Are Portuguese hospitals suitably clustered according to their activity, quality, access, and expenses? An exploratory analysis on behalf of financial sustainability

Rodrigo Amado Pascoal da Graça
rodrigo.graca@tecnico.ulisboa.pt

Instituto Superior Técnico, Lisboa, Portugal

December 2020

Abstract

Financial sustainability has been one of the major challenges faced by healthcare systems. Portugal adopted strategies to promote efficiency in healthcare. One strategy implied the application of the lowest price for the consultations activity found in each of subsets of hospitals belonging to the National Health Service, thus promoting the providers' search for efficiency. This grouping was generated by clustering methodology to the data variables that could explain the costs of their activity. However, the adequacy of the established model concerning the present reality of the hospitals is unknown, as no adjustments have been implemented proceeding its initial application in 2013. Here we show that the current classes do not adequately reflect the current reality of the units. We found that four healthcare units do not integrate the most suitable category. Furthermore, from the 20 tested clustering methodologies, the maximum achieved silhouette is 0.36, indicating that the providers are the healthcare providers are not well grouped. In this context, doubts concerning its implementation for funding and benchmarking can be raised. Our results demonstrate that a revision of the clustering process is necessary. Additionally, this procedure should also encompass variables covering environmental, quality and access of the healthcare services dimensions, since for the present scenario the best results are obtained when these criteria are incorporated. These recommend steps are critical to ensure that the funding and benchmarking are fair, and to avoid the underfunding of the institutions. It is estimated that, due to inadequate grouping, providers receive yearly less €110 million than the due amount.
Keywords: Hospitals, Clustering, Funding, Benchmarking, Portuguese healthcare system

1. Introduction

The National Health Service (NHS) is the main provider of the healthcare services in Portugal. The NHS is funded mainly by taxes and it ensures universal care access to all residents in Portugal, regardless of social, economical, and legal status [17]. The NHS was responsible for 57.2% of the total health expenses in Portugal in 2018. This amounts to more than €10 000 million.[1] The hospitals receive the majority of the financial resources allocated to the NHS [10].

The hospital of the NHS are divided into groups. These are used for benchmarking and funding purposes. The current grouping has been utilized since 2013 [1]. This segregation of the units was performed by applying a clustering methodology to a data set that characterised the hospitals according to the installed capacity and production of healthcare services, variables which are able to explain the costs of the providers.[2]

The pertinence of an exploratory analysis conducted for grouping the Portuguese public hospitals is supported by the the time that has passed since the definition of the established groups. The presentation of the prevailing model occurred in October 2011. One of the arguments put forward for supporting the employment of the proposed hospital grouping was that the previous financing scheme did not reflect the contemporary reality of the institutions. Eight years have passed since the last funding model was instituted (in 2003) and significant transformations occurred in the contemplated hospitals [5]. Within this framework, it is relevant to assess the suitability of the current model against the present reality of the providers, because it has already been the same number of years since the conception of the current model. Additionally, the document that sets the incorporation of the grouping into the funding schemes for the publicly managed hospi-

[1] Instituto Nacional de Estatística - Estatísticas da Saúde : 2018. Lisboa : INE, 2020. Available at `https://www.ine.pt/xurl/pub/257793024` accessed on 18/11/2020

[2] `https://benchmarking-acss.min-saude.pt/BHEnquadramento/AbordagemMetodologica` acessed on 03/02/2020

tals acknowledges the existence of providers that were in frontier zones between two groups [1]. Wherefore slight variations along the years may have placed these units in different groups.

It is fundamental that the division in groups reflects the reality of the institutions, as a critical requirement for ensuring a fair funding and benchmarking.

This document yields two major goals. The first is to identify the parameters that define the clustering model that considering the information available replicates the currently established hospital groups. The second is to generate hospital groups following the same procedure of the previous step, considering solely the features used in the original analysis for both the most recent data and the data contemporary to the first presentation of the model. Each goal is cover by a specific experimental phase.

2. Background

2.1. Hospital clusters for funding and benchmarking in the Portuguese NHS

There is a multitude of healthcare units concerning the level of care, specialization degree, geographical location, population served, legal and managerial frameworks, etc. Ergo, the comparison of providers should be frame considering the similarity of the units [4]. The structure considered by NHS to conduct the analysis is the division of the hospitals depicted in Table 1.

The classes of units illustrated in Table 1 were presented in the 12^{th} [Portuguese] national of health economics conference in October 2011 [5]. The variables regarded in the process of construction of the groups are: medical working hours; nursing working hours; beds; offices for medical appointments; hospitalization episodes; urgency episodes; equivalent patients; complementary and diagnostic tests and therapies (CDTT); number of distinctive medical and surgical Diagnostic Related Groups (DRG); number of complex medical and surgical DRG; number of specialities offering consultations with a high differentiation level; classification of hospitals according to their urgency service; number of different types of CDTT with high differentiation level; beds in specialized units; classification regarding university teaching duties; ratio of resident physicians from the total medical doctors[16].

The methodology that ACSS followed was focus on the efficiency and production point of view. As visible in the set of 22 features listed in the precedent paragraph. The selection of these factors was guided by the effect of these on the structure of costs of the hospitals [1].

In Portugal, the NHS utilizes the hospital groups for benchmarking these units. The information regarding this process is available to any citizen trough an online platform at the responsibility of ACSS.[3] The efficiency comparison of the healthcare providers aims to

[3] `https://benchmarking-acss.min-saude.pt` accessed on 28/08/2020

Table 1: NHS Hospitals and the corresponding categorization

Name of the Health Unit	Abbreviation	Group
CH Médio Ave	CHMA	B
CH Póvoa do Varzim/Vila do Conde	CHPVC	B
HD Figueira da Foz	HDFF	B
H Santa Maria Maior	HSMM	B
CH Oeste	CHO	B
ULS Nordeste	ULSN	B
ULS Castelo Branco	ULSCB	B
ULS Guarda	ULSG	B
ULS Litoral Alentejano	ULSLA	B
H Vila Franca de Xira PPP	HVFX	B
CH Barreiro/Montijo	CHBM	C
H Senhora da Oliveira	HSOG	C
CHU Cova da Beira	CHUCB	C
CH Leiria	CHL	C
CH Setúbal	CHS	C
CH Baixo Vouga	CHBV	C
CH Entre Douro e Vouga	CHEDV	C
CH Médio Tejo	CHMT	C
HD Santarém	HDS	C
CH Tâmega e Sousa	CHTS	C
CH Cascais PPP	CHC	C
H Loures PPP	HL	C
ULS Alto Minho	ULSAL	C
ULS Matosinhos	ULSM	C
ULS Baixo Alentejo	ULSBA	C
ULS Norte Alentejo	ULSNA	C
CH VN Gaia / Espinho	CHVNG	D
H Espirito Santo	HESE	D
H Garcia da Orta	HGO	D
H Fernando da Fonseca	HPDFF	D
CH TM Alto Douro	CHTMAD	D
CH Tondela – Viseu	CHTV	D
CHU Algarve	CHUA	D
H Braga	HB	D
CH Lisboa Ocidental	CHLO	E
CHU de Coimbra	CHUC	E
CHU Lisboa Central	CHULC	E
CHU Lisboa Norte	CHULN	E
CHU do Porto	CHUP	E
CHU de São João	CHUSJ	E
IPO Porto	IPOP	F
IPO Lisboa	IPOL	F
IPO Coimbra	IPOC	F

H - Hospital; CH - Hospital Center; CHU - University Hospital Center; IPO - Portuguese Oncology Institute;

promote the economical and financial performance of hospitals while promoting better access and quality of care.[4]

The categorization of hospitals is also used in the funding schemes of these units. As the calculation of the monetary value to be transferred to each public hospital takes into account the hospital groups that unit belongs to, particularly in the reimbursement of consultations.

The Contract program is the central document concerning the payment of the activity of hospitals EPE since 2002. The document stipulates the volume and type of the healthcare services that the provider will produce and the correspondent financial payment for those services. Each contract has a scope of a year

[4] `https://benchmarking-acss.min-saude.pt/BHEnquadramento/Objetivos` accessed on 28/08/2020

[17].[5]

The information conveyed here regarding the funding schemes is according to the funding terms and schemes for the year of 2020 terms that defined the funding schemes for the of 2020 in [2]. The Contract Program (CP) is composed of three major three elements: delivery of care, incentives and penalties. The grouping depicted in Table 1 directly impacts two components of CP: the relative (benchmarking) incentive and the payment of the consultations.

The incentive above mention is based on the benchmarking approach. This element began to be included in the contract programs of 2017 [6]. It consists on the computation of the Index of Compared Performance, determined according to the achieved results in the set of indicators that evaluate access, quality and efficiency, resulting in an ordered list of providers per hospital category. The hospitals' classes that are utilized in this incentive are the grouping depicted in Table 1.

This incentive is awarded to the organizations that are placed first. The cost of this incentive is supported by all remaining units of the respective group. Therefore this incentive is a penalty for the institutions that have to finance the reward. The values of the award are not disclosed in the terms of contractualization.

The other funding component directly influenced by the hospital groups is medical consultations, as this line of production has its unit price is defined the class of each provider. The practiced unitary price was defined by the minimum unit cost of this service that is found when evaluating all the providers of hospital groups [1] [9].

2.2. Computational background

Clustering is an unsupervised machine learning technique, which aggregates data instances. The resulted groups are designated as clusters. Clustering techniques label objects according to the assessment of the similarity of the observations. Thus, the groups comprise objects that are similar between them and are dissimilar when compared to data items from other clusters [13].

Clustering can be explained as a process that has as input a set of observations unlabeled that produces as output the initial set of observations labelled and organized in groups. The process comprises four steps: dimensionality reduction; selection of the dissimilarity measure; election of the clustering technique and analysis and validation of the results.

Each stage influences the outcome, in such a way that for the same input, a multitude of different set clusters can be achieved. The challenge of clustering is tuning the parameters so that the output reflects the actual structure and relationship between the observations.

This present subsection summarises the state of the art concerning the clustering of hospitals. The employment of this machine learning tool to generate subsets of healthcare providers is driven by multiple motives. Despite the broad range of reasons the scientific literature covering this topic is scare. Albeit this fact, hereby are outlined the most relevant information of the most recognized articles in the theme, eight in total.

Firstly, the study of Hariyanti et al. [12] involves the application of K-means and hierarchical algorithms (with single and complete linkage criteria) to Indonesian hospitals with the purpose of classifying these healthcare units according to the established law.

Secondly, the work of Belfin et al. [3] details the experience to identify the areas where the scarcity of healthcare service is most pressing in India. The study is conducted to support the decision for defining the location of a novel healthcare units in India. Hierarchical clustering with complete linkage criterion is the metric used.

Thirdly, hierarchical clustering with Ward's method was the technique applied to data regarding three hospitals of Singapore with the aim of aggregating hospital medical specialities based on their utilization by patients. This experience is reported in the study of You et al. [19].

Fourthly, the most prevailing reason for performing clustering hospitals is to improve benchmarking results. Data Envelopment Analysis (DEA) is a powerful methodology to generate information on benchmarking. This approach is a non-parametric technique that evaluates and identifies the most efficient entities regarding different dimensions adjusted to each case. There exist constrictions which limit the application and outcomes of the method. A couple of the most significant are: the declining in performance with the increment of variables, and sensitivity of results regarding the input and output features covered in the analysis [6]. There are many approaches possible to overcome these drawbacks. Here are highlighted the approaches of machine learning techniques of unsupervised learning (clustering) [11], the optimizations of the weights in the DEA process [7], and the aggregation of the healthcare units into homogeneous groups considering the environmental factors and volume of outpatients [15].

Fifthly, the aforementioned reported experiences involve the application of a narrow range of techniques, predominated by hierarchical and K-means methods. However, the publication of Byrne et al. [4] is interesting in this context, since it proposes a novel approach for grouping hospital units. The proposed method is based upon Nearest-Neighbour (NN) algorithm, offer-

[5] Decree-Law no.18/2017 `https://data.dre.pt/eli/dec-lei/18/2017/02/10/p/dre/pt/html` accessed on 30/08/2020

[6] Terms of contractualization for 2017 `http://www.acss.min-saude.pt/2016/07/22/metodologia-de-contratualizacao-2/` accessed on 25/09/2020

ing advantages over the more traditional methods.

Taking into account all the supra-mentioned articles is possible to conclude that for clustering hospitals the most applied methods are hierarchical clustering (5/8 of papers) and the K-means (4/8 of studies). Additionally, K-NN clustering method and multiple correlation model were implemented each of them once for clustering this type of health care units.

3. Methodology

Two main goals are proposed to be achieve. First, to define the parameters of the clustering model that led to the currently established hospital classes. Second, to generate hospital groups following the same procedure of the previous step, considering solely the features used in the original analysis for both the most recent data and the data contemporary to the first presentation of the model. The third consists of repeating the preceding process but, considering the dimensions of access, quality and environment as well. A experimental framework that comprises threes phases was designed to meet these objectives.

This project focus on Portuguese general public hospitals that currently have a juridical status of EPE. Thus, being excluded from the study both specialized hospitals and general hospitals with a different juridical and administration frame, such as SPA and LHU. These exclusions are implemented due to the meaningful differences of the providers evidence as consequence of the disparity of the clinical activity and divergence regarding the governance of the units defined by the legal-administrative framework, respectively. The 28 hospitals covered in this work are those belonging to the class B, C, D and E depicted in Table 1, with the exception of the ULS and hospitals PPP. H Braga EPE is not considered as for the period analysed in this study it was a hospital PPP.

The main structure of the study is set upon the two scenarios. The first regards the data considered in the original clustering process that lead to the present hospitals categories, which have been implemented since 2013 in the funding schemes. The guidelines for the production and financial agreement of that year were published on the November 2012 [1]. Although, the creation of the hospital groups was presented in 2011 [5]. The information available leads to some uncertainty concerning the sources of the data. Thus a supposition was made that all the data used was the most recently available at the time contemporary to the creation of the clusters, so 2010 was the year that matches these expectations. This scenario was labelled as CTC, with the initials from the expression "contemporary to [the] creation". The other time frame used was the present scenario. Like the name suggests it corresponds to the actual time-frame. However, due to the delay that exists between the publication of reports and data regarding a certain period, the majority of the data in this time-frame corresponds to 2019.

The set of features utilized in the original process of clustering are listed in the background. The first phase of the work aims replicating this process. However, the existence of restrains conditioned the pursue of this goal. Since it was precluded the access to some of the variables included in the original protocol. Seven proxies variables were used to overcome this issue. Hospital categorization defined in Decree no.82/2014 is used to replace "Range of medical appointments for differentiated care". The range of DRG, the range of complex DRG and the ratio of complex DRG are replaced by the four Case Mix Indexes (CMI) available covering both the ambulatory and hospitalization of the surgical and medical activity. The number of nurses substitutes the value of nursing hours. The number of medical appointments replaces the information on the cardinality of external consultation cabinets, the ratio of medical appointments for differentiated care. Finally, Internal CDTT is the proxy variable utilized to substitute the number of equipment and differential equipment features.

The experience is designed to consider the actual hospital network of the NHS and to implement the same criteria in all the scenarios covered. This guideline is respected with the expection of the "hospital categorization" variable. The initial intention was to consider the classification of the hospitals from which actual referral network were built, established in *Carta Hospital* published in 2014.[7]. Although, it is not reasonable to use this criterion in CTC scenario as at the time this classification did not exist yet. Therefore, it was chosen to use instead the classification of the public hospitals contemporary at that moment: classification of the permanent care service of healthcare units, defined in the dispatch no. 5414/2008 [8]. This information coincides with the already incorporated feature "urgency typology". Consequently, the information that is capture in both scenarios is the same, but the number of features differs: the CTC scenario is tested with 18 variables while the Present scenario uses 19. As the urgency typology is only used one time, and it encompass the information of the categorization of the hospitals in the CTC scenario too.

For this task, several combinations were used. The choice of algorithms to be tested was based on the information available by the ACSS, regarding the procedure applied, and also on the state of art for clustering hospitals. The table 2 illustrates the different combination of approaches used. Two important notes. One, as mentioned previously, there was an innovative application of K-NN clustering to hospital grouping reported

[7] Decree no. 82/2014 https://data.dre.pt/eli/port/82/2014/04/10/p/dre/pt/html accessed on 08/03/2020

[8] Dispatch no. 5414/2008 https://dre.pt/web/guest/pesquisa/-/search/3378909/details/normal?q=5414%2F2008 accessed on 08/03/2020

in [4]. Even though it is an interesting approach, DBSCAN shares the advantages of this and also identifies outliers among samples. This latter aspect, contrasts with all the other algorithms used. Two, multiple correlation algorithm is a clustering technique designed for dealing with data sets with high dimensionality [14]. That is not the case of the data studied in this work, therefore this method is not included.

Table 2: List of all the different clustering methods used for grouping the sample. Each combination is composed of a clustering algorithm, a similarity metric, and for hierarchical clustering methods it is also defined a linkage criterion

ID	Algorithm	Metric	Linkage Criteria
1	K-means	Man.	-
2		Euc.	
3		Euc. + Ham.	
4		Man. + Ham.	
5	Hierarchical	Man.	Complete
6		Euc.	
7		Euc. + Ham.	
8		Man. + Ham.	
9		Man.	Ward
10		Euc.	
11		Euc. + Ham.	
12		Man. + Ham.	
13		Man.	Single
14		Euc.	
15		Euc. + Ham.	
16		Man. + Ham.	
17	DBSCAN	Man.	-
18		Euc.	
19		Euc. + Ham.	
20		Man. + Ham.	

Man. - Manhattan; Euc. - Euclidean; Euc. + Ham. - Euclidean for numerical features and Hamming for categorical; Man. + Ham. - Manhattan for numerical features and Hamming for categorical

Experiences were made with three major types of algorithms: K-means, hierarchical clustering and DBSCAN. The hierarchical clustering algorithms were used with three different linkage criteria: Ward's, single and complete linkage. For all of these different approaches four distances metrics were used, forming the 20 combinations that are expressed in table 2.

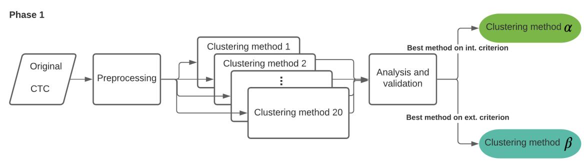

Figure 1: Scheme of the phase 1 of the proposed methodology

Phase 2 focus on the study of grouping of hospitals concerning the original features and its evolution along the time dimension. It computes the clusters that result from the application of the two methods identified in Phase 1. α correspond to the method that best aggregates the providers, and β the method that best replicates the established groups. The selection of the mentioned techniques regards the data set that describes the units in the CTC scenario with the original variables.

Figure 2 depicts the framework for phase 2 of the work, which is detailed in Subsection ??. The clustering method α and β are the techniques that are pick from the set of tested approaches in Phase, as depicted in Figure 1.

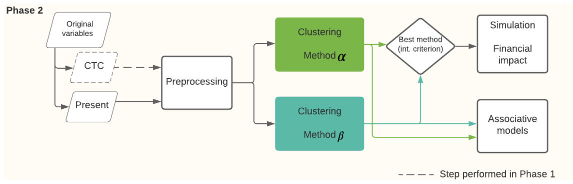

Figure 2: Scheme of the phase 2 of the proposed methodology

4. Results

From the analysis of within-cluster sum of squares (WCSS) according to the elbow method, five is identified as the optimal number of clusters. Therefore, the analysis will cover this number of clusters, as it will include the number of the established grouping: four.

Table 3 and Table 4 reveal the results of the different clustering combination utilized to group the target hospitals, for k = 4 and k = 5, respectively. These charts express the number of clusters considered to be optimal for the data objects and also show the corresponding values for each combinations on both the internal and external validation criteria. Furthermore, these tables include the assessment of the statistical significance regarding the obtained clusters for each of the methods tested.

As stated above, K-means is a non-deterministic algorithm. Therefore, a different approach is ought to be followed for obtaining and analysing the outputs of this method. The aggregation of the K-means algorithms considered in Table 3 and Table 4 are the best evaluation on the external criterion. For each combination the model was ran 30 times.

Table 3: Results from the application of the different tested combination to the data with the original variables, with the number of clusters = 4

ID	Adjusted Rand index	Silhouette	Statistical significance*
1	0.414105	0.226048	Y
2	**0.500339**	0.189837	Y
3	0.327473	0.286912	Y
4	0.389907	0.160098	Y
5	0.294630	0.285205	Y
6	0.155445	0.318714	Y
7	0.296089	0.288527	Y
8	0.255639	0.228400	Y
9	**0.495659**	0.212423	Y
10	0.199359	**0.359867**	Y
11	0.363173	0.271526	Y
12	0.323193	0.226890	Y
13	0.224670	0.225269	Y
14	0.175908	0.230770	Y
15	0.149091	0.169891	Y
16	0.121951	0.225910	Y

Groups P-values 0.01; Y - Yes;

Table 4: Results from the application of the different tested combination to the data with the original variables, with the number of clusters = 5

ID	Adjusted Rand index	Silhouette	Statistical significance*
1	0.432979	0.185248	Y
2	**0.571429**	0.134417	Y
3	0.368421	0.190092	Y
4	0.353705	0.190593	Y
5	0.286344	0.265503	Y
6	0.213758	**0.346319**	Y
7	0.293233	0.280614	Y
8	0.276382	0.229677	Y
9	0.455206	0.219007	Y
10	0.213758	**0.346319**	Y
11	0.363533	0.271466	Y
12	0.322908	0.211125	Y
13	0.197044	0.177740	Y
14	0.147783	0.212554	Y
15	0.197044	0.119127	Y
16	0.169086	0.193322	Y

Groups P-values < 0.01; Y - Yes;

Table 5: Results from the application of the DBSCAN algorithm to the data with the original variables

ID	Number of clusters	Adjusted Rand index	Silhouette	Statistical significance*
17	19.0	0.296142	0.069851	Y
18	19.0	0.296142	0.069851	Y
19	19.0	0.296142	0.069851	Y
20	19.0	0.296142	0.069851	Y

* Groups P-values < 0.01. Y - Yes;

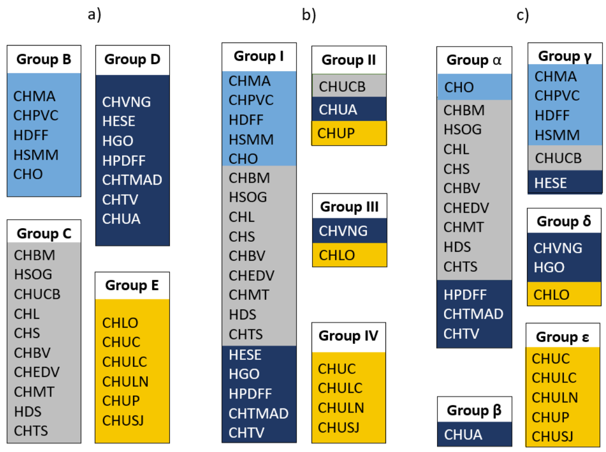

Figure 3: Comparison of the novel hospitals clusters obtained by different methods with the data set that comprises the original features in the CTC scenario. a) established groups; b) hierarchical clustering with Ward's criterion (k=4); c) K-means (k=5)

The results generated by the 20 tested methodologies are presented in the Table 3, Table 4 and Table 5. By observing the values on external and internal criteria depicted in these tables, a pair of methods that produce the clusters with the best performance with respect to the the internal and external criterion can be identified. These are *ID 10* – Hierarchical with the Ward's linkage criterion – for k=4 (having a silhouette score of 0.359867) and *ID2* – K-means with Euclidean metric – for k=5 (with a value of 0.571429 for the adjusted Rand index), respectively. Hence this couple are the methodologies utilized in the clustering tasks that are subsequently presented in this document, corresponding to the method α and β in both Figure 1 and Figure 2.

The output of the hierarchical clustering contemplates a dendrogram. The distance separating the clusters when $K\bar{4}$ is very narrow. This implies that this number of groups is far from being the optimal for the case studied. Hence, it was considered the number of groups to be five. This was the choice considering that this corresponded to the alternative value identified previously as the number for the optimal number of cluster.

This subsection contemplates the aggregation results of the 28 hospitals units produced in Phase 2 of the experimental framework. Figure 4 show the current categorization of the hospitals and the proposed grouping that results from the method α (hierarchical algorithm) and method β (K-means), refereed in the previous subsection. Table 6 depicts the validation assessment of the clustering results concerning the Present scenario of Phase 2. It also cover the statistical significance of these results.

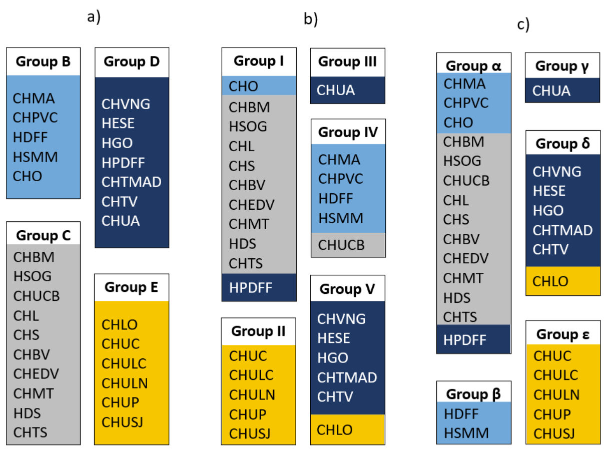

Figure 4: Comparison of the novel hospitals clusters obtained by different methods with the data set that comprises the original features in the Present scenario. a) established groups; b) hierarchical clustering with Ward's criterion (k=4); c) K-means (k=5)

Table 6: Validation and statistical significance of the clusters produced in Present scenario of Phase 2

Clustering method	Adjusted Rand index	Silhouette	Statistical significance*
K-means	0.5517886	0.2580752	Y
Hierarchical clustering	0.5859564	0.2534862	Y

* Groups P-values < 0.01. Y - Yes;

The effect of taken into account the novel groupings on reimbursement of medical appointment the hospital budget is -14 623 927 € (-10.19% regarding the attributed value) and +110 794 012 € (+20.26%) regarding CTC and Present scenario, respectively. The groups considered were the ones with the highest performance regarding the internal validation criterion.

5. Discussion

The initial challenge was the replication of the model that generates the NHS hospital groups. The results that more closely get to those are obtained with the

K-means method with k = 5 (for the CTC secnario), as depicted in Figure 3. In a more quantitative view, the units from the category B, C and E are placed together missing only one unit, which corresponds to 80%, 90% and 83.3 % ratio between the hospitals group in the cluster with others members of their groups of the ACSS grouping over the total of units belonging to the group. Whereas, the providers of the established group D only achieve a value of 43% in this ratio. It can be speculated that this group was less homogeneous than all the defined classes. Thus, due to the slight disparity in the distances caused by the use of proxy variables this produced enough changes in the distances between the units, which culminated in the fragmentation of the group in the k-means approach. This hypothesis is supported by the fact that for all the other situations analysed 2 of the hospitals that appear here in a different group from the core group, they appear clustered in the core groups (raising the ratio to 71%).

As a final note, the fact that certain hospitals are not aggregated according to the ACSS classes should not be interpret as the model being wrong. These divergences can be due to differences concerning the clustering algorithm, similarity criterion and/or variables [18]. The first and second elements can be discarded as responsible, considering the focus of the first phase being on the identification of the methodology that best fits the case study. Even more, when it is taken into account that the pool of tested methods was chosen based upon the state of the art. Thus, the discrepancies are most likely caused by the variables used, as a consequence of the divergences on the features-space. Since there were variables that are represented in our model by proxies, which creates similar variables-space, however not equal.

In light of all the above it is possible to conclude that the main structure of the results of ACSS is captured. Consequently, the model conceived in this work is adequate for the study of the influence of time on the hospital grouping. The procedure that generates the closest results of ACSS is the K-means algorithm with Euclidean distances for k = 5.

5.1. Best division of the samples according to the nature of the data itself

The method that generates the best separation of the hospital units is the hierarchical clustering algorithm with Ward's linkage criterion. The component b) of Figure 3 depicts the grouping that results from applying this approach. Internal criteria assesses the quality of the division of the data instances. Observing and comparing these values in Table 3, Table 4 and Table 5 the *ID 10* - Hierarchical clustering with the Ward's linkage criterion - for k = 4 (having a silhouette score of 0.359867) corresponds to the methodology with the best score, which implies that this is the method that generates the output that best groups the hospitals.

It is paramount to take into account that the method that was used in the original procedure coincides with the best method for grouping the healthcare providers with the experiment data set. Moreover, the number of clusters obtained in the best clustering method is 4, which reinforces the hypothesis that the replication of the process was successful, since this is the identical number of classes that are defined in the operation being reproduced.

The values of silhouette for all the 20 tested methods assume overall low numbers, see Table 3, Table 4 and Table 5. The maximum number on this criterion is 0.359867. This is a number that already implies that clusters have some meaning though not very pronounced, indicating that the healthcare providers under analysis are very distinctive. So any division of the hospitals is not expected to achieve a silhouette value much higher than the maximum number of this study. This raises very relevant questions regarding the adequacy of using the hospital groups obtained by the process that was replicated here, conditioned of course by the information available (both regarding data aspects as methodological issues). These results supports the perspective reported in the [8] that the hospitals are in fact very heterogeneous and this creates challenges for the implementation of fair funding.

5.2. Phase 2: Impact of time in the replicated model

Figure 4 presents the current grouping, the generated by hierarchical clustering with Ward's linkage criterion and K-means algorithm. The analysis of these results is conducted one ACSS group at a time.

Considering solely the established category B, the results are globally consistent in both scenarios and methodologies. These providers present strong affinity with the units of group C. From all the hospitals that constitute this established class B, CHO is the one that would be more likely to being clustered in another group as it is placed in all the methods-scenarios tested in the class in which the most predominant ACSS group is C.

Regarding the subset C, it is visible that with time it becomes better defined. This is evidenced by comparing the constitution of the classes in which these hospitals are put in. In a more quantitative approach this is evidenced by the ratio between the hospitals identified as the original C class over the total of hospitals presented in the same novel(s) class(es). In CTC scenario the value is 0.833, which compares to the 0.933 of the Present scenario. Both these results are the arithmetic average of the pair of methods used.

The providers of D class are the most disperse in the novel grouping (outputs from both the K-means and hierarchical) in the scenario CTC, presented in Figure 3. CHUA is a hospital that in all the scenarios and algorithms is placed into either singles groups or in subsets in which it is the only representative of the established class. This suggests that CHUA should be placed in a different group. HPDFF also appears in the Present Scenario outside the corresponding peers of the ACSS category, this is reported in both methods. It is placed in clusters in which the other hospitals belonging to the current C group. This shows that HPDFF would be better grouped in a subset that contains the C elements and not the D.

The elements of category E exhibit the same pattern in the two temporal scenarios under analysis. All these units aggregate together in a unique class without any other provider, with the exception of CHLO. This unit is placed in the groups that contain the majority of the D class hospitals. Thus, the experimental results suggest that CHLO is not integrated in the most suitable category.

As expected, there was a significant proportion of hospitals aggregated in a very similar way between the scenarios. However, there are divergences in the results between the CTC and the Present scenario. Hospitals such as the CHUA, CHLO, HPDFF and CHO exhibit a pattern that points to changes in the currently grouping performed by ACSS years ago. CHUA and CHO are hospitals that suffered structural changes due to fusions of hospitals and hospital centers between the CTC and Present scenario. Therefore, it seems fair to affirm that fusion or division of hospitals/hospital centers of the grouping should be revisited.

5.3. Financial impact

There were constraints regarding the lack of access to the variables utilized in the original process and that were employed in definition of the values for the reimbursement of the medical consultations, the procedure is declared in the [1]. So an approximation model had to be conceived and implemented.

It is verified that the difference between the value that should be transferred to the hospitals considering the novel groups as opposed to the current group increases with time. In the year of 2013 these differences are estimated to be of 10.19% and it enlarges to 20.26% in 2017. Furthermore, in 2013 the new groups would produce savings, which is inverted in the 2017 scenario, indicating that the effect of the groups in the funding increment with time, which highlights the relevance of reviewing the groups.

The calculations are settled upon the assumption that the ratio of the unitary costs of medical consultations over the standard patient cost is invariant in time. Consequently, it is speculated that the gap between the estimated value that should be paid considering the novel grouping is over 100 million €. This amount should be additionally paid to the providers, solely for the year of 2017. This significant divergent, of 20% of the actual funding for the same activity, means there was an increment of the costs of the medical consultations over the total costs of standard patient. This affirmation only applies to the providers that presented the lowest cost per group. However, if the price practiced was adequate for the activity that is reimbursed, such a high variance (20%) would not be expected.

This permits the speculation that groups utilized for funding are not adequate, reinforcing the need for revisiting them. This position is in accordance with the idea that the root of the problem is located in considerations that were made in the definitions of the prices in 2013. As it is stated certain hospitals were identified to be in a frontier zone. Thus, it is possible that these could be placed in a class, in which the remaining units had higher unit costs. Consequently, the price that was identified is the lowest of the group, but if the unit does not fits properly the group, then it can artificially cause the dropping of the values.

5.4. Principal implications in the management and funding

Hereby are presented the major impacts of this work for the funding and administration of the hospital units covered by this study.

First, a revision should be made to the grouping of the public hospitals to ensure that both the funding and benchmarking of these providers are conducted based upon a model that reflects the current reality of the institutions. This process can also be taken into advantage to improve the fairness of the categorization of healthcare providers, as more data is gathered and available regarding the providers than in the moment of the generation of the first implementation of the established division. Ergo, a broader and more complete vision of the units can be achieved.

Second, the dimension that the incorporate variables cover in the clustering process also should be reconsidered. Efficiency is a very relevant topic that should be continue to be including in the process, as financial sustainability is fundamental for any HS. Although, it can not receive the exclusivity concerning the establishment of the groups for these units, since the pursue of efficiency should not undermine the quality and access of the healthcare services that are defined as acceptable. So, features of the three key-factors should be included in the aggregation procedure: efficiency, quality and access. Moreover, environmental factors should also be enclosed in the process due to their impact not only on the costs of the activity impact, but also social and economical factors influence the activity of the healthcare services, and vice-versa. Furthermore, this decision is supported by the fact that the studies concludes that for the present scenario the general public hospitals EPE are better aggregated when all these dimensions are encompass in the model.

It should also be reevaluated the application of the hospital groups for the important tasks that are currently utilized: benchmarking and funding. The results produced in this work point to the strong distinctiveness of the healthcare units studied. Therefore, the considering the grouping for these tasks should be done with caution. Even more, when the determination of the price for an activity as fundamental for the hospitals as the external medical appointments are based on it. As underneath there is the premise the units belonging to the placed group are so similar that the one that has the lowest cost in providing this service is the most efficient. This is not the most adequate approach for two aspects: One, there can be factors exogenous to the hospital that influence the lower costs: factors of income, more healthy habits, use of healthcare services in private providers. Two, the quality of the service was not consider in the determination of the price, neither was the access.

6. Conclusions

The current grouping of the Portuguese hospitals of NHS should be reviewed given that it diverges from the categorization that best describes the units covered. It is estimated that this inadequacy is causing an overall underfunding of the general public hospitals EPE of about €110 millions/year. The reexamination is necessary, as the established categorization does not adequately translate the reality of the providers, due to the transformations that occurred in the hospital organizations after its publication in 2011.

The main challenges experienced through this work are with respect to the time scenarios and to the variables covered. The identified future work directions aim at overcoming the identified limitations. First, more time scenarios should be covered in the analysis. For instance, all the years between the formation of the hospital classes and the present year could be included in the study. This would permit to identify more precisely when did the change in the grouping occurred and consequently to comprehend the transformation in the healthcare providers that led to this outcome.

Regarding the variables, two major aspects can be addressed. One, categorical ordinal variables such as the "types of urgency" were handled as non-ordinal, resulting in information loss. To overcome this, numeric encoding can be identified to specify the distances between the different categories. Second, efforts can be made to ensure that the information that is gathered by the MH and ACSS is made available for the study. To ensure that the results, and inherently the derived conclusions are more reliable and useful for the polices and management decision makers.

7. Acknowledgement

This document was written and made publically available as an institutional academic requirement and as a part of the evaluation of the MSc thesis in Biomedical Engineering of the author at Instituto Superior Técnico. The work described herein was performed at the INESC-ID, during the period of January-December 2020, under the supervision of Prof. Diogo Ferreira, Prof. Rui Henriques and Prof. Alexandre Nunes from the ISCP.

References

[1] ACSS. *Termos de Referência para contratualização de cuidados de saúde no SNS para 2013*. Administração Central do Sistema de Saúde (ACSS), 2012.

[2] ACSS. *Termos de Referência para contratualização de cuidados de saúde no SNS para 2020*. Administração Central do Sistema de Saúde (ACSS), 2019.

[3] R. Belfin, E. G. M. Kanaga, and P. Bródka. Regional clustering of indian hospitals for better health. In *2018 International Conference on Circuits and Systems in Digital Enterprise Technology (ICCSDET)*, pages 1–5. IEEE, 2018.

[4] M. M. Byrne, C. N. Daw, H. A. Nelson, T. H. Urech, K. Pietz, and L. A. Petersen. Method to develop health care peer groups for quality and financial comparisons across hospitals. *Health services research*, 44(2p1):577–592, 2009.

[5] F. Candoso, A. Lourenço, J. Barata, N. Amaro, T. Nolasco, R. Lavado, and A. S. Ferreira. Proposta de evolução do modelo de financiamento dos hospitais do Serviço Nacional de Saúde: criação de novos grupos de financiamento. In *12ª Conferência Nacional de Economia da Saúde*, Lisbon, 2011.

[6] V. J. M. Cantor and K. L. Poh. Integrated analysis of healthcare efficiency: a systematic review. *Journal of medical systems*, 42(1):1–23, 2018.

[7] S. Cinaroglu. Integrated k-means clustering with data envelopment analysis of public hospital efficiency. *Health Care Management Science*, pages 1–14, 2019.

[8] D. C. Ferreira and R. C. Marques. Should inpatients be adjusted by their complexity and severity for efficiency assessment? evidence from portugal. *Health care management science*, 19(1):43–57, 2016.

[9] D. C. Ferreira, R. C. Marques, and A. M. Nunes. 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.

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

[11] A. Flokou, N. Kontodimopoulos, and D. Niakas. Employing post-dea cross-evaluation and cluster analysis in a sample of greek nhs hospitals. *Journal of medical systems*, 35(5):1001–1014, 2011.

[12] L. S. Hariyanti, K. R. Sungkono, and R. Sarno. Clustering methods based on indicator process model to identify indonesian class hospital. In *2019 International Seminar on Application for Technology of Information and Communication (iSemantic)*, pages 196–201. IEEE, 2019.

[13] D. Karaboga and C. Ozturk. A novel clustering approach: Artificial bee colony (abc) algorithm. *Applied soft computing*, 11(1):652–657, 2011.

[14] H.-P. Kriegel, P. Kröger, and A. Zimek. Clustering high-dimensional data: A survey on subspace clustering, pattern-based clustering, and correlation clustering. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 3(1):1–58, 2009.

[15] H. Najadat, A. Alaiad, S. A. Alasal, G. A. Mrayyan, and I. Alsmadi. Integration of data envelopment analysis and clustering methods. *Journal of Information & Knowledge Management*, 19(01):2040006, 2020.

[16] A. Nunes. *Eficiência e produtividade nas unidades hospitalares em Portugal: análise comparada dos regimes PPP e E.P.E.* PhD thesis, University of Lisbon, 2020.

[17] G. F. I. Simões, J; Augusto, C. Hernández-Quevedo, W. H. O. R. O. for Europe, E. O. on Health Systems, and Policies. *Portugal: health system review.* Health Systems in Transition; vol. 19, n. 2. World Health Organization. Regional Office for Europe, 2017.

[18] B. Yang, X. Fu, N. D. Sidiropoulos, and M. Hong. Towards k-means-friendly spaces: Simultaneous deep learning and clustering. In *international conference on machine learning*, pages 3861–3870. PMLR, 2017.

[19] X. You, B. H. Heng, and K. L. Teow. Application of sparse matrix clustering with convex-adjusted dissimilarity matrix in an ambulatory hospital specialist service. In *2014 IEEE Symposium on Computational Intelligence in Big Data (CIBD)*, pages 1–8. IEEE, 2014.