detailed study about the quality of the SNS entities, which is possible through building an MCDA model. Note that this model must present a non-compensatory character to represent quality better. Acceptable performance levels in a criterion should not compensate for weak performances in other criteria, especially if we look to life-threatening events caused by poor care. The lack of patient safety resulting in death cannot be compensated by other criteria [52]. Likewise, an efficient hospital should not be considered a benchmark if the medical practice within is not well-conducted or centered on the patient or makes available all necessary resources for the patient's healing.

In the present study, we propose an MCDA approach to assess the hospitals quality performance through the construction of an Electre Tri-nC model. The DM asked us to create a scale with qualitative levels for each criterion. Such a scale was constructed from the set of subcriteria of each criterion through a sorting method. The objective was to provide an assessment of each hospital's quality regarding the different criteria, separately. After this, the DM required an aggregation of those performance levels into an overall assessment in terms of a composite qualitative indicator. The DM considered that the available methods were not adequate for such a purpose. In opposition, the Electre Tri-nC method allows fulfilling that goal, by providing a classification for each hospital in terms of quality level. It is assumed that such a model is co-constructed through the interaction between the analyst(s) and the focus group or, in this case, the DM. This cooperation eases gathering preference information, as well as understanding the reasoning and subjectivity associated to the opinions of the stakeholders, who participate in the focus group.

The DM that accompanied and cooperated in this study is a former expert from the Ministry of Health, who possesses know-how in the healthcare sector and performance assessment, adding to his knowledge in health policy-making and administration. This study was conducted during several interactive sessions with the analyst and the DM, who expressed his perceptions and preferences regarding the decision situation here addressed.

As previously mentioned, the possibility of an interplay between the analyst and the DM was a prerequisite for the DM. Indeed, the DM participated and accompanied the whole decision aiding process for building the model. In general, the process took place in an easy way, since the DM has experience with MCDA methods and a deep knowledge of the healthcare system.

### *3.2. Methods*

In this study, the DM created, with our help, a criteria tree based on five criteria (and some subcriteria for each criterion) characterizing both the activity and the quality of each hospital. We propose a new approach for constructing the criteria scales based on those indicators, in order to assess the hospitals performance levels on each criterion. We need to consider such criteria levels and build a decision model aiming to assess the overall quality of the Portuguese public hospitals.

We used two sorting methods to deal with the assessment of the Portuguese Public Hospitals in terms of overall quality:

1. Electre Tri-C [25], which constructs the scales of the criteria used to assess the performances of those hospitals; and
2. Electre Tri-nC [18] for the ultimate goal of assigning each hospital to an ordered category of quality, while a set of criteria and preference parameters are considered. These categories are defined through characteristic reference actions.

Electre Tri-nC is a generalization of Electre Tri-C, since the former imposes no constraints for the number of reference actions per category [18]. Electre Tri-nC allows the DM and the analyst to co-construct the decision process and characterize the categories with paramount freedom comparatively to Electre Tri-C. A brief description is provided in Appendix A.

Both methods use the following set of elements, data and parameters, for constructing a model with the DM:

1. Actions: They are the objects of the decision under evaluation that will be assigned to the categories.
2. Criteria: They are constructed based on different points of view for assessing the actions' performances. A qualitative or quantitative scale composed of various performance levels have to be established. Also a preference direction, *i.e.*, minimize (the lower the level, the better) or maximize (the greater the level, the better), has to be chosen.
3. Performance table: It contains all actions' performances on the criteria.
4. Categories: They are predefined and preferentially ordered to receive the actions according to the model.
5. Reference actions: A set of reference actions (or a single one, in the case of Electre Tri-C) are defined to represent each category.
6. Weights: A weight is associated with each criterion to represent its relative importance among the criteria.
7. Veto threshold: It is associated with a criterion aiming to increase the power of such a criterion; it reinforces the non-compensatory character of the methods by the application of a veto power.

8. Preference and indifference thresholds: These two parameters are used for modelling the imperfect character of the data from the computation of the performances attributed to the actions.
9. Credibility level: This parameter represents the minimal degree of credibility, bearing in mind all criteria, to validate the outranking relation between an action and a reference action or a set of reference actions (it can be compared to the majority level in voting theory).

In both methods, it is assumed that the DM participates in the integral decision aiding process, through guidance of the analyst, providing her/his judgements and relevant information for defining the above-mentioned elements of the decision model.

### 3.3. Data processing

This case study focuses on assessing the quality of the Portuguese public hospitals and, for that, it is important that the data were reliable and significant. The Central Administration of the Health System in Portugal created and maintains a benchmarking database containing several operating data (key performance indicators) regarding most hospitals of the SNS. The main goal of this database creation was to increase the transparency of the operations within the SNS. The benchmarking database is publicly available online.[3] The database contains information for many key performance indicators, measured monthly during several years.

The first task of data processing was selecting the time interval. Even though this study's information in the benchmarking database already included data until September 2019, it was not available for many entities. Therefore, since the data from 2017 and 2018 were complete, we decided to use the data corresponding to these years. This time lag does not contain significant technology shifts that would influence our results. Although the information is provided by month, it is simple to produce the accumulated results per year and institution, allowing to compare the performance between the two whole years.

The second task of the data processing consisted of selecting the hospitals out of the forty-three institutions that initially composed the benchmarking database. The exclusion criteria were as follows:

1. Substantial missing data (*e.g.*, public-private partnerships, smaller hospitals that do not provide one specific service);
2. Specialized hospitals, with specific technologies of production (*e.g.*, oncology centers, maternity), thus not comparable with the others;
3. Vertically integrated hospitals, providing more than one level of health care (primary and secondary care).

This data selection resulted in twenty-five institutions, containing five hospitals and twenty hospital centers. In this case study, the hospitals to be included are the potential actions $a_m,\ m = 1, \ldots, 25,$ detailed in Table 1. Finally, after collecting all data, the result was twenty-five hospitals operating during 2017 and 2018. All in all, this represents a 600-entry sample per year, therefore a total of 1200-entry samples.

### 3.4. Construction of the criteria

This subsection presents the family of criteria and their corresponding family of subcriteria. They were built based on the literature review and discussed with the DM. We selected twenty-four indicators from the thirty-four indicators distributed in six dimensions proposed by the benchmarking database: Access, Performance Assistance, Productivity, Economic-Financial, Safety, and Volume and Usage (we discarded these six dimensions since each indicator will be allocated to a subcriterion for being operationalized). During the construction of the criteria family with the DM, the criteria were easily validated. However, some subcriteria changed from a family of criteria, some were merged, and others were excluded. When it comes to attributing a subcriterion to a criterion, some are more direct; others request a more in depth analysis and interaction with the DM. The resulting family of criteria considered for this assessment consists of five criteria, denoted $g_n$, for $n = 1, \ldots, 5$. The set of criteria and subcriteria resulting from the decisions taken are described below. The indicator that operationalizes each subcriterion is also provided.

**Table 1**
Hospitals considered in the case study.

| Action | Code | Name |
|---|---|---|
| $a_1$ | CHMA | *Centro Hospitalar do Médio Ave* |
| $a_2$ | CHPV | *Centro Hospitalar Póvoa de Varzim/Vila do Conde* |
| $a_3$ | CHBM | *Centro Hospitalar Barreiro/Montijo* |
| $a_4$ | CHL | *Centro Hospitalar de Leiria* |
| $a_5$ | CHS | *Centro Hospitalar de Setúbal* |
| $a_6$ | CHBV | *Centro Hospitalar do Baixo Vouga* |
| $a_7$ | CHDV | *Centro Hospitalar Entre Douro e Vouga* |
| $a_8$ | CHMT | *Centro Hospitalar Médio Tejo* |
| $a_9$ | CHTS | *Centro Hospitalar Tâmega e Sousa* |
| $a_{10}$ | CHUCB | *Centro Hospitalar Universite1~rio Cova da Beira* |
| $a_{11}$ | HSO | *Hospital da Senhora da Oliveira, Guimãres* |
| $a_{12}$ | HDS | *Hospital Distrital de Santarém* |
| $a_{13}$ | CHTV | *Centro Hospitalar Tondela-Viseu* |
| $a_{14}$ | CHTAD | *Centro Hospitalar Te1~rs-os-Montes e Alto Douro* |
| $a_{15}$ | CHUA | *Centro Hospitalar Universite1~rio do Algarve* |
| $a_{16}$ | CHVNG | *Centro Hospitalar Vila Nova de Gaia/Espinho* |
| $a_{17}$ | HESE | *Hospital Espírito Santo de Évora* |
| $a_{18}$ | HFF | *Hospital Fernando da Fonseca* |
| $a_{19}$ | HGO | *Hospital Garcia de Orta* |
| $a_{20}$ | CHLO | *Centro Hospitalar de Lisboa Ocidental* |
| $a_{21}$ | CHUCB | *Centro Hospitalar e Universite1~rio de Coimbra* |
| $a_{22}$ | CHULC | *Centro Hospitalar Universite1~rio de Lisboa Central* |
| $a_{23}$ | CHUSJ | *Centro Hospitalar Universite1~rio de São João* |
| $a_{24}$ | CHUP | *Centro Hospitalar Universite1~rio do Porto* |
| $a_{25}$ | CHULN | *Centro Hospitalar Universite1~rio Lisboa Norte* |

1. *Access* ($g_1$). This criterion models the system's ability to provide care services to any citizen if she/he demands or requires. It is expected that accessible health care services display adequate levels of resources per user or per demanded care act to preserve or improve the citizens' health status. It is a criterion to be maximized and should be modeled through a qualitative scale built from the following set of subcriteria:
   (a) *First medical appointments timeliness* ($g_{1,1}$). An accessible healthcare service handles its patients in a reasonably timely manner whenever required. In Portugal, there is a legislated maximum guaranteed time for the first medical appointments in hospitals that makes possible considerations about accessibility. This subcriterion should be:
      - Maximized; and
      - Modeled through the indicator *rate of non-urgent first medical appointments performed in an adequate time* - Number of first medical appointments executed in tolerable time per 100 first medical appointments.

   (b) *Enrolled patient for surgery* ($g_{1,2}$). The reason for including this subcriterion in the access's family is in line with the previous. It is a matter of time and the healthcare service's ability to deal with its surgical waiting list

[3] http://benchmarking.acss.min-saude.pt

without compromising the maximum legislated time for surgery. This subcriterion should be:
- Maximized; and
- Modeled through the indicator *rate of enrolled patients in the surgical waiting list within the guaranteed response time* - Number of patients enrolled in the surgical waiting list within the guarantee response time per 100 patients enrolled in the surgical waiting list.

(c) *Occupancy of beds* ($g_{1,3}$). This subcriterion concerns the healthcare entities' occupancy rate entirely related to its accessibility and equity of services. The number of hospital beds provides a measure of the resources available to deliver services to inpatients. Beds must be maintained, staffed, and immediately available to use [53]. This subcriterion should be:
- Minimized (as it regards the difference between the real and the ideal occupancy rate); and
- Modeled through the indicator *occupancy rate* - Rate that associates the number of days in acute beds and the number of available acute inpatient beds over time, *i.e.*, 100×hospital days/365×beds. According to the DM, the ideal value for the occupancy rate is 85%. The absolute of the difference to the excellent value was used to measure hospital performance.

(d) *Availability of doctors* ($g_{1,4}$). Access to medical care demands the existence of strong characters: the doctors. More doctors available imply less waiting time for a medical appointment, a surgery, or an emergency. Thus, this subcriterion represents well the access to the proper healthcare services, allowing timely care, and reducing patients' dissatisfaction. This subcriterion should be:
- Maximized; and
- Modeled through the indicator *doctors per 1000 inhabitants* - Number of doctors in a hospital per 1000 inhabitants in that hospital's influence area.

(e) *Availability of nurses* ($g_{1,5}$). Regarding care delivery and quality, no less critical is the nurses' presence, also fundamental to provide accessible care [54]. This subcriterion should be maximized and modeled through the indicator. This subcriterion should be:
- Maximized; and
- Modeled through the indicator *nurses per 1000 inhabitants* - Number of nurses in a hospital per 1000 inhabitants in that hospital's influence area.

2. *Care Appropriateness* ($g_2$). This criterion models the ability to deliver patient-centered care services supported by evidence-based guidelines or scientific knowledge [55]. It is expected that when following evidence-base guidelines, the intervention or service results in health benefits (*e.g.*, increased life expectancy, improved functional capacity, and pain relief) exceeding the expected health risks (*e.g.*, mortality, morbidity, and pain resulting from the intervention) by a wide fair enough margin to take the intervention or service worth doing. Whether the healthcare services are not appropriate we anticipate the resolution of the patient's main problem, resulting in excessive staying delay, which can also result in other diseases appearances (*e.g.*, pressure ulcers and hospital-acquired infections, unstable therapy at discharge, unsuitable postdischarge care and, of course, the last scenario, readmission of the patient) [55]. It is a criterion to be maximized and should be modeled through a qualitative scale built from the following set of subcriteria:

(a) *Minor surgeries appropriateness* ($g_{2,1}$). Some major surgeries can be executed as minor procedures supported by clinical evidence without harming to the patient as a subcriterion indicating for care appropriateness. This subcriterion should be:
- Maximized; and
- Modeled through the indicator *number of outpatient surgeries per 100 potential outpatient procedure* - Number of outpatient surgeries per 100 potential outpatient procedures.

(b) *Avoidable readmission prior 30 days after discharge* ($g_{2,2}$). It represents the result of inappropriate preparedness and postdischarge care. This subcriterion should be:
- Minimized; and
- Modeled through the indicator *number of readmissions in 30 days after discharge per 100 inpatients* - Number of patients readmitted within 30 days after discharge per 100 inpatient episodes.

(c) *Excessive staying delay* ($g_{2,3}$). It reflects the inadequacy of the provided care because more extended stays may result in-hospital infections (septicemia) and pressure ulcers. This subcriterion should be:
- Minimized; and
- Modeled through the indicator *number of long-stay inpatients per 100 admissions* - inpatients staying more than 30 days per 100 inpatient episodes.

(d) *Hip surgeries timeliness* ($g_{2,4}$). This subcriterion was first included in the Access family of criteria, but we included this subcriterion within the Appropriate Care criterion after a discussion with the DM. According to the DM's vision, even though hip fractures, especially in elderly patients, represent a significant cause of morbidity and mortality, there is still no consensus about the proceeding's optimal waiting time; however, it should round two days [53]. It is still difficult to identify whether the timeliness in this type of procedures conditions access or only patient's health status specially unnecessary pain and for that reason in this study it is considered a measure of care appropriateness. This subcriterion should be:
- Minimized; and
- Modeled through the indicator *number of hip surgeries performed in the first 48 hours per 100 hip surgeries* - Number of hip surgical procedures on elderly, in the first 48 hours per 100 elderly patients with hip surgeries.

(e) *Delay before surgery* ($g_{2,5}$). It concerns the time between the patient's admission and the surgery. Also, this subcriterion was first included in the *Access* criterion; yet, after a discussion with the DM, it included in the inclusion of this subcriterion in Care Appropriateness. The reason pointed out was the same as the previous subcriterion, bearing in mind whether timeliness conditions access patients' health. Thus, the DM, decided to consider it a measure of care appropriateness because delaying surgeries interferes with the medical guidelines. This subcriterion should be:
- Minimized; and
- Modeled through the indicator *average waiting time before surgery* - Number of days until a surgical episode occurs per scheduled surgical episode.

3. *Patient safety* ($g_3$). This criterion models the absence of preventable harm to a patient during health care and reduction of risk of unnecessary harm associated with health care to an acceptable minimum. Therefore, it is expected that healthcare processes do not harm patients, not exposing them to chemicals, foreign bodies, trauma, or infectious agents. It is a criterion to be maximized and should be modeled through a qualitative scale built from the following set of subcriteria:

(a) *Bedsores* ($g_{3,1}$). These episodes are preventable, resulting in less harm or no harm to the patients. Therefore, bedsores' presence indicates a lack of clinical safety, and should be minimized [56]. This subcriterion should be:
  - Minimized; and
  - Modeled through the indicator *number of bedsores per 100 inpatients*.

(b) *Bloodstream infections related to central venous catheter (CVC)* ($g_{3,2}$). Any in-hospital infection is an avoidable harmful event, being directly related to poor clinical safety. This subcriterion should be:
  - Minimized; and
  - Modeled through the indicator *bloodstream infection rate related to CVC per 100 inpatients*.

(c) *Postoperative pulmonary embolisms or thrombosis* ($g_{3,3}$). Any of these events are highly fatal complications for patients during or after surgery [57]. This subcriterion should be:
  - Minimized; and
  - Modeled through the indicator *postoperative pulmonary embolism/deep venous thrombosis cases per 100 surgical procedures*.

(d) *Postoperative septicaemia* ($g_{3,4}$). Septicemia or sepsis is a life-threatening condition triggered by an infection, resulting in tissues and organs' injuries [58]. For instance, sepsis events mainly occur after surgery, leading to organ failure and death. Thus, it results from poor safety within the hospital (namely, in the operating theater and the ward). This subcriterion should be:
  - Minimized; and
  - Modeled through the indicator *postoperative septicaemia cases per 100 inpatients*.

(e) *Non-instrumental vaginal deliveries with severe laceration* ($g_{3,5}$). During childbirth, patient safety can be assessed by bearing in mind the potentially avoidable tearing of the perineum during vaginal delivery [59]. Thus, vaginal lacerations are directly related to a lack of clinical safety. This subcriterion should be:
  - Minimized; and
  - Modeled through the indicator *cases of trauma on vaginal delivery (third and fourth degree lacerations), without instrumentation, per 100 assisted deliveries*.

(f) *Assisted vaginal deliveries with severe laceration* ($g_{3,6}$), in line with the previous subcriterion. Note that a perineal laceration risk is significantly increased when instruments are used to assist the delivery [59]. This subcriterion should be:
  - Minimized; and
  - Modeled through the indicator *cases of trauma on vaginal delivery (third and fourth degree lacerations), with instrumentation, per 100 assisted deliveries*.

4. *Efficiency* ($g_4$). This criterion models the system's ability to achieve the objectives concerning the resources consumed (*e.g.*, labor and capital) to produce valued outputs (*e.g.*, inpatients and outpatients). The health care systems providers have a genuine interest in seeking out best practices and identifying room for improvement. However, there are some cases where those health care providers are technically efficient not because the best practices are being followed, but because they divest on safety, care appropriateness, and access to increasing the number of treated patients mitigating the lack of investment. The health entities' primary goal should be a financially sustainable management of resources and deliver the best care possible. Efficiency is a criterion to be maximized and should be modeled through a qualitative scale built from the following set of subcriteria:

(a) *Expenses with staff* ($g_{4,1}$). Staff is an essential resource to deliver care. It includes doctors, nurses, ancillary staff, technicians, and administrative staff, to name a few. All have roles within the hospital value chain and take part in the patient healing process. For the sake of sustainability, any waste of resources, including staff, is to be minimized. This subcriterion should be:
  - Minimized; and
  - Modeled through the indicator *expenses with staff per (complexity and severity adjusted) patient*.

(b) *Expenses with drugs, pharmaceutical products, and clinical consumables* ($g_{4,2}$). It is worth to mention that the three types of expenses were being considered independently. However, with the DM's endorsement, it was considered that data were far more complete in case the three types of costs were evaluated together. Nonetheless, they belong to the efficiency domain, in which a careful managing of the resources must be performed, and the (waste of) expense is to be minimized. This subcriterion should be:
  - Minimized; and
  - Modeled through the indicator *expenses with drugs, pharmaceutical products, and clinical consumables per (complexity and severity-adjusted) patient*.

(c) *Expenses with supplies and external services* ($g_{4,3}$). This subcriterion is directly related to efficiency since it demands careful management of the resources that ideally should be the minimum value without compromising the patients. This subcriterion should be:
  - Minimized; and
  - Modeled through the indicator *Expenses with supplies and external services per severity-adjusted patient* (describes the value in euros of the expenses with supplies and external services per severity-adjusted patient).

(d) *Expenses with overtime* ($g_{4,4}$). This subcriterion is directly related to inefficiency, since the health professionals work extra hours to fulfill an individual hospital entity's needs. This subcriterion should be:
  - Minimized; and
  - Modeled through the indicator *Expenses with overtime per total expenses with staff* (represents the percentage of the value in euros of the expenses with overtime within the value of expenses with staff).

(e) *Expenses with outsourcing* ($g_{4,5}$). The need to require outsourcing entities to fulfill specific hospital entity's needs is linked to inefficiency. Beware that according to the DM, the cost per hour of outsourcing services in health is much upper than with overtime and, of course, new hires. This subcriterion should be:
  - Minimized; and
  - Modeled through the indicator *Expenses with outsourcing per total expenses with staff* (shows the percentage of the value in euros of outsourcing expenses within the value of expenses with staff).

5. *Caesarean Appropriateness* ($g_5$). This criterion models the ability to deliver patient-centered care services in caesarean sections supported by evidence-based guidelines. Although this criterion is related to *Care Appropriateness*, $g_2$, the fact that it has its technology of production made us opt for separating them. When medically justified, a caesarean section can effectively prevent and decrease maternal and perinatal mortality and morbidity. Neverthless, there is no evidence describing caesarean delivery benefits for women or infants who do not require the procedure. Caesarean sections can cause significant and sometimes permanent complications,

**Table 2**
Performances of the reference hospitals per category for *Care Appropriateness* ($g_2$).

| Category | Performance | Reference hospital | Subcriterion | | | | |
|---|---|---|---|---|---|---|---|
| | | | $g_{2.1}$ | $g_{2.2}$ | $g_{2.3}$ | $g_{2.4}$ | $g_{2.5}$ |
| $C_5$ | Very Good | $b_5^1$ | 90.00 | 5.00 | 2.70 | 90.00 | 0.50 |
| $C_4$ | Good | $b_4^1$ | 85.00 | 6.50 | 3.20 | 80.00 | 0.60 |
| $C_3$ | Neutral | $b_3^1$ | 80.00 | 7.40 | 3.70 | 50.00 | 0.90 |
| $C_2$ | Poor | $b_2^1$ | 75.00 | 8.30 | 4.50 | 30.00 | 1.10 |
| $C_1$ | Very Poor | $b_1^1$ | 70.00 | 9.80 | 5.20 | 20.00 | 1.40 |

disability or even death, especially in situations related to lack of the facilities or capacity to properly conduct safe surgery and treat surgical complications [60]. Nonetheless, the rates of caesarean delivery have increased over time in nearly all OECD countries and globally, WHO reported that annually there is an overuse of ineffective care of more than six million caesarean sections [9,53]. It led us to consider *Caesarean Appropriateness* as a criterion, since its practice commits quality in cases where it is not medically justified. It is a criterion to be maximized and will be modeled through a qualitative scale built from the following set of subcriteria:

(a) *Volume of caesarean sections* ($g_{5,1}$). According to the literature, the performance of caesarean sections is related to inadequate care. This subcriterion should be:
  - Minimized; and
  - Modeled through the indicator *Number of cesarean sections per 100 deliveries* (is the percentage of cesarean deliveries in the universe of deliveries).

(b) *Caesarean sections in UCFTPs* ($g_{5,2}$). According to the literature, the performance of caesarean sections in UCFTPs is related to inadequate care. This subcriterion should be:
  - Minimized; and
  - Modeled through the indicator *Number of cesarean sections in UCFTPs per 100 sections in UCFTPs* (represents in percentage the number of cesarean sections in UCFTPs in the universe of deliveries).

(c) *First caesarean sections in UCFTPs* ($g_{5,3}$). According to the literature, the performance of the first caesarean sections in UCFTPs is related to inadequate care. This subcriterion should be:
  - Minimized; and
  - Modeled through the indicator *Number of first cesarean sections in UCFTPs per 100 deliveries in UCFTPs without cesarean section before* (rate of first cesarean sections in UCFTPs with no cesarean section before within the universe of deliveries).

The description above is summarized in Table B.1 (Appendix B).

### 3.5. Criteria scales

In this section the characteristics for each of the five criteria and the twenty-four subcriteria are presented. When a criterion has more than one subcriterion, it is known as a built-in criterion. Consequently, there is a need for building scales to cover the multidimensionality of the criterion. Concerning our subcriteria, with the DM's help, it was easy to create scales based on indicators' scales. However, in what concerns the criteria, each of them has more than two subcriteria (multidimensional), which is a significant problem to tackle in this case study.

Although some approaches have been suggested through the years to build scales combining multiple dimensions, none of them was applicable to our case. It arrives from the fact that each criterion aggregates several subcriteria and each of them has several levels. There are multiple possible combinations to establish the levels of the subcriteria. We propose an innovative approach that uses the results of the application of the Electre Tri-C method to the subcriteria set (the DM was only able to define one reference action per category) to define the levels for the criteria set. The following steps have been followed for each criterion:

1. Define the levels for all the subcriteria;
2. Apply the Electre Tri-C method to assess the hospitals according to the set of subcriteria;
3. Convert the categories assessed in the previous step to each action to a level between 1 and 5 on an ordinal scale, where level 1 is the minimum and level 5 is the best. As we defined five categories, this conversion is direct ($C_1$ represents level 1, $C_2$ represents level 2, until $C_5$ that represents a level 5), unless the Electre method had assessed an interval of categories. In the case of that, we decided to create a *lower-level view* and an *upper-level view*. The lower-level view corresponds to the worst category of an interval of categories (minimum) attributed to action. The upper-level view corresponds to the best category of an interval of categories (or maximum) attributed to an action. Both these views can include the cases where only one category was assigned to an action, presenting both the same category.

Let us illustrate this procedure for the case of criterion *Care Appropriateness*, $g_2$. To obtain all twenty-five hospitals' performances on the five criteria, consider the original performances of the respective subcriteria (corresponding to the levels assigned to the indicators), we built five models with Electre Tri-C. The categories correspond to the final criteria scale levels, as described in Step 3 above. Regarding criterion $g_2$, in interaction with the DM, a set of reference hospitals (one per category) were defined by providing the performances on the respective five subcriteria, displayed in Table 2.

A set of weights were assigned to the subcriteria using the procedure SRF described in Subsection 3.7.2 (for the case of the criteria). Thus, the values considered in the case under analysis are presented in Table 3.

Indifference and preference thresholds were considered for all the subcriteria (see Table B.12). The veto effect can be used for mitigating compensation (i.e., avoiding systematic compensatory effects) among some subcriteria [61], but in this application the DM did not consider relevant the use of veto thresholds for the subcriteria. In this case, the focus was on the construction of the scales of the five main criteria from the available data regarding the indicators/subcriteria. Defining the reference profiles and weights was relatively easy to the decision maker, due to his knowledge about the healthcare system and those indicators. However, at the criteria level (upper aggregation), the DM expressed that it was crucial that the model presents a non compensatory character to better represent quality, as good performances in a criterion do not compensate weak performances in other criteria, especially looking to life threatening ones, where for instance a poor patient safety resulting in death cannot be compensated by other criteria. For the credibility level, the DM established $\lambda = 0.60$. Taken into account

**Table 3**
Weight attributed by the DM to each subcriterion of *Care Appropriateness* ($g_2$).

| Criterion | Subcriterion | Non-normalized weight | Normalized weight |
|---|---|---|---|
| $g_2$, Care Appropriateness | $g_{2,1}$ | 7 | 26.92 |
| | $g_{2,2}$ | 10 | 38.46 |
| | $g_{2,3}$ | 7 | 26.92 |
| | $g_{2,4}$ | 1 | 3.85 |
| | $g_{2,5}$ | 1 | 3.85 |

**Table 4**
Assessment of the hospitals considering the subcriteria of $g_2$ (Electre Tri-C).

| Hospital | 2017 | | 2018 | |
|---|---|---|---|---|
| | Minimum | Maximum | Minimum | Maximum |
| $a_1$ | $C_3$ | $C_4$ | $C_3$ | $C_4$ |
| $a_2$ | $C_3$ | $C_4$ | $C_3$ | $C_4$ |
| $a_3$ | $C_2$ | $C_2$ | $C_1$ | $C_2$ |
| $a_4$ | $C_3$ | $C_3$ | $C_4$ | $C_4$ |
| $a_5$ | $C_3$ | $C_4$ | $C_4$ | $C_4$ |
| $a_6$ | $C_3$ | $C_4$ | $C_3$ | $C_4$ |
| $a_7$ | $C_3$ | $C_4$ | $C_4$ | $C_4$ |
| $a_8$ | $C_3$ | $C_3$ | $C_2$ | $C_2$ |
| $a_9$ | $C_4$ | $C_4$ | $C_3$ | $C_4$ |
| $a_{10}$ | $C_1$ | $C_2$ | $C_2$ | $C_3$ |
| $a_{11}$ | $C_3$ | $C_3$ | $C_2$ | $C_3$ |
| $a_{12}$ | $C_2$ | $C_3$ | $C_3$ | $C_3$ |
| $a_{13}$ | $C_2$ | $C_3$ | $C_4$ | $C_5$ |
| $a_{14}$ | $C_4$ | $C_4$ | $C_4$ | $C_4$ |
| $a_{15}$ | $C_2$ | $C_3$ | $C_2$ | $C_3$ |
| $a_{16}$ | $C_2$ | $C_3$ | $C_3$ | $C_4$ |
| $a_{17}$ | $C_3$ | $C_4$ | $C_3$ | $C_3$ |
| $a_{18}$ | $C_1$ | $C_2$ | $C_4$ | $C_4$ |
| $a_{19}$ | $C_3$ | $C_4$ | $C_2$ | $C_4$ |
| $a_{20}$ | $C_2$ | $C_3$ | $C_3$ | $C_3$ |
| $a_{21}$ | $C_2$ | $C_2$ | $C_3$ | $C_3$ |
| $a_{22}$ | $C_2$ | $C_3$ | $C_3$ | $C_4$ |
| $a_{23}$ | $C_2$ | $C_3$ | $C_4$ | $C_4$ |
| $a_{24}$ | $C_3$ | $C_3$ | $C_5$ | $C_5$ |
| $a_{25}$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ |

**Table 5**
Example of conversion from categories to viewpoint for three hospitals of criterion $g_2$ in 2018.

| Hospital | Category | | Viewpoint | |
|---|---|---|---|---|
| | Minimum | Maximum | Lower-level | Upper-level |
| $a_2$ | $C_3$ | $C_4$ | 3 | 4 |
| $a_3$ | $C_1$ | $C_2$ | 1 | 2 |
| $a_4$ | $C_4$ | $C_4$ | 4 | 4 |

all data (25 hospitals and 5 categories) and parameters (references, weights, thresholds, and credibility level), the results of this Electre Tri-C model for $g_2$ were obtained using appropriate software (MCDA-ULaval v0.6.16). That is, a category or a range of possible categories was assigned to each hospital.

Both the lower-level view and upper-level view have been considered for all criteria in the next steps (model construction). We studied a scenario considering the performances based on a lower-level view and another with performances based on an upper-level view, using data from 2017 and 2018. Table 4 displays the hospitals' assessment considering the subcriteria of $g_2$ using the Electre Tri-C model described above. Table 5 shows the conversion of the assigned categories concerning criterion $g_2$, in 2018, for four hospitals, where occurred the case of being assigned two possible categories and, consequently, we attributed two viewpoints.

The procedure proposed above was performed for all criteria. The five criteria are expressed on an ordinal scale with five levels and they are to be maximized. The data related to the remaining subcriteria used for constructing the criteria scales is shown in Appendix B. Tables B.2 and B.3 contain the performances of the hospitals on all subcriteria in the years of 2017 and 2018, respectively. Tables B.4–B.7 show the performances of the reference actions. Tables B.8–B.11 present the subcriteria weights. Table B.12 provides the discriminating thresholds associated to the subcriteria scales. The data were obtained over several interactions with the DM, who expressed the preferences and knowledge based on his experience and expertise.

### 3.6. Performance tables

In this study, we use the performance tables presented in Table 6, corresponding to the hospitals' assessment of the five criteria for the four considered scenarios and the scales previously constructed (see Subsection 3.5).

### 3.7. Construction of the parameters

This subsection describes the decision model constructed in interaction with the DM using the Electre Tri-nC method. The parameters are related to the criteria previously defined, considering already the final scale obtained by applying Electre Tri-C, as explained in Subsection 3.5.

#### 3.7.1. Categories and their reference actions

In line with the DM, a set of five categories to receive the hospitals regarding quality was predefined as $C_5$ - Very Good, $C_4$ - Good, $C_3$ - Neutral, $C_2$ - Poor, and $C_1$ - Very Poor.

Later, based on the knowledge of the DM, we defined the reference (dummy) hospitals for each category, as displayed in Table 7.

#### 3.7.2. Criteria weights and veto thresholds

In order to determine the weights, we applied the Deck of Cards Method based on the revised Simos procedure, also known by the name of its original software SRF (from Simos-Roy-Figueira) [62]. It allows any DM to easily rank a set of criteria in a given context and provide the analyst the information needed to attribute a numerical value to each criterion's weight.

In general, the SRF procedure considers two phases: a meeting with the DM to gather all the information needed to apply the method and afterwards the calculation of each criterion's weight.

Regarding this case study, the procedure was applied as many times as the number of existing families of criteria and subcriteria. It means one time for the criteria and five times for the subcriteria.

In the first phase, we collected the necessary information for obtaining the weight values with SRF, according to the DM's expertise, following the steps below.

1. Firstly, we provided the DM a set of $n$ cards, according to the $n$ criteria, containing the criterions name previously defined and written on it;
2. Secondly, we asked the DM to rank the cards (or criteria) provided in ascending order. It resulted in a hierarchy, where the first criterion was the least important (lowest

**Table 6**
Performance tables for the years of 2017 and 2018 and respective viewpoints.

| Hospital | 2017 | | | | | | | | | | 2018 | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Lower-level | | | | | Upper-level | | | | | Lower-level | | | | | Upper-level | | | | |
| | $g_1$ | $g_2$ | $g_3$ | $g_4$ | $g_5$ | $g_1$ | $g_2$ | $g_3$ | $g_4$ | $g_5$ | $g_1$ | $g_2$ | $g_3$ | $g_4$ | $g_5$ | $g_1$ | $g_2$ | $g_3$ | $g_4$ | $g_5$ |
| $a_1$ | 2 | 3 | 4 | 2 | 2 | 3 | 4 | 4 | 2 | 3 | 3 | 3 | 4 | 3 | 2 | 3 | 4 | 4 | 4 | 3 |
| $a_2$ | 4 | 3 | 5 | 1 | 3 | 5 | 4 | 5 | 1 | 3 | 4 | 3 | 5 | 2 | 3 | 5 | 4 | 5 | 2 | 3 |
| $a_3$ | 2 | 2 | 3 | 3 | 3 | 2 | 2 | 3 | 3 | 4 | 3 | 1 | 4 | 3 | 3 | 3 | 2 | 4 | 3 | 3 |
| $a_4$ | 2 | 3 | 5 | 4 | 2 | 2 | 3 | 5 | 4 | 2 | 2 | 4 | 5 | 3 | 2 | 2 | 4 | 5 | 3 | 2 |
| $a_5$ | 3 | 3 | 4 | 2 | 4 | 3 | 4 | 4 | 2 | 4 | 3 | 4 | 3 | 3 | 3 | 3 | 4 | 4 | 3 | 3 |
| $a_6$ | 2 | 3 | 3 | 3 | 3 | 2 | 4 | 3 | 3 | 3 | 2 | 3 | 3 | 3 | 3 | 2 | 4 | 3 | 3 | 3 |
| $a_7$ | 2 | 3 | 4 | 3 | 4 | 2 | 4 | 4 | 3 | 4 | 2 | 4 | 3 | 3 | 3 | 2 | 4 | 3 | 3 | 4 |
| $a_8$ | 2 | 3 | 3 | 2 | 3 | 2 | 3 | 4 | 2 | 3 | 2 | 2 | 5 | 1 | 3 | 2 | 2 | 5 | 2 | 3 |
| $a_9$ | 2 | 4 | 3 | 4 | 4 | 2 | 4 | 3 | 4 | 4 | 2 | 3 | 4 | 3 | 4 | 2 | 4 | 4 | 3 | 5 |
| $a_{10}$ | 2 | 1 | 1 | 2 | 2 | 3 | 2 | 1 | 2 | 2 | 3 | 2 | 3 | 1 | 2 | 3 | 3 | 3 | 1 | 2 |
| $a_{11}$ | 1 | 3 | 4 | 4 | 3 | 1 | 3 | 4 | 4 | 4 | 1 | 2 | 4 | 3 | 3 | 1 | 3 | 4 | 3 | 3 |
| $a_{12}$ | 2 | 2 | 4 | 2 | 3 | 2 | 3 | 4 | 2 | 3 | 2 | 3 | 3 | 1 | 3 | 2 | 3 | 3 | 2 | 3 |
| $a_{13}$ | 2 | 2 | 4 | 4 | 3 | 2 | 3 | 4 | 4 | 4 | 3 | 4 | 2 | 4 | 4 | 3 | 5 | 2 | 4 | 5 |
| $a_{14}$ | 2 | 4 | 3 | 3 | 2 | 2 | 4 | 4 | 3 | 2 | 2 | 4 | 4 | 3 | 1 | 2 | 4 | 4 | 3 | 2 |
| $a_{15}$ | 3 | 2 | 2 | 2 | 3 | 3 | 3 | 2 | 2 | 4 | 3 | 2 | 3 | 2 | 3 | 3 | 3 | 3 | 2 | 4 |
| $a_{16}$ | 3 | 2 | 3 | 4 | 2 | 3 | 3 | 3 | 4 | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 4 | 3 | 3 | 3 |
| $a_{17}$ | 2 | 3 | 2 | 3 | 3 | 2 | 4 | 3 | 3 | 3 | 2 | 3 | 2 | 2 | 2 | 2 | 3 | 2 | 2 | 2 |
| $a_{18}$ | 1 | 1 | 1 | 3 | 2 | 1 | 2 | 1 | 3 | 3 | 2 | 4 | 2 | 3 | 3 | 2 | 4 | 3 | 3 | 4 |
| $a_{19}$ | 3 | 3 | 3 | 4 | 4 | 3 | 4 | 3 | 4 | 4 | 3 | 2 | 3 | 4 | 3 | 3 | 4 | 3 | 4 | 4 |
| $a_{20}$ | 3 | 2 | 2 | 4 | 3 | 3 | 3 | 2 | 4 | 4 | 2 | 3 | 2 | 4 | 3 | 2 | 3 | 2 | 4 | 3 |
| $a_{21}$ | 2 | 2 | 3 | 4 | 3 | 2 | 2 | 3 | 4 | 3 | 2 | 3 | 4 | 4 | 3 | 2 | 3 | 4 | 4 | 3 |
| $a_{22}$ | 3 | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 2 | 3 | 3 | 3 | 4 | 2 | 3 | 3 |
| $a_{23}$ | 3 | 2 | 2 | 3 | 3 | 3 | 3 | 2 | 5 | 3 | 2 | 4 | 3 | 4 | 3 | 3 | 4 | 3 | 4 | 4 |
| $a_{24}$ | 2 | 3 | 3 | 3 | 3 | 2 | 3 | 3 | 5 | 3 | 3 | 5 | 3 | 5 | 3 | 3 | 5 | 3 | 5 | 3 |
| $a_{25}$ | 2 | 2 | 3 | 4 | 3 | 3 | 2 | 3 | 4 | 3 | 2 | 2 | 2 | 3 | 3 | 3 | 2 | 2 | 4 | 4 |

**Table 7**
Criteria performances of the reference actions per category.

| Category | Performance | Reference hospital | Criterion | | | | |
|---|---|---|---|---|---|---|---|
| | | | $g_1$ | $g_2$ | $g_3$ | $g_4$ | $g_5$ |
| $C_5$ | Very Good | $b_5^1$ | 5 | 5 | 5 | 5 | 5 |
| | | $b_5^2$ | 5 | 4 | 5 | 4 | 5 |
| | | $b_5^3$ | 5 | 4 | 5 | 4 | 4 |
| $C_4$ | Good | $b_4^1$ | 4 | 4 | 5 | 4 | 5 |
| | | $b_4^2$ | 4 | 4 | 5 | 4 | 4 |
| | | $b_4^3$ | 4 | 4 | 4 | 4 | 4 |
| $C_3$ | Neutral | $b_3^1$ | 4 | 4 | 4 | 3 | 4 |
| $C_2$ | Poor | $b_2^1$ | 3 | 3 | 4 | 3 | 4 |
| | | $b_2^2$ | 3 | 3 | 3 | 3 | 3 |
| $C_1$ | Very Poor | $b_1^1$ | 3 | 2 | 3 | 2 | 3 |
| | | $b_1^2$ | 2 | 2 | 2 | 2 | 3 |
| | | $b_1^3$ | 2 | 2 | 2 | 1 | 3 |

**Table 8**
Ranking of the criteria.

| Card/#Blank cards | $g_3$ | 0 | $g_2$ | 1 | $g_1$ | 3 | $g_5$ | 3 | $g_4$ | Ratio-$z$ |
|---|---|---|---|---|---|---|---|---|---|---|
| Position | 5 | | 4 | | 3 | | 2 | | 1 | 10 |

**Table 9**
Weight attributed to each criterion.

| Criterion | Non-normalized weight | Normalized weight |
|---|---|---|
| $g_1$, Access | 7.55 | 23.60 |
| $g_2$, Care Appropriateness | 9.18 | 28.69 |
| $g_3$, Patient Safety | 10.00 | 31.25 |
| $g_4$, Efficiency | 1.00 | 3.12 |
| $g_5$, Caesarean Appropriateness | 4.27 | 13.34 |

weight), and the last was the most important (most significant weight). In the case of criteria having the same importance (same weight), they were grouped. According to the ranking's position, the cards were attributed to different ranks being the lowest level Rank 1, the second-lowest level Rank 2, etc.;

3. Afterward, we asked the DM if there were any consecutive cards or groups of cards where the importance was greater. According to the difference between the rankings, the DM added a blank card (or more) between the consecutive card (or group of cards). In case no blank card was added, it meant the difference of two consecutive levels was one unit, one blank card meant the difference of importance was two units, and so on;
4. In the last step, we asked the DM, how many times the criterion/criteria presented in the most important position, was/were more important than the criterion/criteria present in the least significant position. The result was described by a numerical value called the ratio-$z$ [62].

Table 8 illustrates the ranking constructed by the DM and the respective value of ratio-$z$. In case of existence, blank cards were numerically represented between two consecutive ranking levels, numerically ordered (1 stands for the least important position).

In the second phase of the procedure, we used the DecSpace platform[4] to execute the DCM-SRF method. The data previously collected was inserted, namely the criteria name, the ranking of the criteria cards, the blank cards, and the ratio-$z$.

As output, the values of the non-normalized weights and normalized weight were provided, as presented in Table 9.

The DM felt relevant to assign values to the veto threshold ($v_j$) to *Patient Safety* and *Care Appropriateness* criteria, mainly due to the importance of both assuming when evaluating healthcare quality, which reinforces the importance of both criteria in a quality assessment and the non-compensatory character of the method, as shown in Table 10.

[4] Available at http://decspace.sysresearch.org

**Table 10**
Veto thresholds attibuted to the criteria.

| | $g_1$ | $g_2$ | $g_3$ | $g_4$ | $g_5$ |
|---|---|---|---|---|---|
| $v$ | ∅ | 2 | 3 | ∅ | ∅ |

*3.7.3. Preference and indifference thresholds*

In the present case, we did not establish values for the preference and indifference thresholds. These technical parameters were defined in the construction of the criteria scales based on the subcriteria scales using our innovative approach. However, they were not considered in the case study itself. The reason is twofold: the criteria scales have a few levels, and the imperfect knowledge is already modeled by considering two possible viewpoints (the lower-level view and the upper-level view).

*3.7.4. Credibility level*

Regarding the credibility level, we started by explaining to the DM that this parameter should take a value greater than 0.5 and lower or equal to 1. We explained that it could be viewed as a majority measure for validating the outranking relation between a hospital and a reference hospital. Thus, initially, the DM expressed some hesitation for choosing a value within the range [0.55,0.65], and, in the end, the DM validated the credibility level with a value of 0.60. He considered it as an adequate value for this decision situation. As previously stated, $\lambda = 0.60$ was also used in the application of Electre Tri-C for constructing the criteria scales based on the subcriteria information (explained in Section 3.5).

## 4. Results

This section is devoted to the analysis and discussion of the results of our decision model. We describe the construction of an outranking relation and analyze its exploitation, and discuss the main achieved results.

*4.1. Analysis of the construction of an outranking relation*

The objective of the model's execution is to assign the actions to one of the five categories previously defined. This assignment proceeds to the establishment of the outranking relations. Electre Tri-nC builds one or more outranking relations that consider each action's performance on each criterion. These relations enable to state whether an action $a$, is preferred to an action $b$, denoted $aSb$, in case there are enough arguments to decide that $a$ is at least as good as $b$ according to a criterion $g_j$. The credibility index measures these relations' credibility, $\sigma$, and when considering a set of reference actions, it is called a categorical credibility index [18] (see A.4).

In our case, we used the software MCDA-ULaval v0.6.16 to obtain the assignment results. The software also provides those categorical credibility indices. In what follows, for the sake of simplicity, we only analyze the results for the lower-level view of 2018. Table 11 shows the corresponding categorical credibility indices.

While observing this table, we can quickly notice the presence of outranking relations justified by categorical credibility indices equal to 1, $\sigma(a, B_h) = 1$ or $\sigma(B_h, a) = 1$. Other important conclusions regarding the categories are described below:

1. Most potential actions denote categorical credibility equal or close to one for the case $\sigma(a, B_1)$. In other words, 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;
2. The opposite happened for $C_5$ as expected: the category demonstrated an outranking relation over all the potential actions, $\sigma(B_5, a) = 1, \forall a$. It sustains the fact that $C_5$ is the best category from the existing five.

Concerning the comparison between the credibility level $\lambda$ and the categorical credibility indices $\sigma$ that allows establishing one of the four $\lambda$-binary relations (see A.4): (a) $\lambda$-outranking, (b) $\lambda$-preference, (c) $\lambda$-indifference, and (d) $\lambda$-incomparability. The com-

**Table 11**
Categorical credibility indices between the hospitals and the subsets of reference actions for the lower-level view of 2018.

| Hospital | $\sigma(a, B_h)$ | | | | | $\sigma(B_h, a)$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $B_1$ | $B_2$ | $B_3$ | $B_4$ | $B_5$ | $B_1$ | $B_2$ | $B_3$ | $B_4$ | $B_5$ |
| $a_1$ | 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 | 0.71 | 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 |

**Table 12**
Assignment results for the years of 2017 and 2018 and respective viewpoints.

| Hospital | 2017 | | | | 2018 | | | |
|---|---|---|---|---|---|---|---|---|
| | Lower-level | | Upper-level | | Lower-level | | Upper-level | |
| | Min. | Max. | Min. | Max. | Min. | Max. | Min. | Max. |
| $a_1$ | $C_2$ | $C_2$ | $C_2$ | $C_3$ | $C_2$ | $C_2$ | $C_2$ | $C_4$ |
| $a_2$ | $C_3$ | $C_4$ | $C_4$ | $C_5$ | $C_3$ | $C_3$ | $C_4$ | $C_5$ |
| $a_3$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_2$ |
| $a_4$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_3$ | $C_3$ | $C_3$ | $C_3$ |
| $a_5$ | $C_2$ | $C_3$ | $C_2$ | $C_4$ | $C_2$ | $C_2$ | $C_2$ | $C_3$ |
| $a_6$ | $C_1$ | $C_2$ | $C_2$ | $C_2$ | $C_1$ | $C_2$ | $C_2$ | $C_2$ |
| $a_7$ | $C_2$ | $C_2$ | $C_2$ | $C_4$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ |
| $a_8$ | $C_1$ | $C_2$ | $C_2$ | $C_2$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ |
| $a_9$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_3$ | $C_4$ |
| $a_{10}$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_2$ |
| $a_{11}$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_1$ | $C_1$ | $C_2$ | $C_2$ |
| $a_{12}$ | $C_1$ | $C_1$ | $C_2$ | $C_2$ | $C_1$ | $C_2$ | $C_1$ | $C_2$ |
| $a_{13}$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_3$ | $C_3$ | $C_4$ |
| $a_{14}$ | $C_1$ | $C_1$ | $C_2$ | $C_3$ | $C_2$ | $C_3$ | $C_2$ | $C_3$ |
| $a_{15}$ | $C_1$ | $C_1$ | $C_2$ | $C_2$ | $C_1$ | $C_2$ | $C_2$ | $C_2$ |
| $a_{16}$ | $C_1$ | $C_2$ | $C_1$ | $C_2$ | $C_1$ | $C_2$ | $C_2$ | $C_2$ |
| $a_{17}$ | $C_1$ | $C_1$ | $C_2$ | $C_2$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ |
| $a_{18}$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ |
| $a_{19}$ | $C_2$ | $C_2$ | $C_2$ | $C_3$ | $C_1$ | $C_2$ | $C_2$ | $C_3$ |
| $a_{20}$ | $C_1$ | $C_1$ | $C_2$ | $C_2$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ |
| $a_{21}$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ |
| $a_{22}$ | $C_1$ | $C_2$ | $C_1$ | $C_2$ | $C_1$ | $C_2$ | $C_2$ | $C_2$ |
| $a_{23}$ | $C_1$ | $C_1$ | $C_1$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_3$ |
| $a_{24}$ | $C_1$ | $C_2$ | $C_1$ | $C_2$ | $C_3$ | $C_3$ | $C_3$ | $C_3$ |
| $a_{25}$ | $C_1$ | $C_1$ | $C_1$ | $C_2$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ |

parison between an action and the subsets of reference actions of each category, $B_h$, results in one and only one of the following cases [18]:

1. Action $a$ is neither $\lambda$-indifferent nor $\lambda$-incomparable to $B_h, h = 1, \ldots, q$;
2. Action $a$ is $\lambda$-indifferent to at least one subset of reference actions $B_h$. Furthermore, if $B_h$ is not unique, then the subsets of reference actions, which are $\lambda$-indifferent to an action $a$ set of categories consecutive;
3. Action $a$ is $\lambda$-incomparable to at least one subset of reference actions $B_h$. Furthermore, if $B_h$ is not unique, then the subsets of reference actions, which are $\lambda$-incomparable to action $a$, define a subset of consecutive categories.

In our case, the credibility level chosen was $\lambda = 0.6$ and the results of the $\lambda$-binary relations are represented in Table B.13: $I$ represents the $\lambda$-indifference relations between the action and the set of reference actions that define a category; $R$ represents the $\lambda$-incomparability relations; and the inequality symbols $\prec$ and $\succ$ denote preference relations, $\prec$ in case the set of reference actions is preferred to the action and $\succ$, otherwise.

### 4.2. Analysis of the exploitation of an outranking relation

In this step, we finally present the predefined categories to which the hospitals (actions) were assigned to, i.e., the assignment results. As stated in Appendix A, the assignment procedure is done by comparing an action $a$ with the reference actions, bearing in mind the credibility level $\lambda$, in our case $\lambda = 0.6$, and applying the two joint rules (the ascending and the descending rules).

The assignment procedure using both rules conjointly selects a lowest and a highest possible categories to an action $a$, generating a range of possible categories, $\Gamma(a)$ [18]:

1. When an action $a$ is neither $\lambda$-indifferent nor $\lambda$-incomparable to $B_h$, $h = 1, \ldots, q$, $\Gamma(a)$ is composed of one or two consecutive categories;
2. When $a$ is $\lambda$-indifferent to at least one subset of reference actions $B_h$, $\Gamma(a)$ is composed of the subset of consecutive categories defined by such $\lambda$-indifference, and, possibly, by including one or two of the adjacent categories to them;
3. When $a$ is $\lambda$-incomparable to at least one subset of reference actions $B_h$, $\Gamma(a)$ is composed of the subset of consecutive categories defined by such $\lambda$-incomparability, and, possibly, by including one or two of the adjacent categories to them.

Let us recall that we proposed an innovative approach that used the Electre Tri-C assignment procedure results from the subcriteria families to construct the criteria scales; in the assignment procedure, some of the actions were assigned to an interval of categories, which originated two viewpoints: the lower-level view, which considered the worst category assigned to an action, and the upper-level view, which considered the best category of the interval assigned to an action. Hence, the assignment procedure results (considering the five criteria) include both viewpoints for 2017 and 2018, which are all represented in Table 12.

### 4.3. Discussion of the main results

The number of hospitals (actions) assigned to each interval of categories, and the respective percentages for each of the years and viewpoints, are summed up in Table 13. Once observed this table, we notice variations on the hospitals' assignment to the categories regarding the view and the years.

The main variations on the assignment results are as follows:

1. In 2017, only one interval of categories, $[C_1, C_2]$, maintained the same number of hospitals assigned for both the views with five hospitals; and also three categories with no hospitals assigned, $C_3$, $C_4$, and $C_5$.
2. In the lower-level view, the most represented category was the worst category, $C_1$, with eleven hospitals and 92% of the hospitals were assigned to a category equal or lower than $C_2$. In respect to the highest category assigned, it was in the interval of $[C_3, C_4]$ with only one action, $a_2$.

**Table 13**
Number and respective percentage of hospitals assigned per interval of category.

| Interval of category | | 2017 | | 2018 | |
|---|---|---|---|---|---|
| Min. | Max. | Lower-level view | Upper-level view | Lower-level view | Upper-level view |
| $C_1$ | $C_1$ | 11 (44%) | 4 (16%) | 7 (28%) | 4 (16%) |
| $C_1$ | $C_2$ | 5 (20%) | 5 (20%) | 6 (24%) | 3 (12%) |
| $C_2$ | $C_2$ | 7 (28%) | 10 (40%) | 7 (28%) | 8 (32%) |
| $C_2$ | $C_3$ | 1 (4%) | 3 (12%) | 2 (8%) | 4 (16%) |
| $C_2$ | $C_4$ | 0 (0%) | 2 (8%) | 0 (0%) | 1 (4%) |
| $C_3$ | $C_3$ | 0 (0%) | 0 (0%) | 3 (12%) | 2 (8%) |
| $C_3$ | $C_4$ | 1 (4%) | 0 (0%) | 0 (0%) | 2 (8%) |
| $C_4$ | $C_4$ | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) |
| $C_4$ | $C_5$ | 0 (0%) | 1 (4%) | 0 (0%) | 1 (4%) |
| $C_5$ | $C_5$ | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) |

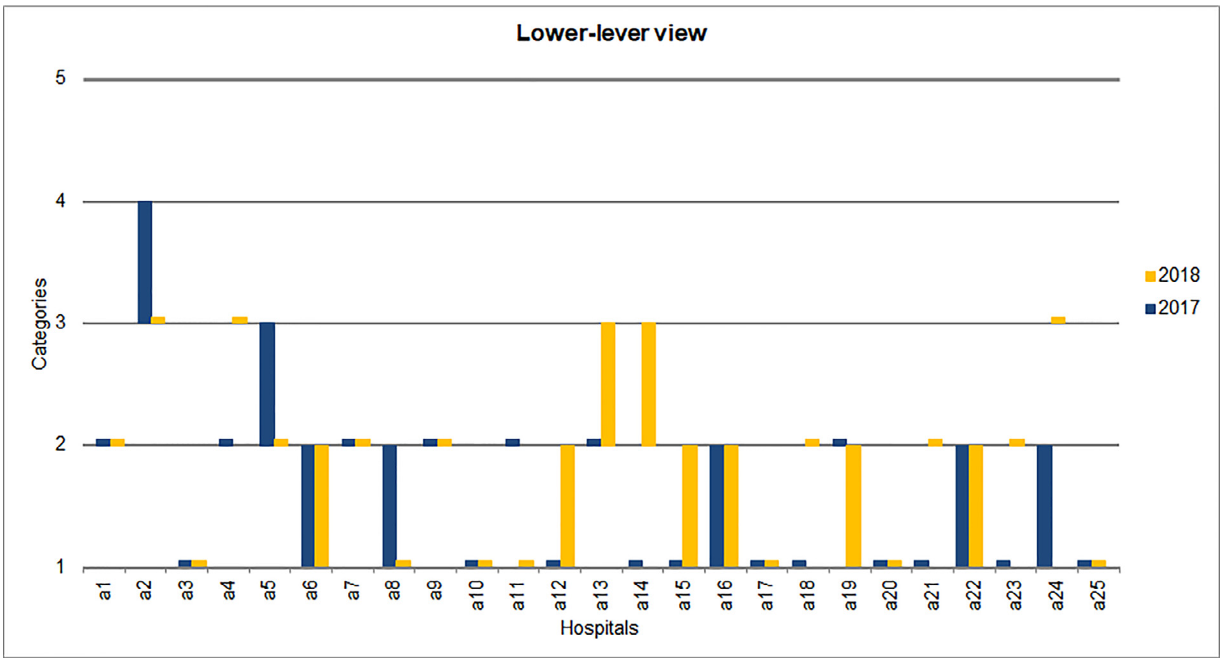


**Fig. 1.** Comparison of the assignments between the years of 2017 and 2018 for the lower-level view.

3. In the upper-level view, the most represented category was $C_2$ with ten hospitals, leaving behind $C_1$ with four hospitals representing a reduction of 60% compared to the lower-level view and 76% of the hospitals were assigned to a category equal or lower than $C_2$ representing a reduction of 17% compared to the lower-level view. Furthermore, in the upper-level view the highest category assigned to action was in the interval of $[C_4, C_5]$ with one action assigned, again $a_2$.
4. In 2018, only two categories, $C_4$ and $C_5$, maintained the same number of hospitals assigned for both the views with no hospitals.
5. In the lower-level view, the most represented category was the worst category, $C_1$, as well as category $C_2$, both with seven hospitals assigned and 80% of the hospitals were assigned to a category equal or lower than $C_2$, in respect to the highest category assigned it was $C_3$ with three hospitals assigned.
6. In the upper-level view, the most represented category was $C_2$, with eight hospitals, leaving behind $C_1$ with four hospitals assigned, representing a reduction of 43% comparing to the lower-level view. 60% of the hospitals were assigned to a category equal to or lower than $C_2$, representing a 25% reduction compared to the lower-level view. Furthermore, in the upper-level view, the highest category assigned to action was in the interval of $[C_4, C_5]$ with one action assigned, $a_1$.

The hospitals in the lower-level view were assigned to equal or worst categories compared to the upper-level views, corroborating what was expected in line with the approach created in Subsection 3.5 regarding the scales. In general, there was maintenance or an improvement of the quality in the hospitals between 2017 and 2018, expressed when comparing lower-level and upper-level views.

Concerning all the four assignment results, no hospitals were assigned to the best category $C_5$. It allows us to conclude that there was no under evaluation of the reference hospitals that define that category. In opposition, regarding the worst category, $C_1$, there were several hospitals assigned to it.

To provide better visualization of the assignment procedure, we included in Figs. 1 and 2 two plots representing the category or interval of categories attributed to each hospital for 2017 and 2018, one for each viewpoint. Once again, it is visible that the upper-level view achieved better categories than the lower-level view, for obvious reasons. The categories attributed to the hospitals in 2018 tend to be equal or better than in 2017, which is more visible in the upper-level viewpoint.In Appendix C we include Fig. C.3 for

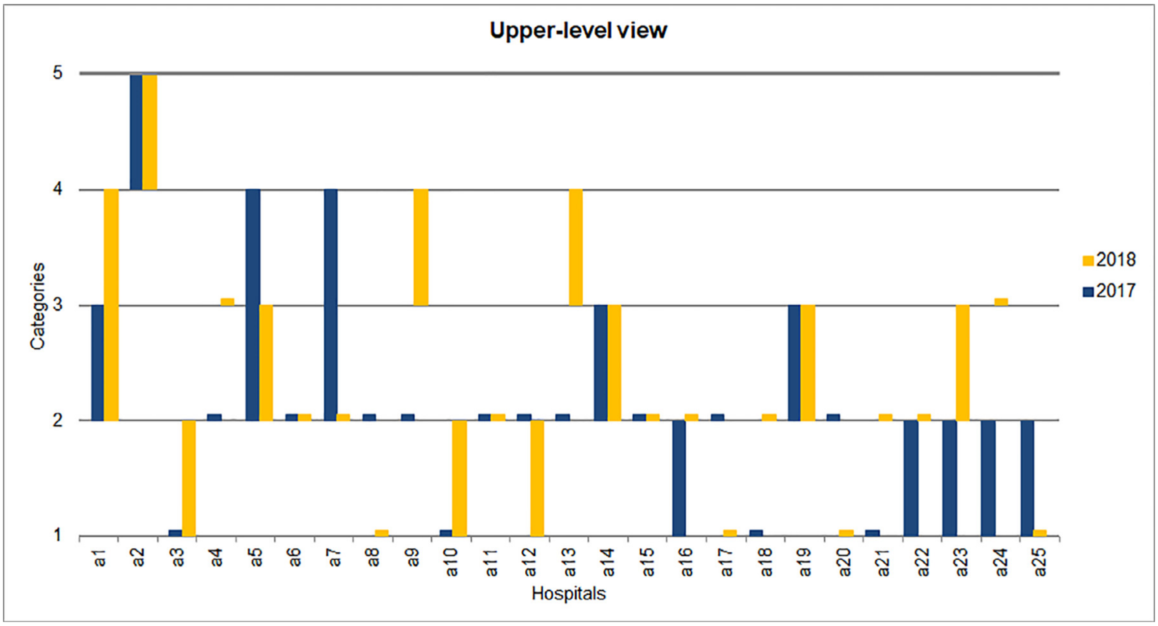


**Fig. 2.** Comparison of the assignments between the years of 2017 and 2018 for the upper-level view.

showing the evolution of each hospital from 2017 to 2018, that is, if a hospital improves, stays equal or decreases in quality.

Looking at the hospitals with better categories assigned, $a_2$ (Póvoa do Varzim/Vila do Conde Hospital Centre), it was considered: the best action for both years in the upper-level view; the best in the lower-level view in 2017; and, together with $a_{24}$, the best in the lower-level view for 2018. These results show that the action $a_2$ is consistently assigned to the 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 $a_{24}$ (Porto University Hospital Centre) together with $a_2$ it is assigned with the best category for the lower-level view of 2018 with a $C_3$, having the same evaluation for the upper-level view. It suggests that $a_{24}$ presents better quality healthcare among the largest hospital centers of the NHS. Hospitals $a_9$ and $a_{13}$ for 2018 in the upper-level view were assigned with an interval of categories $[C_3, C_4]$, still far from the results presented by $a_2$, among the best categories attributed in this case study.

Although, it is visible that there were hospitals 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 centers $a_{25}$ and $a_{20}$. Those represent the hospitals with the lowest quality healthcare performances according to the case study. Overall, there are other hospitals where the categories assigned are not much better than the ones assigned to the previous hospitals, which suggests that the quality provided by the hospitals of the NHS is low, in the majority of the cases between the categories $C_1$ and $C_2$, as it is visible in Table 13.

## 5. Robustness analysis

This section performs a robustness analysis creating scenarios by changing some preference parameters' values, namely the credibility level and the criteria weights. For this scenario analysis, we firstly tested two different values for the credibility level: $\lambda = 0.55$ and $\lambda = 0.65$. In a second analysis, we tested three different scenarios that were created when the DM established the weights of the family of criteria using the SRF procedure, where we applied changes in the number of blank cards and the ratio-$z$. In the last test, we varied at the same time, the credibility level $\lambda$, the number of blank cards, and the ratio-$z$. Thus, we evaluate how changes in the inputs of our model influence its outputs. We discuss the results of this analysis.

### 5.1. Changing the credibility level

The change of the credibility level induces changes on the assignments, since it directly influences both ascending and descending joint rules:

1. If the credibility level $\lambda$ increases, the ascending rule tends to increase the category assigned for an action $a$. In contrast, the descending rule tends to decrease the category assigned for that same action. In general, there is a tendency for both rules to converge in the category to be assigned to an action $a$, possibly coinciding in a unique category;
2. If the credibility level $\lambda$ decreases, the ascending rule tends to decrease the category assigned for an action $a$. In contrast, the descending rule tends to increase the category assigned for that same action. In general, there is a tendency for both rules to diverge in the category to be assigned to an action $a$, possibly generating an interval of categories.

It suggests the existence of a critical $\lambda$ responsible for assigning the majority of the actions to a particular category [63]. As for that, in this robustness analysis, the credibility level was varied two times, assuming two distinct values, $\lambda = 0.55$ and $\lambda = 0.65$, to be compared to our analysis where $\lambda = 0.60$, to find a solution close to the critical value of $\lambda$ that assigned the majority of the actions to a unique category.

In Table 14 and Table 15 we present the assignment results for $\lambda = 0.55$ and $\lambda = 0.60$, and the results for $\lambda = 0.60$ and $\lambda = 0.65$, respectively. Note that in both tables, whenever it was registered a difference between the category or interval of categories assigned when $\lambda = 0.60$ and the two variations of $\lambda$, the cell was represented with a grey background for a better visualization.

Taking into account Table 14, considering the lower-level view of 2017, there were three alterations (12%), where $a_2$ was assigned to a unique category (rather than an interval of categories) resulting in a more substantial assignment relation comparing to $\lambda = 0.60$, as for $a_{13}$ and $a_{14}$ were assigned to a different category (but not an interval of categories).

**Table 14**
Assignment results for $\lambda = 0.55$ and $\lambda = 0.60$.

| | $\lambda = 0.55$ | | | | | | | | $\lambda = 0.60$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 2017 | | | | 2018 | | | | 2017 | | | | 2018 | | | |
| | Lower-level view | | Upper-level view | | Lower-level view | | Upper-level view | | Lower-level view | | Upper-level view | | Lower-level view | | Upper-level view | |
| Hospital | Min. | Max. | Max. | Max. | Min. | Max. | Min. | Max. | Min. | Max. | Min. | Max. | Min. | Max. | Min. | Max. |
| $a_1$ | $C_2$ | $C_2$ | $C_2$ | $C_4$ | $C_2$ | $C_3$ | $C_2$ | $C_4$ | $C_2$ | $C_2$ | $C_2$ | $C_3$ | $C_2$ | $C_2$ | $C_2$ | $C_4$ |
| $a_2$ | $C_3$ | $C_3$ | $C_4$ | $C_5$ | $C_3$ | $C_3$ | $C_4$ | $C_5$ | $C_3$ | $C_4$ | $C_4$ | $C_5$ | $C_3$ | $C_3$ | $C_4$ | $C_5$ |
| $a_3$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_2$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_2$ |
| $a_4$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_3$ | $C_5$ | $C_3$ | $C_5$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_3$ | $C_3$ | $C_3$ | $C_3$ |
| $a_5$ | $C_2$ | $C_3$ | $C_2$ | $C_4$ | $C_2$ | $C_3$ | $C_2$ | $C_4$ | $C_2$ | $C_3$ | $C_2$ | $C_4$ | $C_2$ | $C_2$ | $C_2$ | $C_3$ |
| $a_6$ | $C_1$ | $C_2$ | $C_2$ | $C_2$ | $C_1$ | $C_2$ | $C_2$ | $C_2$ | $C_1$ | $C_2$ | $C_2$ | $C_2$ | $C_1$ | $C_2$ | $C_2$ | $C_2$ |
| $a_7$ | $C_2$ | $C_2$ | $C_2$ | $C_4$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_4$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ |
| $a_8$ | $C_1$ | $C_2$ | $C_2$ | $C_2$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_2$ | $C_2$ | $C_2$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ |
| $a_9$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_4$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_3$ | $C_4$ |
| $a_{10}$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_2$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_2$ |
| $a_{11}$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_1$ | $C_1$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_1$ | $C_1$ | $C_2$ | $C_2$ |
| $a_{12}$ | $C_1$ | $C_1$ | $C_2$ | $C_2$ | $C_1$ | $C_2$ | $C_1$ | $C_2$ | $C_1$ | $C_1$ | $C_2$ | $C_2$ | $C_1$ | $C_2$ | $C_1$ | $C_2$ |
| $a_{13}$ | $C_1$ | $C_1$ | $C_2$ | $C_2$ | $C_2$ | $C_3$ | $C_3$ | $C_3$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_3$ | $C_3$ | $C_4$ |
| $a_{14}$ | $C_2$ | $C_2$ | $C_2$ | $C_4$ | $C_2$ | $C_4$ | $C_2$ | $C_4$ | $C_1$ | $C_1$ | $C_2$ | $C_3$ | $C_2$ | $C_3$ | $C_2$ | $C_3$ |
| $a_{15}$ | $C_1$ | $C_1$ | $C_1$ | $C_2$ | $C_1$ | $C_2$ | $C_1$ | $C_2$ | $C_1$ | $C_1$ | $C_2$ | $C_2$ | $C_1$ | $C_2$ | $C_2$ | $C_2$ |
| $a_{16}$ | $C_1$ | $C_2$ | $C_1$ | $C_2$ | $C_1$ | $C_2$ | $C_2$ | $C_3$ | $C_1$ | $C_2$ | $C_1$ | $C_2$ | $C_1$ | $C_2$ | $C_2$ | $C_2$ |
| $a_{17}$ | $C_1$ | $C_1$ | $C_2$ | $C_2$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_2$ | $C_2$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ |
| $a_{18}$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ |
| $a_{19}$ | $C_2$ | $C_2$ | $C_2$ | $C_3$ | $C_1$ | $C_2$ | $C_2$ | $C_3$ | $C_2$ | $C_2$ | $C_2$ | $C_3$ | $C_1$ | $C_2$ | $C_2$ | $C_3$ |
| $a_{20}$ | $C_1$ | $C_1$ | $C_2$ | $C_2$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_2$ | $C_2$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ |
| $a_{21}$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ |
| $a_{22}$ | $C_1$ | $C_2$ | $C_1$ | $C_2$ | $C_1$ | $C_2$ | $C_2$ | $C_2$ | $C_1$ | $C_2$ | $C_1$ | $C_2$ | $C_1$ | $C_2$ | $C_2$ | $C_2$ |
| $a_{23}$ | $C_1$ | $C_1$ | $C_1$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_3$ | $C_1$ | $C_1$ | $C_1$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_3$ |
| $a_{24}$ | $C_1$ | $C_2$ | $C_1$ | $C_2$ | $C_3$ | $C_3$ | $C_3$ | $C_3$ | $C_1$ | $C_2$ | $C_1$ | $C_2$ | $C_3$ | $C_3$ | $C_3$ | $C_3$ |
| $a_{25}$ | $C_1$ | $C_1$ | $C_1$ | $C_2$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_2$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ |

**Table 15**
Assignment results for $\lambda = 0.60$ and $\lambda = 0.65$.

| | $\lambda = 0.60$ | | | | | | | | $\lambda = 0.65$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 2017 | | | | 2018 | | | | 2017 | | | | 2018 | | | |
| | Lower-level view | | Upper-level view | | Lower-level view | | Upper-level view | | Lower-level view | | Upper-level view | | Lower-level view | | Upper-level view | |
| Hospital | Min. | Max. | Max. | Max. | Min. | Max. | Min. | Max. | Min. | Max. | Min. | Max. | Min. | Max. | Min. | Max. |
| $a_1$ | $C_2$ | $C_2$ | $C_2$ | $C_3$ | $C_2$ | $C_2$ | $C_2$ | $C_4$ | $C_2$ | $C_2$ | $C_2$ | $C_3$ | $C_2$ | $C_2$ | $C_2$ | $C_3$ |
| $a_2$ | $C_3$ | $C_4$ | $C_4$ | $C_5$ | $C_3$ | $C_3$ | $C_4$ | $C_5$ | $C_3$ | $C_3$ | $C_4$ | $C_5$ | $C_3$ | $C_3$ | $C_4$ | $C_5$ |
| $a_3$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_2$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_2$ |
| $a_4$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_3$ | $C_3$ | $C_3$ | $C_3$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_3$ | $C_2$ | $C_3$ |
| $a_5$ | $C_2$ | $C_3$ | $C_2$ | $C_4$ | $C_2$ | $C_2$ | $C_2$ | $C_3$ | $C_2$ | $C_3$ | $C_2$ | $C_4$ | $C_2$ | $C_2$ | $C_2$ | $C_3$ |
| $a_6$ | $C_1$ | $C_2$ | $C_2$ | $C_2$ | $C_1$ | $C_2$ | $C_2$ | $C_2$ | $C_1$ | $C_2$ | $C_2$ | $C_2$ | $C_1$ | $C_2$ | $C_2$ | $C_2$ |
| $a_7$ | $C_2$ | $C_2$ | $C_2$ | $C_4$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_4$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ |
| $a_8$ | $C_1$ | $C_2$ | $C_2$ | $C_2$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_2$ | $C_2$ | $C_2$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ |
| $a_9$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_3$ | $C_4$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_3$ | $C_4$ |
| $a_{10}$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_2$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_2$ |
| $a_{11}$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_1$ | $C_1$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_1$ | $C_1$ | $C_2$ | $C_2$ |
| $a_{12}$ | $C_1$ | $C_1$ | $C_2$ | $C_2$ | $C_1$ | $C_2$ | $C_1$ | $C_2$ | $C_1$ | $C_1$ | $C_2$ | $C_2$ | $C_1$ | $C_2$ | $C_1$ | $C_2$ |
| $a_{13}$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_3$ | $C_3$ | $C_4$ | $C_1$ | $C_1$ | $C_2$ | $C_2$ | $C_2$ | $C_3$ | $C_3$ | $C_4$ |
| $a_{14}$ | $C_1$ | $C_1$ | $C_2$ | $C_3$ | $C_2$ | $C_3$ | $C_2$ | $C_3$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ |
| $a_{15}$ | $C_1$ | $C_1$ | $C_2$ | $C_2$ | $C_1$ | $C_2$ | $C_2$ | $C_2$ | $C_1$ | $C_1$ | $C_2$ | $C_2$ | $C_1$ | $C_2$ | $C_2$ | $C_2$ |
| $a_{16}$ | $C_1$ | $C_2$ | $C_1$ | $C_2$ | $C_1$ | $C_2$ | $C_2$ | $C_2$ | $C_1$ | $C_1$ | $C_1$ | $C_2$ | $C_1$ | $C_2$ | $C_2$ | $C_2$ |
| $a_{17}$ | $C_1$ | $C_1$ | $C_2$ | $C_2$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_2$ | $C_2$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ |
| $a_{18}$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ |
| $a_{19}$ | $C_2$ | $C_2$ | $C_2$ | $C_3$ | $C_1$ | $C_2$ | $C_2$ | $C_3$ | $C_2$ | $C_2$ | $C_2$ | $C_3$ | $C_1$ | $C_2$ | $C_2$ | $C_3$ |
| $a_{20}$ | $C_1$ | $C_1$ | $C_2$ | $C_2$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_2$ | $C_2$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ |
| $a_{21}$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ |
| $a_{22}$ | $C_1$ | $C_2$ | $C_1$ | $C_2$ | $C_1$ | $C_2$ | $C_2$ | $C_2$ | $C_1$ | $C_2$ | $C_1$ | $C_2$ | $C_1$ | $C_2$ | $C_2$ | $C_2$ |
| $a_{23}$ | $C_1$ | $C_1$ | $C_1$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_3$ | $C_1$ | $C_1$ | $C_1$ | $C_2$ | $C_2$ | $C_2$ | $C_2$ | $C_3$ |
| $a_{24}$ | $C_1$ | $C_2$ | $C_1$ | $C_2$ | $C_3$ | $C_3$ | $C_3$ | $C_3$ | $C_1$ | $C_2$ | $C_1$ | $C_2$ | $C_3$ | $C_3$ | $C_3$ | $C_3$ |
| $a_{25}$ | $C_1$ | $C_1$ | $C_1$ | $C_2$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ | $C_2$ | $C_1$ | $C_1$ | $C_1$ | $C_1$ |

**Table 16**
Number of alterations in the assignment results comparing case 'DM' and $Z = 10$ and each of the scenarios varying ratio-$z$.

| | 2017 Lower-level view | | | 2017 Upper-level view | | | 2018 Lower-level view | | | 2018 Upper-level view | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Case | 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 |

Regarding the upper-level view of 2017, there were three alterations (12%), where all the hospitals were assigned to an interval of categories comparing to $\lambda = 0.60$ resulting in a weaker assignment, note that in the case of $a_1$ and $a_{14}$ both were assigned to a non-consecutive interval of categories.

Regarding the lower-level view of 2018, there were four alterations (16%), where the hospitals were assigned to an interval of categories comparing to $\lambda = 0.60$ resulting in a weaker assignment, note that in the case of $a_4$ and $a_{14}$ both were assigned to a non-consecutive interval of categories.

As to the upper-level viewpoint of 2018, there were seven alterations (28%), where six of the hospitals were again assigned to an interval of categories from which only $a_{15}$ and $a_{16}$ were assigned to an interval of consecutive categories comparing to $\lambda = 0.60$ resulting in a weaker assignment, the exception was $a_{13}$ that was assigned to a unique category resulting in a more robust assignment.

All in all, there were eighteen alterations, where fifteen of them (83%) resulted in hospitals assigned to an interval of categories and only three of them (17%) resulted in assignments to a unique category.

In the case where it was used $\lambda = 0.65$, considering the lower-level view of 2017, there were four alterations (16%) where $a_2$ and $a_{16}$, were assigned to a unique category resulting in a more substantial assignment relation comparing to $\lambda = 0.60$, $a_{13}$ and $a_{14}$ changed from a unique category to another. Regarding the upper-level view of 2017, there was only one alteration (4%) assigned to a unique category comparing to $\lambda = 0.60$, resulting in a more robust assignment.

Regarding the lower-level view of 2018, there were two alterations (8%), where $a_{14}$ was assigned to a unique category compared to $\lambda = 0.60$, resulting in a more substantial assignment relation; the opposite happened that $a_4$ was assigned to an interval of categories.

As to the upper-level viewpoint of 2018, there were three alterations (12%), where only $a_{14}$ was assigned to a unique category comparing to $\lambda = 0.60$, resulting in a more robust assignment, $a_1$ and $a_4$ were assigned to an interval of categories, resulting in a weaker assignment.

All in all, considering these ten alterations, three of them (30%) resulted in hospitals assigned to an interval of categories and five of them (50%) resulted in hospitals assigned to a unique category, the other two alterations (20%) represent a change of unique category to another.

In the case where $\lambda = 0.55$, there was an increase in hospitals assigned to an interval of categories, suggesting that this level of credibility is not close to the critical level. Regarding the case where $\lambda = 0.65$, the balance reflects three more hospitals assigned to a unique category than for the case where $\lambda = 0.60$, which suggests that the critical level of credibility might be between these two values of $\lambda$.

### 5.2. Changing the weights

We tested how variations the SRF parameters, which were responsible for attributing the weight to the criteria, influenced our model's results. The SRF procedure makes use of the DM knowledge, for ranking the criteria, adding the blank cards, and choosing the ratio-$z$, as stated in Section 3.7.2. At the time, the DM ranked the five criteria, and when he was adding blank cards, he was not sure whether to add two or three blank cards between the *Access* ($g_1$) and the *Caesarean Appropriateness* ($g_5$) criteria; and also whether to add three or four cards blank cards between the *Caesarean Appropriateness* ($g_5$) and the *Efficiency* ($g_4$) criteria. In the end, the DM opted for adding three blank cards between each of the criteria in these cases, however this situation allowed us to create three different cases:

1. Case A: Two blank cards were added between *Access* ($g_1$) and the *Caesarean Appropriateness* ($g_5$), instead of three cards;
2. Case E: Four blank cards were added between *Caesarean Appropriateness* ($g_1$) and *Efficiency* ($g_4$), instead of three cards;
3. Case I: Two blank cards were added between *Access* ($g_1$) and *Caesarean Appropriateness* ($g_5$) criterion, instead of three, and also four blank cards between *Caesarean Appropriateness* ($g_5$) and the *Efficiency* ($g_4$), instead of three.

Furthermore, we opted to include two variations of the ratio-$z$, since this ratio influences the weight attributed to each of the criteria. Thus, the variations of the ratio-$z$ were $Z = 9$ and $Z = 11$, to contrast with the value chosen by the DM, $Z = 10$. Note that these variations were only tested for the attribution of weight of the criteria and not to the groups of subcriteria.

Due to the extensively of the results, we resume the number of alterations in Table 16. We observe that in a universe of 1100 assignments, only seven alterations of categories (0.64%) were registered, which shows that the results are genuinely consistent, even when using different parameters in the model.

### 5.3. Changing both the credibility level and weights

Here we tested both previous changes on the parameters at the same time, that is, we analyzed several scenarios that resulted from the combinations of the following:

1. Credibility level: $\lambda = 0.55$, $\lambda = 0.60$, and $\lambda = 0.65$;
2. Blank cards: Three cases, A, E, and I, created in the previous section;
3. Ratio-$z$: $Z = 9$, $Z = 10$, and $Z = 11$.

Table 17 summarizes the number of alterations resulting from comparing the new conditions and the DM case with $Z = 10$ and $\lambda = 0.60$.

As observed in Table 17, there are 192 alterations in 2400 possibilities, which represents 8% of alterations, a small percentage taking into account that in these tests it was changed from one parameter ($\lambda$) until three parameters ($\lambda$, case, and ratio-$z$). There were less alterations for $\lambda = 0.65$, than for $\lambda = 0.55$, what once again it is related to the fact that the critical $\lambda$ is closer to $\lambda = 0.6$ and $\lambda = 0.65$. The maximum number of alterations of assignments was registered for all the scenarios of the upper-level view of 2018 with $\lambda = 0.55$, with six alterations per scenario, and the minimum

**Table 17**
Number of alterations in the assignment results comparing case 'DM' and $Z = 10$ and each of the scenarios varying ratio-$z$ and the credibility level $\lambda$.

| | 2017 Lower-level view | | | 2017 Upper-level view | | | 2018 Lower-level view | | | 2018 Upper-level view | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Z=9 | Z=10 | Z=11 | Z=9 | Z=10 | Z=11 | Z=9 | Z=10 | Z=11 | Z=9 | Z=10 | Z=11 |
| Case | | | | | | | $\lambda$=0.55 | | | | | |
| DM | 1 | 1 | 1 | 3 | 3 | 3 | 4 | 4 | 4 | 6 | 6 | 6 |
| A | 1 | 1 | 1 | 3 | 3 | 3 | 3 | 3 | 3 | 6 | 6 | 6 |
| E | 1 | 1 | 1 | 3 | 3 | 3 | 4 | 3 | 3 | 6 | 6 | 6 |
| I | 1 | 1 | 1 | 3 | 3 | 3 | 3 | 5 | 3 | 6 | 6 | 6 |
| Case | | | | | | | $\lambda$=0.65 | | | | | |
| DM | 0 | 0 | 0 | 1 | 1 | 1 | 2 | 2 | 2 | 3 | 3 | 3 |
| A | 0 | 0 | 0 | 1 | 1 | 1 | 2 | 2 | 2 | 3 | 3 | 3 |
| E | 0 | 0 | 0 | 1 | 1 | 1 | 2 | 2 | 2 | 3 | 3 | 3 |
| I | 0 | 0 | 0 | 1 | 1 | 1 | 2 | 2 | 2 | 3 | 3 | 3 |

number of alterations registered was zero, which happened for all the scenarios of the lower-level view of 2017 with $\lambda = 0.65$.

All in all, this analysis with 96 scenarios, once again, proved that our model is consistent, because even though we inserted so many changes in the inputs, the output was nearly the same for all of them.

### 5.4. Discussion of the robustness analysis results

Aiming to test the robustness of the decision model constructed with the DM, we performed an extensive scenario analysis. It focused on changes in the credibility level and the criteria weights, creating three cases and inducing changes in the SRF procedure, namely in the order of the cards and in ratio-$z$.

We tested 96 scenarios, in total, and in 2400 assignments generated only 192 alterations were produced, equivalent to 8%, to our original model. It suggests that the model created is robust. Moreover, there were fewer changes alterations when considering the credibility level $\lambda = 0.65$, than when considering $\lambda = 0.55$, suggesting that the credibility level, validated by the DM, is a consistent value.

## 6. Health policy and managerial implications

Evaluating the hospitals' performance has been considered a *must-do* practice to ensure the financial sustainability of healthcare systems worldwide. In general, such an exercise bases itself on a single dimension of quality (efficiency), through benchmarking techniques, including (but not limited to) Data Envelopment Analysis and models alike, and Stochastic Frontier Analysis. Although coined as more objective, these benchmarking techniques hardly account for the DM's subjective assessment, which should be ubiquitous in any performance and policy recommendations resulting from it. Besides, the inclusion of qualitative data (like satisfaction with healthcare) in those techniques is, at least, questionable. Not to mention the compensatory/pure mathematical nature of these techniques in the computation of coefficients or multipliers, from which one tends to retrieve (likely biased) economic implications, including marginal products and marginal rates of substitution. Because of this, decision-makers may consider these techniques like *black-boxes*, not entirely trusting on their outputs. As the one carried out in this research, MCDA appears, then, as an alternative to evaluating hospital performance as a whole (not just efficiency). It is more appealing for the DM, who understand it more straightforwardly as its fundamental points of view and judgments are included in the analysis. The resulting policy and managerial implications are naturally under subjectivity. However, such an issue may be mitigated by considering at least one individual (*bona fide* and independent) representing each group of stakeholders (policymakers, hospital managers, clinical staff, non-clinical staff, labor unions, and citizens).

Our results may impact hospital financing. According to a budget defined every year, these entities are financed, fixing prices and payments based on the minimal observed unitary costs [64]. The indicators used in this research to quantify patient safety, care appropriateness, access, and cesarean appropriateness are also monitored by the Health Ministry and used for financed purposes [65]. proposed a pay-for-performance alternative to optimize payments, exploring the concepts of quality, access, and environmental subsets. The optimization of payments follows Data Envelopment Analysis, with additional constraints related to those subsets. Concerning the environmental subsets, the authors suggest fixing bandwidths, easily computed through well-known methods (such as the Silverman's bandwidth), to compare like with like. In the case of quality and access subsets, they suggest comparing each hospital with the ones as good as the former, or at least with the quality or access levels above a threshold defined by the DM. Although the proposed optimization tool gets promising results in substantial cost savings, there are some difficulties related to the definition of thresholds and maybe the bandwidths. Indeed, the DM may struggle with the imposition of a minimal level per quality or access indicator, which the thresholds intend to model. Therefore, instead of adopting not-so-straightforward methodologies to create those quality and access comparability subsets, we may use the outputs of Electre-Tri-nC and Electre-Tri-C together to form those subsets. Indeed, the procedure would be relatively simple: *(i)* start by classifying hospitals based on their performance, namely in terms of their quality and access; *(ii)* fix the lowest classification allowing a hospital to be part of the comparability subset, *e.g.*, the DM may impose that no hospital with the average performance or below can belong to the subset – thus, the subset contains hospitals with good or very good performance levels, only; *(iii)* in line with [64,65], construct a frontier composed of the hospitals forming the comparability subset using any benchmarking technique and the appropriate inputs (*e.g.*, operating costs) and outputs (*e.g.*, patients adjusted by complexity); *(iv)* project each hospital in the frontier and obtain the optimal values for inputs and outputs; and *(v)* estimate optimal payments using the optimal inputs and outputs from the last step.

The results achieved in this research are worrying, especially concerning the pandemic outbreak we are living in nowadays. Recently one of the most prestigious medical journals globally, The Lancet, published an editorial about the Portuguese SNS [13]. The publication concluded that it no longer meets the needs of the population. Portugal is one of four countries (out of 33 analyzed) where the public health expenditure was reduced between 2000 and 2017. As a result of that, the hospitals were not modernized, and the medical equipment became obsolete.

Moreover, the public care has been losing space to private care, which has enjoyed the migration of public medical workforce from the public service because of poor work conditions, lack of motivation from managers, and staff burnout. The application of a multicriteria approach to perform a quality assessment of the Portuguese public hospitals proved to be a reliable tool. If there is a need to reformulate and invest in the SNS, quality must be considered to assure patients' needs and safety, the workforce, and the health institutions. Our results support that the article's claims as hospitals of the Portuguese SNS exhibited relatively low performance levels. An intensive effort should be carried out to search for the best practices within this field, not only in Portugal but also in other countries, especially those whose healthcare systems are Beveridgian [66]. Nonetheless, and recalling the SARS-COV-2 pandemic outbreak, the lack of quality in Portuguese hospitals casts doubts about their capacity to treat patients most safely and appropriately. This effect is aggravated by the reduced response to chronic and acute patients, motivated by the exaggerated focus directed to the cases of COVID-19 and the inattention to the other cases.

## 7. Conclusions

According to our model and results, the hospitals were assigned to a category or an interval of categories. In general, there was maintenance or improvement of quality in the healthcare entities between 2017 and 2018, expressed both in lower and upper-level views. Based on the results, we can draw the following conclusions:

1. No hospitals were assigned to the best category. Thus, there was no under evaluation of the reference hospitals used to characterize such a category.
2. Regarding the worst category there were several hospitals assigned to. This fact was also sustained by the results, where actions were continuously being assigned to the two worst categories.
3. For example, considering the year with better results, 2018, and the upper-level view, 60% of hospitals were assigned to a category or an interval of categories equal or less than the worst performance category. Only three hospitals were assigned to a category higher than the neutral performance category;
4. *Centro Hospitalar de Póvoa do Varzim/Vila do Conde* was considered the best hospital in both years and views, inclusively being assigned to the highest categories twice. It suggests the possibility of using this entity as a benchmark.
5. After a robustness analysis, we verified that only 8% of the assignments could change, suggesting that our results are robust.

Models are developed to represent real situations. They are not perfect, neither our model is. Thereby the major limitations are hereby presented. In the case study, it was used the data present in the ACSS benchmarking, as it includes clinical quality indicators that were significant and reliable to the DM. Even though the benchmarking presents several quality indicators, according to the literature review there are more indicators, than the ones we had access to. Therefore, to increase the robustness of our criteria tree and consequently our model, it would be needed more data gathered and provided by the benchmark. Information centered in the outcomes, such as patient satisfaction, which is of difficult access, or about the infrastructures would be interesting to apply on this method. The results of this work highly depend on criteria preference, a task that it is developed conjointly between the analyst and the DM. Once we only had one DM to collaborate with us, we were deprived of generating more robust results from our model that would allow us to compare assignments and different perspectives. Having access to other DMs to test our model would provide information about the models adaptability to other opinions and scenarios.

The application of a multicriteria approach to perform a quality assessment of the Portuguese public hospitals proved to be a reliable tool. If there is a need to reformulate and invest in the SNS, quality must be considered to assure the patients' needs and safety, the workforce, and the health institutions.

## CRediT authorship contribution statement

**António Rocha:** Conceptualization, Methodology, Formal analysis, Investigation, Writing - original draft, Writing - review & editing. **Ana Sara Costa:** Writing - original draft, Writing - review & editing, Visualization. **José Rui Figueira:** Conceptualization, Methodology, Investigation, Writing - original draft, Writing - review & editing, Supervision, Funding acquisition. **Diogo Cunha Ferreira:** Writing - original draft, Writing - review & editing. **Rui Cunha Marques:** Writing - original draft, Writing - review & editing, Supervision, Funding acquisition.

## Acknowledgements

This work was supported by national funds through Fundação para a Ciência e a Tecnologia (FCT) with reference UIDB/GES/00097/2020 and in the framework of the project PTDC/EGE-OGE/30546/2017 (hSNS: Portuguese public hospital performance assessment using a multi-criteria decision analysis framework).

## Appendix A. ELECTRE TRI-nC: A multicriteria method

### A1. Overview

The ELECTRE methods family, stands for ELimination and Choice Expressing the REality (in French, *ELimination Et Choix Traduisant la REalit*), which has been developing since the introduction of ELECTRE I [67]. These methods have been used to deal with a wide range of real-world MCDA situations and are based on two main components [61,68–70]:

1. A multiple criteria aggregation procedure, which allows to build one or more outranking relations based on the performances of each action on each criterion with the perspective of comparing in a more comprehensive way each pair of actions;
2. An exploitation procedure, which is used to obtain adequate results, meaning that they bear in mind the nature of the *problematic* (choosing, ranking, or sorting).

An *outranking relation* is a binary relation, $S$, defined on the set of potential actions, $A$, such that $a$ is preferred to $b$ ($aSb$) if there are enough arguments to decide that $a$ is at least as good as $b$, while there is no essential argument to refute that statement [71].

ELECTRE TRI-nC is a sorting MCDA method that assigns a set of actions to a set of ordered and pre-defined categories, according to the performance of each action in a set of criteria [18]. Categories can be defined by one or more reference actions, while in ELECTRE TRI-C only one reference action per category is possible [25]. Note that an increase of the reference actions for the same category can contribute for enriching the definition of the category and allows to obtain more narrow intervals of categories to which an action can be assigned to [61].

ELECTRE TRI-nC uses other preference parameters, such as weights and veto thresholds. An interaction process between the

DM and the analyst should be adopted with the aim of assigning value to these parameters. The method uses two joint rules (the ascending and the descending rules), each of them responsible for attributing one category to an action, what can result in one single category or an interval of categories in case both rules do not coincide. Moreover, this method was conceived to verify a set of fundamental structural requirements (conformity, homogeneity, monotonicity, and stability) [18].

In Electre Tri-nC, there are two main components: the construction of outranking relations, through comparing the level of credibility with the credibility index, calculated between each action and the set of reference actions of each category; and the exploitation of the outranking relations through the two joint rules, which assigned each action to a category.

### A2. Main data

Let $A = \{a_1, a_2, \ldots, a_i\}$ denote the set of *potential actions*, which can be fully known a priori or be progressively built up during the decision aiding process [72,73]. The method intends to assign these actions to a set of *completely ordered categories*, denoted $C = \{C_1, C_2, \ldots, C_h, \ldots, C_q\}$, with $q \geqslant 2$. As for that, it is necessary to build a *coherent family of criteria* [74], denoted $G = \left\{g_1, g_2, \ldots, g_j, \ldots, g_n\right\}$, which will be used to assess the *performance* of the potential actions, $g_j(a)$.

### A3. Parameters

A set of preference parameters is necessary for constructing a Electre Tri-nC model. The categories are defined by a set of *reference actions*, denoted $B = \{B_1, B_2, \ldots, B_h, \ldots, B_q, \ldots\}$, where $B_h = \left\{b_h^r, r = 1, \ldots, m_h\right\}$ is a subset of the reference actions (actual or dummy) that characterize category $C_h$, such that $m_h \geqslant 1$ and $h = 1,...,q$ [18] (in Electre Tri-C, $m_h = 1$, for all $h$). The method uses pairwise comparison between each action from $A$ and each action form $B$. A *weight*, $w_j$, such that $w_j > 0$, with $j = 1, \ldots, n$, representing the relative importance of the criterion, has to be defined for each criterion. Also a veto threshold can be associated to each criterion, $v_j$, contributing to avoid systematic compensatory effects. The last preference parameter is the *credibility level*, denoted $\lambda$ (see Subsection A.4).

Two technical parameters, *discriminating thresholds*, can be associated to the criteria [75]. Each criterion, $g_j$, is considered a pseudo-criterion or criterion with thresholds, since it is associated with two thresholds: the *preference threshold*, $p_j$, between the performance of two actions, corresponds to the smallest performance difference that, when exceeded, the best performing action is considered to be strictly preferable, and the *indifference threshold*, $q_j$, between the performance of two actions, corresponds to the largest performance difference that is judged compatible, with a situation of indifference between two actions, with different performances. Notice that $p_j \geqslant q_j \geqslant 0$. The purpose of these thresholds is to take in account the imperfect character of the data from the computation of the performances $g_j(a)$, for all $a \in A$, as well as the arbitrariness that affects the definition of the criteria. It is expected that all criteria $g_j \in G$ are to be maximized, thus the preferences increase when the criteria performances increase too [25]. When using the mentioned thresholds, the following binary relations can be derived for each criterion when comparing an action $a$ and action $b$ (they represent an action from set $A$ and a generic reference action from set $B$, respectively, for the sake of simplicity of notation):

1. $\left|g_j(a) - g_j(b)\right| \leqslant q_j$ where $a$ is indifferent to $b$ according to criterion $g_j$, in the right notation written $aI_jb$;
2. $g_j(a) - g_j(b) > p_j$ where $a$ is strictly preferable to $b$ according to criterion $g_j$, represented by $aP_jb$;
3. $q_j < q_j(a) - q_j(b) \leqslant p_j$ where the judgment is ambiguous, and there are no sufficient reasons to conclude an indifference situation, nor a strict preference between the two action. There is a hesitation between indifference and strict preference, meaning that $a$ it is weakly preferable to $b$, represented by $aQ_jb$.

Regarding the statement, $p_j \geqslant q_j \geqslant 0$, $q_j$ may be equal to zero and/or equal to $p_j$. If $p_j = 0$, any difference of performances in favor of one action over another can be considered as significant for a strict preference on criterion $g_j$. On the other hand, this is not always true as a result of the imperfect data characteristics or to arbitrariness that may affect the definition of the criteria.

### A4. Construction of an outranking relation

As previously stated, an outranking relation is represented by $aS_jb$, which means that "the action $a$ is at least as good as $b$", according to a criterion $g_j$. For the construction of outranking relations one needs to consider three concepts that allow to justify the same construction: agreement, non-disagreement and degree of credibility [18]:

1. *Concordance*: It refers to the conformity between criteria that favors $aS_jb$ to be accepted, meaning a sufficient majority of criteria must be in favor of this relation. The concordance is estimated by the global concordance index, $c(a, b)$, that associates each criterion to a weight $w_j$, such that $w_j > 0$, with $j = 1, \ldots, n$ and $\sum_{j=1}^{n} w_j = 1$ (assuming that the sum of all weights is 1). By definition:

$$c(a,b) = \sum_{j\in C(aPb)} w_j + \sum_{j\in C(aQb)} w_j + \sum_{j\in C(aIb)} w_j + \sum_{j\in C(aQb)} w_j\varphi_j \tag{A.1}$$

being the parameter $\varphi_j$ defined by:

$$\varphi_j = \frac{p_j - \left(g_j(b) - g_j(a)\right)}{p_j - q_j} \in [0, 1[. \tag{A.2}$$

2. *Non-discordance*: When none of the minority criteria that opposes $aS_jb$ exercises its power to veto this assertion, in other words refuting it. The discordance is estimated by the discordance index that associates each criterion to a veto power, $v_j$, such that $v_j > p_j$. The veto effect is modeled using the partial discordance index, $d_j(a, b)$, $j = 1, \ldots, n$, and is defined as:

$$d_j(a,b) = \begin{cases} 0 & \text{if } g_j(a) - g_j(b) \geqslant -p_j \\ \frac{g_j(a)-g_j(b)+p_j}{p_j - v_j} & \text{if } -v_j \leqslant g_j(a) - g_j(b) < -p_j \\ 1 & \text{if } g_j(a) - g_j(b) < -v_j \end{cases}$$

3. *Credibility index*: Denoted by $\sigma(a, b)$, it is the degree of credibility to consider that the action "$a$ is at least as good as the action $b$", taking into account the family of criteria, defined by F. To estimate this index is considered the global concordance index and the partial discordance index, following expression:

$$\sigma(a,b) = c(a,b)\prod_{j=1}^{n} T_j(a,b) \tag{A.3}$$

being $T_j(a, b)$ defined by:

$$T_j(a,b) = \begin{cases} \frac{1-d_j(a,b)}{1-c(a,b)} & \text{if } d_j(a,b) > c(a,b) \\ 1 & \text{otherwise} \end{cases}$$

The Electre Tri-nC method, defines a credibility level as the minimum degree of credibility denoted by $\lambda$, which is necessarily considered by the DM for validating or not an outranking statement taking in account all criteria from $G$. The credibility level can be seen as a cutting level, since it converts a fuzzy relation into a crisp outranking relation [76]. Typically, $\lambda$ takes a value within the range ]0.5,1] [18].

For the definition of the following outranking relations, the credibility level $\lambda$ is compared to the credibility indices of the different actions and to the set of reference actions of each category: $\sigma(a, B_h) = max_{r=1,\ldots,m_h}\{\sigma(a, b_h^r)\}$ and $\sigma(B_h, a) = max_{s=1,\ldots,m_h}\{\sigma(b_h^s, a)\}$. The credibility level allows to define four $\lambda$-binary relations that later assume a role to propose an assignment, the relations are presented below:

a) $\lambda$-outranking: $aS^{\lambda}B_h \Leftrightarrow \sigma(a, B_h) \geqslant \lambda$;
b) $\lambda$-preference: $aP^{\lambda}B_h \Leftrightarrow \sigma(a, B_h) \geqslant \lambda \ \wedge \ \sigma(B_h, a) < \lambda$;
c) $\lambda$-indifference: $aI^{\lambda}B_h \Leftrightarrow \sigma(a, B_h) \geqslant \lambda \ \wedge \ \sigma(B_h, a) \geqslant \lambda$;
d) $\lambda$-incomparability: $aR^{\lambda}B_h \Leftrightarrow \sigma(a, B_h) < \lambda \ \wedge \ \sigma(B_h, a) < \lambda$.

*A5. Exploitation of an outranking relation: The assignment procedure*

The assignment procedure is performed to attribute one category or an interval of categories, to an action $a$, which is compared to the reference actions $B_h$ considering the level of credibility $\lambda$. As for that, Electre Tri-nC makes use of two joint rules: the ascending rule and the descending rule. Both rules include a function $\rho(a, B_h)$ that allows the choice of one of two consecutive categories to be assigned to an action $a$. The selection function is following presented:

$$\rho(a, B_h) = min\{\sigma(a, B_h), \sigma(B_h, a)\}. \tag{A.4}$$

The two joint rules are defined in order to assign one or more possible categories to an action $a$ [18]:

1. *Descending rule*: Choose a credibility level $\lambda$ in the range of ]0.5,1]; and decrease $h$ from $(q+1)$ until the first value, $t$, such that $\sigma(a, B_t) \geqslant \lambda$ ($C_t$ is called the descending pre-selected category):
   a) If $t = q$, select $C_q$ as possible category to assign action $a$;
   b) For $0 < t < q$, if $\rho(a, B_t) > \rho(a, B_{t+1})$, then select $C_t$ as a possible category to assign $a$; otherwise, select $C_{t+1}$;
   c) For $t = 0$, select $C_1$ as a possible category to assign $a$.
2. *Ascending rule*: Choose a credibility level $\lambda$ in the range of ]0.5,1]; and increase $h$ from 0 until the first value of $k$, such that $\sigma(B_k, a) \geqslant \lambda$ ($C_k$ is called the ascending pre-selected category):
   a) For $k = 1$, select $C_1$ as a possible category to assign action $a$;
   b) For $1 < k < (q+1)$, if $\rho(\{a\}, B_k) > \rho(a, B_{k-1})$, then select $C_k$ as a possible category to assign $a$; otherwise, select $C_{k-1}$;
   c) For $k = (q+1)$, select $C_1$ as a possible category to assign $a$.

Each rule selects one possible category to each action. However, the fact that rules act simultaneously, can result in two different possibilities: the minimum and maximum categories are overlapped that results in the attribution of one single category; or the minimum and maximum categories are different, which results in an interval of categories.

## Appendix B. Additional tables

**Table B.1**
Criteria and subcriteria.

| Criterion | Subcriterion | Measure | Direction |
|---|---|---|---|
| $g_1$, Access | $g_{1,1}$: First medical appointments timeliness | Number of non-urgent first medical appointments performed in adequate time per 100 first medical appointments | Maximize |
| | $g_{1,2}$: Enrolled patients for surgery | Number of enrolled in the surgical waiting list within the mean guaranteed response time | Maximize |
| | $g_{1,3}$: Availability of beds | Difference between the real occupancy rate and the ideal occupancy rate | Minimize |
| | $g_{1,4}$: Availability of doctors | Doctors per 1000 inhabitants | Maximize |
| | $g_{1,5}$: Availability of nurses | Nurses per 1000 inhabitants | Maximize |
| $g_2$, Care Appropriateness | $g_{2,1}$: Minor surgeries appropriateness | Number of outpatient surgeries per 100 potential outpatient procedure | Maximize |
| | $g_{2,2}$: Avoidable re-admission prior 30 days after discharge | Number of readmissions in 30 days after discharge per 100 inpatients | Minimize |
| | $g_{2,3}$: Excessive staying delay | Number of long-stay inpatients per 100 admissions | Minimize |
| | $g_{2,4}$: Hip surgery timeliness | Number of hip surgeries performed in the first 48 hours per 100 hip surgeries | Maximize |
| | $g_{2,5}$: Delay before surgery | Average waiting time before surgery | Minimize |

(*continued on next page*)

**Table B.1** (*continued*)

| Criterion | Subcriterion | Measure | Direction |
|---|---|---|---|
| $g_3$, Patient Safety | $g_{3,1}$: Bedsores | Number of bedsores per 100 inpatients | Minimize |
| | $g_{3,2}$: Bloodstream infections related to CVC | Bloodstream infection rate related to CVC per 100 inpatients | Minimize |
| | $g_{3,3}$: Postoperative pulmonary embolisms or thrombosis | Postoperative pulmonary embolism/deep venous thrombosis cases per 100 surgical procedures | Minimize |
| | $g_{3,4}$: Postopertative septicaemia | Postoperative septicemia cases per 100 inpatients | Minimize |
| | $g_{3,5}$: Non-instrumental vaginal deliveries with severe laceration | Cases of trauma on vaginal delivery (third and fourth degree lacerations), without instrumentation, per 100 assisted deliveries | Minimize |
| | $g_{3,6}$: Assisted vaginal deliveries with severe laceration | Cases of trauma on vaginal delivery (third and fourth degree lacerations), with instrumentation, per 100 assisted deliveries | Minimize |
| $g_4$, Efficiency | $g_{4,1}$: Expenses with staff | Expenses with staff per severity-adjusted patient | Minimize |
| | $g_{4,2}$: Expenses with drugs, pharmaceutical products and clinical consumables | Expenses with drugs, pharmaceutical products and clinical consumables per severity-adjusted patient | Minimize |
| | $g_{4,3}$: Expenses with supplies and external services | Expenses with supplies and external services per severity-adjusted patient | Minimize |
| | $g_{4,4}$: Expenses with overtime | Expenses with overtime per total expenses with staff | Minimize |
| | $g_{4,5}$: Expenses with outsourcing | Expenses with outsourcing per total expenses with staff | Minimize |
| $g_5$, Caesarean Appropriatenesss | $g_{5,1}$: Volume of caesarean sections | Number of cesarean sections per 100 deliveries | Minimize |
| | $g_{5,2}$: Caesarean sections in UCFTPs | Number of cesarean sections in UCFTPs per 100 sections in UCFTPs | Minimize |
| | $g_{5,3}$: First caesarean sections in UCFTPs | Number of first cesarean sections in UCFTPs per 100 deliveries in UCFTPs without cesarean section before | Minimize |

**Table B.2**
Performance table for 2017 for all subcriteria.

| Hospital | $g_{1,1}$ | $g_{1,2}$ | $g_{1,3}$ | $g_{1,4}$ | $g_{1,5}$ | $g_{2,1}$ | $g_{2,2}$ | $g_{2,3}$ | $g_{2,4}$ | $g_{2,5}$ | $g_{3,1}$ | $g_{3,2}$ | $g_{3,3}$ | $g_{3,4}$ | $g_{3,5}$ | $g_{3,6}$ | $g_{4,1}$ | $g_{4,2}$ | $g_{4,3}$ | $g_{4,4}$ | $g_{4,5}$ | $g_{5,1}$ | $g_{5,2}$ | $g_{5,3}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $a_1$ | 69.2 | 94.58 | 2.6 | 1.13 | 1.68 | 82.37 | 7.46 | 3.18 | 52.84 | 0.75 | 0.00 | 0.00 | 0.10 | 0.33 | 0.47 | 0.64 | 2089 | 544.99 | 649.00 | 13.4 | 7.80 | 31.7 | 33.0 | 33.0 |
| $a_2$ | 96.8 | 97.75 | 7.9 | 1.16 | 1.80 | 70.26 | 6.36 | 0.85 | 92.17 | 0.61 | 0.00 | 0.00 | 0.00 | 0.18 | 0.21 | 8.80 | 2298 | 334.63 | 660.00 | 14.4 | 8.40 | 29.0 | 26.1 | 26.1 |
| $a_3$ | 76.2 | 59.81 | 2.9 | 1.60 | 3.08 | 66.44 | 7.81 | 4.47 | 51.14 | 0.90 | 0.04 | 0.06 | 0.11 | 0.59 | 0.37 | 3.16 | 1941 | 699.15 | 511.00 | 11.6 | 6.20 | 29.1 | 29.9 | 29.9 |
| $a_4$ | 65.6 | 77.74 | 7.1 | 1.03 | 2.14 | 82.13 | 9.22 | 2.19 | 32.38 | 0.79 | 0.00 | 0.00 | 0.00 | 0.05 | 0.08 | 0.79 | 1745 | 578.45 | 519.00 | 11.0 | 8.00 | 28.7 | 48.3 | 53.4 |
| $a_5$ | 75.0 | 61.97 | 0.8 | 1.72 | 3.07 | 88.25 | 4.94 | 3.90 | 67.09 | 1.03 | 0.02 | 0.00 | 0.09 | 0.34 | 1.36 | 3.21 | 1866 | 1153.02 | 572.00 | 13.2 | 8.90 | 31.9 | 40.6 | 41.8 |
| $a_6$ | 61.3 | 76.14 | 0.6 | 1.36 | 2.44 | 82.28 | 7.22 | 2.72 | 58.29 | 0.52 | 0.17 | 0.00 | 0.15 | 0.67 | 1.41 | 2.22 | 2030 | 746.17 | 488.00 | 12.0 | 5.60 | 23.5 | 29.0 | 28.2 |
| $a_7$ | 57.6 | 78.79 | 1.0 | 1.27 | 2.05 | 80.03 | 6.29 | 2.68 | 22.11 | 0.59 | 0.00 | 0.01 | 0.05 | 0.10 | 0.62 | 3.18 | 1799 | 689.04 | 468.00 | 14.8 | 7.90 | 25.2 | 28.0 | 28.0 |
| $a_8$ | 67.4 | 73.84 | 5.9 | 1.38 | 3.19 | 82.07 | 9.83 | 3.36 | 36.06 | 0.74 | 0.00 | 0.39 | 0.03 | 0.45 | 0.63 | 1.53 | 2042 | 670.77 | 846.00 | 11.7 | 16.10 | 27.4 | 28.2 | 28.2 |
| $a_9$ | 47.2 | 83.25 | 6.8 | 0.79 | 1.26 | 83.41 | 6.31 | 3.17 | 53.31 | 0.89 | 0.03 | 0.01 | 0.15 | 0.12 | 0.41 | 1.51 | 1535 | 539.33 | 546.00 | 12.0 | 9.30 | 22.7 | 22.8 | 22.8 |
| $a_{10}$ | 69.8 | 77.59 | 6.8 | 2.40 | 4.78 | 73.22 | 8.74 | 3.52 | 63.36 | 0.64 | 0.32 | 57.16 | 1.52 | 9.23 | 1.76 | 3.24 | 2286 | 766.22 | 619.00 | 13.9 | 2.90 | 36.7 | 48.0 | 48.0 |
| $a_{11}$ | 60.9 | 63.64 | 14.7 | 1.01 | 1.59 | 80.73 | 8.27 | 4.10 | 49.03 | 0.63 | 0.00 | 0.02 | 0.11 | 0.27 | 0.49 | 3.79 | 1402 | 988.09 | 449.00 | 11.0 | 5.20 | 28.5 | 29.5 | 29.5 |
| $a_{12}$ | 67.8 | 54.71 | 4.4 | 1.82 | 3.61 | 85.56 | 10.43 | 3.27 | 25.25 | 0.67 | 0.01 | 0.00 | 0.08 | 0.37 | 1.14 | 4.31 | 1945 | 843.51 | 653.00 | 15.0 | 5.70 | 28.2 | 29.6 | 29.5 |
| $a_{13}$ | 67.9 | 57.80 | 9.6 | 2.83 | 3.48 | 90.96 | 8.42 | 4.48 | 34.64 | 1.86 | 0.03 | 0.02 | 0.07 | 0.43 | 0.78 | 4.65 | 1695 | 754.83 | 374.00 | 12.6 | 2.20 | 22.4 | 28.3 | 28.3 |
| $a_{14}$ | 67.5 | 70.91 | 1.5 | 1.20 | 2.12 | 88.84 | 10.78 | 2.99 | 78.56 | 0.98 | 0.01 | 0.09 | 0.04 | 0.21 | 0.14 | 0.59 | 1719 | 709.42 | 515.00 | 14.0 | 4.20 | 37.7 | 48.7 | 48.7 |
| $a_{15}$ | 74.0 | 60.24 | 5.2 | 2.01 | 3.72 | 73.76 | 7.30 | 5.76 | 23.42 | 1.46 | 0.08 | 1.80 | 0.21 | 0.69 | 0.42 | 1.94 | 1977 | 920.05 | 608.00 | 14.6 | 7.00 | 26.6 | 25.8 | 25.5 |
| $a_{16}$ | 59.1 | 81.58 | 1.6 | 2.90 | 3.69 | 75.98 | 7.49 | 3.61 | 60.36 | 0.83 | 0.03 | 0.03 | 0.15 | 0.82 | 0.66 | 1.67 | 1540 | 991.53 | 306.00 | 12.0 | 3.10 | 29.4 | 30.6 | 30.6 |
| $a_{17}$ | 61.1 | 65.29 | 5.1 | 2.37 | 3.42 | 71.80 | 6.42 | 3.41 | 24.34 | 0.52 | 0.14 | 0.03 | 0.04 | 1.22 | 0.49 | 4.63 | 1687 | 850.57 | 661.00 | 14.0 | 4.70 | 34.0 | 38.8 | 31.0 |
| $a_{18}$ | 54.9 | 59.55 | 3.9 | 1.39 | 1.80 | 72.88 | 8.32 | 5.04 | 32.69 | 0.69 | 0.06 | 0.11 | 0.42 | 1.29 | 0.79 | 2.76 | 1363 | 826.27 | 569.00 | 15.5 | 7.30 | 31.6 | 51.8 | 38.0 |
| $a_{19}$ | 90.2 | 54.20 | 2.5 | 1.48 | 2.57 | 85.50 | 7.44 | 4.03 | 9.00 | 1.05 | 0.01 | 0.04 | 0.31 | 0.70 | 0.87 | 4.70 | 1413 | 903.95 | 411.00 | 12.1 | 5.60 | 26.3 | 25.8 | 25.4 |
| $a_{20}$ | 77.2 | 64.81 | 4.4 | 3.88 | 5.73 | 76.74 | 8.16 | 5.14 | 35.43 | 1.39 | 0.09 | 0.08 | 0.18 | 0.48 | 0.37 | 0.00 | 1575 | 1220.87 | 405.00 | 10.6 | 3.00 | 27.5 | 31.7 | 31.7 |
| $a_{21}$ | 73.3 | 66.10 | 8.6 | 3.38 | 5.23 | 79.04 | 9.15 | 4.14 | 39.04 | 1.30 | 0.02 | 0.03 | 0.10 | 0.34 | 0.23 | 0.34 | 1530 | 1023.26 | 416.00 | 12.4 | 0.60 | 29.9 | 33.2 | 33.2 |
| $a_{22}$ | 82.6 | 59.31 | 4.2 | 4.93 | 7.19 | 84.83 | 7.90 | 5.26 | 29.99 | 1.31 | 0.04 | 0.01 | 0.16 | 0.72 | 0.55 | 2.67 | 1640 | 1101.73 | 383.00 | 13.7 | 1.00 | 30.9 | 30.8 | 30.7 |
| $a_{23}$ | 61.2 | 78.34 | 3.4 | 4.64 | 7.09 | 77.67 | 6.83 | 3.35 | 55.60 | 0.95 | 0.09 | 0.04 | 0.23 | 0.61 | 0.70 | 2.21 | 1317 | 1012.11 | 339.00 | 13.8 | 1.40 | 28.0 | 27.2 | 26.7 |
| $a_{24}$ | 68.1 | 69.21 | 9.1 | 3.30 | 6.40 | 79.09 | 6.21 | 3.42 | 42.17 | 0.65 | 0.05 | 0.02 | 0.15 | 0.95 | 0.45 | 2.55 | 1372 | 1281.35 | 346.00 | 14.0 | 1.40 | 27.6 | 29.2 | 26.3 |
| $a_{25}$ | 69.5 | 71.77 | 1.4 | 4.45 | 5.58 | 80.70 | 4.89 | 4.49 | 48.73 | 0.98 | 0.02 | 0.02 | 0.18 | 1.35 | 0.62 | 4.49 | 1409 | 1361.12 | 375.00 | 11.8 | 1.10 | 24.9 | 28.8 | 28.8 |

**Table B.3**
Performance table for 2018 for all subcriteria.

| Hospital | $g_{1,1}$ | $g_{1,2}$ | $g_{1,3}$ | $g_{1,4}$ | $g_{1,5}$ | $g_{2,1}$ | $g_{2,2}$ | $g_{2,3}$ | $g_{2,4}$ | $g_{2,5}$ | $g_{3,1}$ | $g_{3,2}$ | $g_{3,3}$ | $g_{3,4}$ | $g_{3,5}$ | $g_{3,6}$ | $g_{4,1}$ | $g_{4,2}$ | $g_{4,3}$ | $g_{4,4}$ | $g_{4,5}$ | $g_{5,1}$ | $g_{5,2}$ | $g_{5,3}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $a_1$ | 76.1 | 96.00 | 3.2 | 1.14 | 1.80 | 83.10 | 7.30 | 3.96 | 29.70 | 0.75 | 0.02 | 0.00 | 0.14 | 0.46 | 0.17 | 0.93 | 2559 | 541.65 | 566.72 | 10.6 | 6.80 | 31.7 | 33.0 | 33.0 |
| $a_2$ | 95.8 | 99.20 | 7.6 | 1.19 | 1.83 | 69.80 | 6.56 | 1.52 | 84.30 | 0.54 | 0.00 | 0.00 | 0.00 | 0.22 | 0.34 | 1.52 | 2566 | 341.12 | 598.84 | 14.4 | 7.30 | 29.0 | 26.1 | 26.1 |
| $a_3$ | 83 | 69.40 | 0.1 | 1.63 | 3.09 | 70.20 | 8.38 | 4.90 | 45.40 | 0.99 | 0.02 | 0.01 | 0.08 | 0.48 | 0.28 | 2.97 | 2063 | 845.06 | 526.06 | 9.3 | 6.04 | 29.1 | 29.9 | 29.9 |
| $a_4$ | 56.2 | 82.70 | 0.6 | 1.06 | 2.15 | 88.50 | 8.18 | 2.31 | 32.40 | 0.74 | 0.01 | 0.00 | 0.00 | 0.09 | 0.00 | 0.27 | 1759 | 2543.22 | 537.76 | 11.9 | 6.72 | 28.7 | 48.3 | 53.4 |
| $a_5$ | 74.1 | 62.80 | 2.8 | 1.81 | 3.09 | 81.80 | 5.36 | 3.85 | 65.00 | 0.87 | 0.10 | 0.00 | 0.08 | 1.02 | 0.99 | 4.35 | 2022 | 691.12 | 534.50 | 10.5 | 8.43 | 31.9 | 40.6 | 41.8 |
| $a_6$ | 66.4 | 82.20 | 1.1 | 1.48 | 2.46 | 83.10 | 6.93 | 2.96 | 66.50 | 0.48 | 0.05 | 0.00 | 0.23 | 0.65 | 1.35 | 2.26 | 2135 | 525.45 | 508.70 | 12.1 | 5.25 | 23.5 | 29.0 | 28.2 |
| $a_7$ | 59 | 80.90 | 4.1 | 1.15 | 1.65 | 84.60 | 6.90 | 2.65 | 23.80 | 0.63 | 0.00 | 0.00 | 0.20 | 0.21 | 0.37 | 2.04 | 2091 | 633.31 | 490.63 | 13.2 | 5.86 | 25.2 | 28.0 | 28.0 |
| $a_8$ | 83.9 | 72.70 | 10.4 | 1.41 | 3.16 | 85.20 | 9.78 | 4.13 | 31.30 | 0.64 | 0.00 | 0.00 | 0.03 | 0.65 | 0.46 | 0.57 | 2373 | 139.59 | 897.92 | 11.8 | 13.91 | 27.4 | 28.2 | 28.2 |
| $a_9$ | 60.4 | 85.90 | 6.2 | 0.87 | 1.35 | 85.70 | 6.69 | 3.50 | 59.00 | 0.85 | 0.04 | 0.00 | 0.14 | 0.36 | 0.33 | 0.00 | 1739 | 616.93 | 624.30 | 12.0 | 8.48 | 22.7 | 22.8 | 22.8 |
| $a_{10}$ | 76.3 | 68.00 | 5.1 | 2.45 | 4.84 | 72.90 | 7.20 | 3.71 | 53.10 | 0.70 | 0.13 | 0.00 | 0.11 | 0.39 | 0.45 | 0.00 | 2472 | 862.13 | 628.96 | 16.0 | 2.74 | 36.7 | 48.0 | 48.0 |
| $a_{11}$ | 57.1 | 74.10 | 1.5 | 1.09 | 1.57 | 79.60 | 7.98 | 4.57 | 43.50 | 0.60 | 0.00 | 0.00 | 0.07 | 0.64 | 0.64 | 3.22 | 1474 | 1200.58 | 499.27 | 11.5 | 5.04 | 28.5 | 29.5 | 29.5 |
| $a_{12}$ | 67.6 | 57.20 | 8.6 | 1.88 | 3.51 | 81.80 | 10.50 | 3.13 | 33.20 | 0.54 | 0.01 | 0.00 | 0.15 | 0.54 | 0.35 | 3.21 | 2702 | 945.41 | 827.94 | 13.7 | 6.58 | 28.2 | 29.6 | 29.5 |
| $a_{13}$ | 80.6 | 49.50 | 3.4 | 2.92 | 3.52 | 96.80 | 5.40 | 4.05 | 36.70 | 1.67 | 0.12 | 0.00 | 0.19 | 1.09 | 0.98 | 3.86 | 1736 | 841.26 | 419.30 | 11.3 | 2.43 | 22.4 | 28.3 | 28.3 |
| $a_{14}$ | 64.1 | 70.10 | 2.6 | 1.25 | 2.21 | 86.20 | 11.44 | 2.91 | 73.50 | 1.11 | 0.02 | 0.01 | 0.08 | 0.44 | 0.29 | 0.62 | 1932 | 696.23 | 611.23 | 12.0 | 4.33 | 37.7 | 48.7 | 48.7 |
| $a_{15}$ | 73.2 | 66.60 | 4.1 | 2.03 | 3.75 | 79.80 | 7.09 | 5.64 | 12.50 | 1.38 | 0.05 | 0.01 | 0.26 | 0.64 | 0.54 | 0.90 | 2318 | 1019.48 | 492.76 | 12.8 | 7.63 | 26.6 | 25.8 | 25.5 |
| $a_{16}$ | 55.7 | 80.80 | 1.9 | 2.78 | 3.51 | 84.60 | 7.44 | 3.76 | 71.80 | 0.96 | 0.08 | 0.02 | 0.18 | 0.61 | 0.32 | 1.99 | 1617 | 1108.94 | 329.66 | 13.2 | 2.58 | 29.4 | 30.6 | 30.6 |
| $a_{17}$ | 61.5 | 65.40 | 3.4 | 3.56 | 3.48 | 76.00 | 6.07 | 3.94 | 16.50 | 0.44 | 0.11 | 0.06 | 0.18 | 0.75 | 1.01 | 2.43 | 1872 | 987.87 | 625.34 | 14.4 | 4.23 | 34.0 | 38.8 | 31.0 |
| $a_{18}$ | 76.7 | 66.10 | 3.0 | 1.38 | 1.73 | 82.70 | 4.84 | 5.53 | 34.90 | 0.49 | 0.03 | 0.00 | 0.32 | 2.01 | 0.60 | 2.48 | 1412 | 822.24 | 573.02 | 15.1 | 7.26 | 31.6 | 51.8 | 38.0 |
| $a_{19}$ | 86.5 | 51.80 | 4.4 | 1.57 | 2.60 | 90.20 | 7.09 | 4.50 | 26.80 | 1.29 | 0.02 | 0.01 | 0.24 | 0.49 | 1.17 | 3.62 | 1378 | 874.07 | 404.06 | 11.4 | 5.13 | 26.3 | 25.8 | 25.4 |
| $a_{20}$ | 70.2 | 63.90 | 6.7 | 3.98 | 5.76 | 80.20 | 7.05 | 4.50 | 51.20 | 1.35 | 0.07 | 0.05 | 0.24 | 0.85 | 0.09 | 0.70 | 1878 | 1210.46 | 405.98 | 9.6 | 3.56 | 27.5 | 31.7 | 31.7 |
| $a_{21}$ | 66 | 65.70 | 8.0 | 3.45 | 5.38 | 81.10 | 8.83 | 4.06 | 39.20 | 1.35 | 0.03 | 0.00 | 0.14 | 0.39 | 0.14 | 1.03 | 1612 | 151.75 | 430.98 | 11.6 | 0.63 | 29.9 | 33.2 | 33.2 |
| $a_{22}$ | 75.1 | 60.90 | 5.7 | 5.08 | 7.04 | 83.80 | 6.54 | 4.82 | 35.40 | 1.30 | 0.07 | 0.00 | 0.21 | 0.56 | 0.82 | 3.27 | 2019 | 1216.67 | 435.95 | 10.2 | 1.14 | 30.9 | 30.8 | 30.7 |
| $a_{23}$ | 50 | 74.60 | 2.3 | 4.82 | 7.11 | 81.40 | 5.11 | 3.25 | 62.70 | 0.95 | 0.04 | 0.03 | 0.31 | 0.55 | 0.18 | 3.19 | 1594 | 1149.94 | 348.36 | 10.7 | 1.96 | 28.0 | 27.2 | 26.7 |
| $a_{24}$ | 72.7 | 79.70 | 11.3 | 3.50 | 5.20 | 78.40 | 1.40 | 2.73 | 58.30 | 0.37 | 0.09 | 0.00 | 0.16 | 0.09 | 0.41 | 1.65 | 1433 | 1410.01 | 310.35 | 11.9 | 1.45 | 27.6 | 29.2 | 26.3 |
| $a_{25}$ | 63.5 | 68.40 | 0.5 | 4.52 | 5.46 | 79.80 | 9.44 | 4.53 | 43.80 | 1.05 | 0.02 | 0.01 | 0.25 | 1.85 | 0.69 | 5.97 | 1491 | 1538.55 | 403.17 | 14.7 | 0.97 | 24.9 | 28.8 | 28.8 |

**Table B.4**
Performances of the reference hospitals per category for *Access* ($g_1$).

| Category | Performance | Reference hospital | Subcriterion | | | | |
|---|---|---|---|---|---|---|---|
| | | | $g_{1,1}$ | $g_{1,2}$ | $g_{1,3}$ | $g_{1,4}$ | $g_{1,5}$ |
| $C_5$ | Very Good | $b_5^1$ | 95.0 | 95.00 | 0.0 | 4.30 | 6.40 |
| $C_4$ | Good | $b_4^1$ | 85.0 | 85.00 | 2.0 | 2.70 | 4.80 |
| $C_3$ | Neutral | $b_3^1$ | 80.0 | 80.00 | 5.0 | 2.10 | 3.50 |
| $C_2$ | Poor | $b_2^1$ | 70.0 | 75.00 | 7.0 | 1.20 | 2.10 |
| $C_1$ | Very Poor | $b_1^1$ | 60.0 | 70.00 | 9.0 | 1.00 | 1.70 |

**Table B.5**
Performances of the reference hospitals per category for *Patient Safety* ($g_3$).

| Category | Performance | Reference hospital | Subcriterion | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | $g_{3,1}$ | $g_{3,2}$ | $g_{3,3}$ | $g_{3,4}$ | $g_{3,5}$ | $g_{3,6}$ |
| $C_5$ | Very Good | $b_5^1$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.05 | 0.15 |
| $C_4$ | Good | $b_4^1$ | 0.03 | 0.02 | 0.08 | 0.03 | 0.16 | 0.35 |
| $C_3$ | Neutral | $b_3^1$ | 0.05 | 0.04 | 0.15 | 0.06 | 0.33 | 0.64 |
| $C_2$ | Poor | $b_2^1$ | 0.10 | 0.06 | 0.21 | 0.76 | 0.64 | 1.00 |
| $C_1$ | Very Poor | $b_1^1$ | 0.12 | 0.09 | 0.31 | 1.20 | 0.81 | 2.05 |

**Table B.6**
Performances of the reference hospitals per category for *Efficiency* ($g_4$).

| Category | Performance | Reference hospital | Subcriterion | | | | |
|---|---|---|---|---|---|---|---|
| | | | $g_{4,1}$ | $g_{4,2}$ | $g_{4,3}$ | $g_{4,4}$ | $g_{4,5}$ |
| $C_5$ | Very Good | $b_5^1$ | 1408 | 537.94 | 348.12 | 10.6 | 1.14 |
| $C_4$ | Good | $b_4^1$ | 1536 | 689.56 | 407.24 | 11.6 | 2.62 |
| $C_3$ | Neutral | $b_3^1$ | 1829 | 847.82 | 509.85 | 12.08 | 5.22 |
| $C_2$ | Poor | $b_2^1$ | 2039 | 1082.11 | 605.71 | 13.98 | 7.29 |
| $C_1$ | Very Poor | $b_1^1$ | 2381 | 1287.73 | 660.08 | 14.82 | 8.51 |

**Table B.7**
Performances of the reference hospitals per category *Caesarean Appropriateness* ($g_5$).

| Category | Performance | Reference hospital | Subcriterion | | |
|---|---|---|---|---|---|
| | | | $g_{5,1}$ | $g_{5,2}$ | $g_{5,3}$ |
| $C_5$ | Very Good | $b_5^1$ | 15.0 | 15.0 | 80.0 |
| $C_4$ | Good | $b_4^1$ | 20.0 | 20.0 | 85.0 |
| $C_3$ | Neutral | $b_3^1$ | 28.4 | 31.2 | 90.0 |
| $C_2$ | Poor | $b_2^1$ | 30.0 | 30.0 | 95.0 |
| $C_1$ | Very Poor | $b_1^1$ | 35.0 | 35.0 | 100.0 |

**Table B.8**
Weight attributed by the DM to each subcriterion of *Access* ($g_1$).

| Criterion | Subcriterion | Non-normalized weight | Normalized weight |
|---|---|---|---|
| $g_1$, Access | $g_{1,1}$ | 2.00 | 27.10 |
| | $g_{1,2}$ | 1.88 | 25.47 |
| | $g_{1,3}$ | 1.50 | 20.33 |
| | $g_{1,4}$ | 1.00 | 13.55 |
| | $g_{1,5}$ | 1.00 | 13.55 |

**Table B.9**
Weight attributed by the DM to each subcriterion of *Patient Safety* ($g_3$).

| Criterion | Subcriterion | Non-normalized weight | Normalized weight |
|---|---|---|---|
| $g_3$, Patient Safety | $g_{3,1}$ | 5.09 | 14.69 |
| | $g_{3,2}$ | 7.55 | 21.79 |
| | $g_{3,3}$ | 10.00 | 28.87 |
| | $g_{3,4}$ | 10.00 | 28.87 |
| | $g_{3,5}$ | 1.00 | 2.89 |
| | $g_{3,6}$ | 1.00 | 2.89 |

**Table B.10**
Weight attributed by the DM to each subcriterion of *Efficiency* ($g_4$).

| Criterion | Subcriterion | Non-normalized weight | Normalized weight |
|---|---|---|---|
| $g_4$, Efficiency | $g_{4,1}$ | 7.43 | 20.72 |
| | $g_{4,2}$ | 10.00 | 27.88 |
| | $g_{4,3}$ | 10.00 | 27.89 |
| | $g_{4,4}$ | 7.43 | 20.72 |
| | $g_{4,5}$ | 1.00 | 2.79 |

**Table B.11**
Weight attributed by the DM to each subcriterion of *Caesarean Appropriateness* ($g_5$).

| Criterion | Subcriterion | Non-normalized weight | Normalized weight |
|---|---|---|---|
| $g_5$, Caesarean Appropriateness | $g_{5,1}$ | 1 | 20 |
| | $g_{5,2}$ | 2 | 40 |
| | $g_{5,3}$ | 2 | 40 |

**Table B.12**
Indifference and preference thresholds of the subcriteria.

| Threshold | Subcriterion | | | | | |
|---|---|---|---|---|---|---|
| | $g_{1,1}$ | $g_{1,2}$ | $g_{1,3}$ | $g_{1,4}$ | $g_{1,5}$ | |
| $q$ | 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 | | | |

**Table B.13**
The existing $\lambda$-binary relations between the hospitals and the categories for $\lambda = 0.6$.

| Hospital | $C_1$ | $C_2$ | $C_3$ | $C_4$ | $C_5$ |
|---|---|---|---|---|---|
| $a_1$ | $\succ$ | $I$ | $I$ | $I$ | $\prec$ |
| $a_2$ | $\succ$ | $\succ$ | $\succ$ | $I$ | $\prec$ |
| $a_3$ | $I$ | $I$ | $\prec$ | $\prec$ | $\prec$ |
| $a_4$ | $\succ$ | $\succ$ | $I$ | $\prec$ | $\prec$ |
| $a_5$ | $\succ$ | $I$ | $I$ | $\prec$ | $\prec$ |
| $a_6$ | $\succ$ | $I$ | $\prec$ | $\prec$ | $\prec$ |
| $a_7$ | $\succ$ | $I$ | $\prec$ | $\prec$ | $\prec$ |
| $a_8$ | $I$ | $\prec$ | $\prec$ | $\prec$ | $\prec$ |
| $a_9$ | $\succ$ | $\succ$ | $I$ | $I$ | $\prec$ |
| $a_{10}$ | $I$ | $I$ | $\prec$ | $\prec$ | $\prec$ |
| $a_{11}$ | $\succ$ | $I$ | $\prec$ | $\prec$ | $\prec$ |
| $a_{12}$ | $I$ | $I$ | $\prec$ | $\prec$ | $\prec$ |
| $a_{13}$ | $\succ$ | $\succ$ | $\succ$ | $R$ | $\prec$ |
| $a_{14}$ | $\succ$ | $I$ | $I$ | $\prec$ | $\prec$ |
| $a_{15}$ | $\succ$ | $I$ | $\prec$ | $\prec$ | $\prec$ |
| $a_{16}$ | $\succ$ | $I$ | $\prec$ | $\prec$ | $\prec$ |
| $a_{17}$ | $I$ | $\prec$ | $\prec$ | $\prec$ | $\prec$ |
| $a_{18}$ | $\succ$ | $I$ | $\prec$ | $\prec$ | $\prec$ |
| $a_{19}$ | $\succ$ | $I$ | $I$ | $\prec$ | $\prec$ |
| $a_{20}$ | $I$ | $\prec$ | $\prec$ | $\prec$ | $\prec$ |
| $a_{21}$ | $\succ$ | $I$ | $\prec$ | $\prec$ | $\prec$ |
| $a_{22}$ | $\succ$ | $I$ | $\prec$ | $\prec$ | $\prec$ |
| $a_{23}$ | $\succ$ | $I$ | $I$ | $\prec$ | $\prec$ |
| $a_{24}$ | $\succ$ | $\succ$ | $\prec$ | $\prec$ | $\prec$ |
| $a_{25}$ | $I$ | $\prec$ | $\prec$ | $\prec$ | $\prec$ |

## Appendix C. Additional figures

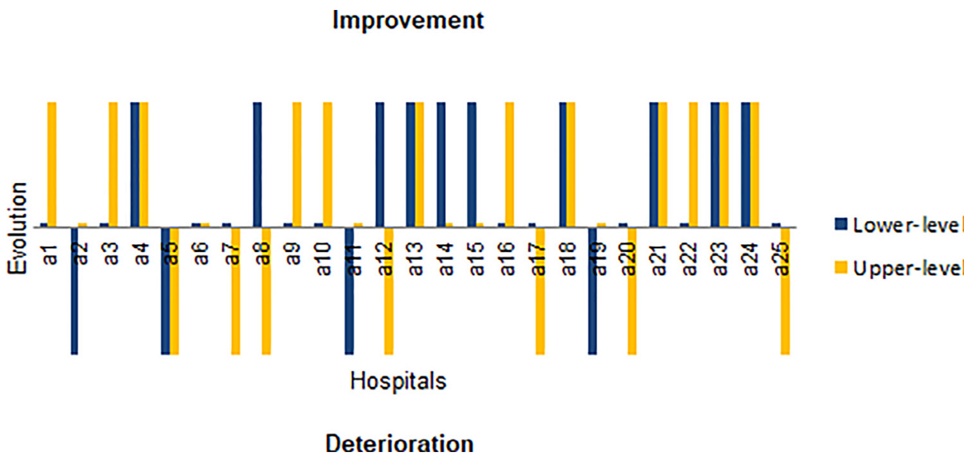


**Fig. C.3.** Evolution of the hospitals' quality from 2017 to 2018 for the lower-level and the upper-level views.

## References


[1] Boussemart J-P, Ferrier GD, Leleu H, Shen Z. An expanded decomposition of the luenberger productivity indicator with an application to the chinese healthcare sector. Omega: The International Journal of Management Science 2020;91:102010. doi:10.1016/j.omega.2018.11.019.

[2] Atun R, de Andrade LOM, Almeida G, Cotlear D, Dmytraczenko T, Frenz P, Garcia P, Gmez-Dants O, Knaul FM, Muntaner C, de Paula JB, Rgoli F, Serrate PC-F, Wagstaff A. Health-system reform and universal health coverage in latin america. The Lancet 2015;385(9974):1230–47. doi:10.1016/S0140-6736(14)61646-9.

[3] Potrafke N. The growth of public health expenditures in oecd countries: do government ideology and electoral motives matter? J Health Econ 2010;29(6):797–810. doi:10.1016/j.jhealeco.2010.07.008.

[4] Mossialos E, Le Grand J. Health care and cost containment in the european union, vol 50; 2000.

[5] Barros P. Economia da saúde - Conceitos e comportamentos. Coimbra (Almedina); 2013. ISBN 9789724054018.

[6] Qureshi MI, Qayyum S, Nassani AA, Aldakhil AM, Qazi Abro MM, Zaman K. Management of various socio-economic factors under the united nations sustainable development agenda. Resour Policy 2019;64:101515. doi:10.1016/j.resourpol.2019.101515.

[7] Gostin LO, Friedman EA. The sustainable development goals: one-Health in the worlds development agenda. JAMA 2015;314(24):2621–2. doi:10.1001/jama.2015.16281.

[8] Okereke M, Ukor NA, Adebisi YA, Ogunkola IO, Favour Iyagbaye E, Adiela Owhor G, et al. Impact of covid-19 on access to healthcare in low- and middle-income countries: current evidence and future recommendations. Int J Health Plann Manage 2020;36(1):13–17. doi:10.1002/hpm.3067.

[9] National Academies of SciencesEngineering, Medicine. Crossing the global quality chasm: improving health care worldwide; 2018. doi:1017226/25152.

[10] Di Matteo L. The sustainability of public health expenditures: evidence from the canadian federation. The European Journal of Health Economics 2010;11:569–84.

[11] Ferreira D, Nunes A, Marques R. Operational efficiency *vs* clinical safety, care appropriateness, timeliness, and access to health care. Journal of Productivity Analysis 2020;53:355–75.

[12] Azevedo H, Mateus C. Cost effects of hospital mergers in portugal. The European Journal of Health Economics 2014;15(9):999–1010.

[13] The Lancet. National health care in portugal: a new opportunity. The Lancet 2019;394(10206):12–98. doi:10.1016/S0140-6736(19)32278-0.

[14] Nunes A, Ferreira DC. Reforms in the portuguese health care sector: challenges and proposals. International Journal of Health Planning and Management 2019;34:21–33. doi:10.1002/hpm.2695.

[15] Ferreira DC, Marques RC, Nunes A. Economies of scope in the health sector: the case of portuguese hospitals. Eur J Oper Res 2018;266(2):716–35. doi:10.1016/j.ejor.2017.09.044.

[16] Reis M, Costa C, Mendes R, Lopes S. The funding system and incentives in portuguese hospitals. BMC Health Serv Res 2010;10:n/a.

[17] Ferreira DC, Nunes AM, Marques RC. Optimizing payments based on efficiency, quality, complexity, and heterogeneity: the case of hospital funding. International Transactions in Operational Research 2020;27(4):1930–61.

[18] Almeida-Dias J, Figueira JR, Roy B. A multiple criteria sorting method where each category is characterized by several reference actions: the ELECTRE TRI-NC method. Eur J Oper Res 2012;217(3):567–79. doi:10.1016/j.ejor.2011.09.047.

[19] Corrente S, Greco S, Słowiński R. Multiple criteria hierarchy process with electre and promethee. Omega: The International Journal of Management Science 2013;41(5):820–46.

[20] Corrente S, Greco S, Słowiński R. Multiple criteria hierarchy process for electre tri methods. Eur J Oper Res 2016;252(1):191–203.

[21] Keeney RL. Value-focused thinking. Harvard University Press; 1996.

[22] Barron FH, Person HB. Assessment of multiplicative utility functions via holistic judgments. Organ Behav Hum Perform 1979;24(2):147–66.

[23] Green PE, Srinivasan V. Conjoint analysis in consumer research: issues and outlook. Journal of consumer research 1978;5(2):103–23.

[24] Bana e Costa C, Corrêa ÉC, De Corte J-M, Vansnick J-C. Facilitating bid evaluation in public call for tenders: a socio-technical approach. Omega: The International Journal of Management Science 2002;30(3):227–42.

[25] Almeida-Dias J, Figueira JR, Roy B. ELECTRE TRI-C: a multiple criteria sorting method based on characteristic reference actions. Eur J Oper Res 2010;204(3):565–80. doi:10.1016/j.ejor.2009.10.018.

[26] Institute of Medicine. Crossing the quality chasm: A New health system for the 21st century. Washington, DC, USA: National Academy Press; 2001.

[27] Institute of Medicine. Performance measurement: accelerating improvement. Washington, DC, USA: National Academy Press; 2005.

[28] Beauchamp T, Childress J. Principles of biomedical ethics. Oxford, UK: Oxford University Press; 2008.

[29] Fee E, Garofalo M. Florence nightingale and the crimean war. Am J Public Health 2010;100(9):1591.

[30] Rothschild JM, Hurley AC, Landrigan CP, Cronin JW, Martell-Waldrop K, Foskett C, et al. Recovery from medical errors: the critical care nursing safety net. The Joint Commission Journal on Quality and Patient Safety 2006;32(2):63–72. doi:10.1016/S1553-7250(06)32009-0.

[31] Umscheid CA, Mitchell MD, Doshi JA, Agarwal R, Williams K, Brennan PJ. Estimating the proportion of healthcare-associated infections that are reasonably preventable and the related mortality and costs. Infection Control & Hospital Epidemiology 2011;32(2):101–14.

[32] McCradden MD, Joshi S, Anderson JA, Mazwi M, Goldenberg A, Zlotnik Shaul R. Patient safety and quality improvement: ethical principles for a regulatory approach to bias in healthcare machine learning. Journal of the American Medical Informatics Association 2020. doi:10.1093/jamia/ocaa085.

[33] Classen DC, Resar R, Griffin F, Federico F, Frankel T, Kimmel N, et al. Global trigger tool shows that adverse events in hospitals may be ten times greater than previously measured. Health Aff 2011;30(4):581–9.

[34] Andel C, Davidow S, Hollander M, Moreno D. The economics of health care quality and medical errors. J Health Care Finance 2012;39(1):39–50.

[35] Robertson-Preidler J, Biller-Andorno N, Johnson T. What is appropriate care? an integrative review of emerging themes in the literature. BMC Health Serv Res 2017;17(452):1–17.

[36] Ferreira DC, Marques RC, Nunes A, Figueira JR. Patients' Satisfaction: the medical appointments valence in portuguese public hospitals. Omega: The International Journal of Management Science 2018;80:58–76. doi:10.1016/j.omega.2017.08.009.

[37] Braspenning J, Hermens R, Calsbeek H, Campbell S, van der Wees P, Grol R. Indicators for Quality and Safety of Care. John Wiley & Sons, Ltd; 2020. p. 131–54.

[38] Paul R, Miller D. Cesarean birth: how to reduce the rate. Am J Obstet Gynecol 1995;172(6):1903–11.

[39] Moore M. Reducing the rate of cesarean birth. J Perinat Educ 2002;11(2):41–3.

[40] Chaillet N, Dumont A. Evidence-based strategies for reducing cesarean section rates: a meta-analysis. Birth 2007;34(1):53–64.

[41] Vadnais M, Hacker M, Shah N, Jordan J, Modest A, Siegel M, Golen T. Quality improvement initiatives lead to reduction in nulliparous term singleton vertex cesarean delivery rate. Joint Commission Journal on Quality and Patient Safety 2017;43(2):41–3.

[42] Lane H, Sarkies M, Martin J, Haines T. Equity in healthcare resource allocation decision making: a systematic review. Social Science & Medicine 2017;175:11–27.
[43] Kreng VB, Yang C-T. The equality of resource allocation in health care under the national health insurance system in taiwan. Health Policy 2011;100(2):203–10.
[44] Bambra C, Garthwaite K, Hunter D. All things being equal: does it matter for equity how you organize and pay for health care? a review of the international evidence. International Journal of Health Services 2014;44(3):457–77.
[45] Gulliford M, Figueroa-Munoz J, Morgan M, Hughes D, Gibson B, Beech R, Hudson M. What does 'access to health care' mean? Journal of Health Services Research & Policy 2002;7(3):186–8.
[46] Peters D, Garg A, Bloom G, Walker D, Brieger W, Rahman M. Poverty and access to health care in developing countries. Ann N Y Acad Sci 2008;1136:161–71.
[47] Donabedian A. The seven pillars of quality. Archives of Pathology & Laboratory Medicine 1990;114(11):1115–18.
[48] Donabedian A. Evaluating the quality of medical care. Milbank Quarterly 2005;83(4):691–729.
[49] Donabedian A. The quality of care: how can it be assessed? J Am Med Assoc 1988;260(12):1743–8.
[50] Berwick D, Fox D. "Evaluating the quality of medical care": Donabedian's classic years later. Milbank Quarterly 2016;94(2):237–41.
[51] Ferreira DC, Nunes A. Technical efficiency of portuguese public hospitals: a comparative analysis across the five regions of portugal. International Journal of Health Planning and Management 2018;34:411–22. doi:10.1002/hpm.2658.
[52] Otani K, Waterman B, Faulkner K, Boslaugh S, Dunagan C. 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 2010;55(1):25–38. doi:10.1097/00115514-201001000-00006.
[53] OECD. Health at a glance 2017: OECD indicators; 2017. doi:101787/health_glance-2017-en.
[54] Laurant M, Reeves D, Hermens R, Braspenning J, Grol R, Sibbald B. Substitution of doctors by nurses in primary care. Cochrane database of systematic reviews 2005. doi:10.1002/14651858.CD001271.pub2.
[55] Ferreira DC, Marques RC. Do quality and access to hospital services impact on their technical efficiency? Omega: The International Journal of Management Science 2019;86:218–36. doi:10.1016/j.omega.2018.07.010.
[56] Sullivan N, Schoelles K. Preventing in-facility pressure ulcers as a patient safety strategy: a systematic review. Ann Intern Med 2013;158:410–16. doi:10.7326/0003-4819-158-5-201303051-00008.
[57] Shonyela FS, Yang S, Liu B, Jiao J. Postoperative acute pulmonary embolism following pulmonary resections. Annals of thoracic and cardiovascular surgery 2015;21(5):409–17.
[58] Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. The third international consensus definitions for sepsis and septic shock (sepsis-3). JAMA 2016;315(8):801–10.
[59] Harvey M, Pierce M, Walter J, Chou Q, Diamond P, Epp A, Geoffrion R, Larochelle A, Maslow K, Neustaedter G. Obstetrical anal sphincter injuries (OASIS): prevention, recognition, and repair. Journal of Obstetrics and Gynaecology Canada 2015;37(12):1131–48.
[60] World Health Organization. WHO Statement on caesarean section rates. Reprod Health Matters 2015;23.
[61] Figueira JR, Greco S, Roy B, Słowiński R. An overview of ELECTREmethods and their recent extensions. Journal of Multi-Criteria Decision Analysis 2013;20(1–2):61–85. doi:10.1002/mcda.
[62] Figueira JR, Roy B. Determining the weights of criteria in the ELECTRE type methods with a revised Simos' procedure. Eur J Oper Res 2002;139:317–26. doi:10.1016/S0377-2217(01)00370-8.
[63] Almeida-Dias J. Multiple Criteria Decision Aiding for Sorting Problems: Concepts, Methodologies, and Applications. Université Paris IX Dalphine- Lisboa:IST; 2011.
[64] Ferreira DC, Marques RC, Nunes AM. Pay for performance in health care: a new best practice tariff-based tool using a log-linear piecewise frontier function and a dual–primal approach for unique solutions. Operational Research 2019:1–46.
[65] Ferreira DC, Nunes AM, Marques RC. Optimizing payments based on efficiency, quality, complexity, and heterogeneity: the case of hospital funding. International Transactions in Operational Research 2020;27(4):1930–61.
[66] Pereira MA, Machete IF, Ferreira DC, Marques RC. Using multi-criteria decision analysis to rank european health systems: the beveridgian financing case. Socioecon Plann Sci 2020:100913.
[67] Roy B. Classement et choix en présence de points de vue multiples. Revue française d'informatique et de recherche opérationnelle 1968;2(8):57–75.
[68] Figueira JR, Greco S, Roy B, Słowiński R. ELECTRE methods: main features and recent developments. In: Zopounidis C, Pardalos P, editors. Handbook of Multicriteria Analysis. Springer; 2010. p. 51–89.
[69] Figueira JR, Mousseau V, Roy B. ELECTRE methods. In: Multiple Criteria Decision Analysis: State of the Art Surveys, vol. 1. New York, USA: Springer Science+Business Media LLC; 2016. p. 155–85.
[70] Govindan K, Jepsen M. ELECTRE: a comprehensive literature review on methodologies and applications. Eur J Oper Res 2016;250(1):1–29. doi:10.1016/j.ejor.2015.07.019.
[71] Roy B. The outranking approach and the foundations of ELECTRE methods. In: Bana e Costa C, editor. Readings in multiple criteria decision aid. Springer; 1991. p. 155–83.
[72] Roy B. Méthodologie multicritère d'aide à la décision. Paris, France: Economica; 1985.
[73] Roy B. Multicriteria methodology for decision aiding. Dordrecht, The Netherlands: Kluwer Academic Publishers; 1996.
[74] Roy B, Bouyssou D. Aide multicritère à la décision: méthodes et cas.. Paris, France: Economica; 1993.
[75] Roy B, Figueira JR, Almeida-Dias J. Discriminating thresholds as a tool to cope with imperfect knowledge in multiple criteria decision aiding: theoretical results and practical issues. Omega: The International Journal of Management Science 2014;43:9–20. doi:10.1016/j.omega.2013.05.003.
[76] Greco S, Ehrgott M, Figueira JR. Multiple criteria decision analysis: state of the art surveys. Springer Science+Business Media LLC; 2016. doi:101007/978-1-4939-3094-4.