J Med Syst (2016) 40: 103
DOI 10.1007/s10916-016-0458-9

SYSTEMS-LEVEL QUALITY IMPROVEMENT

A New Strategy to Evaluate Technical Efficiency in Hospitals Using Homogeneous Groups of Casemix

How to Evaluate When There is Not DRGs?

Manuel Villalobos-Cid[1] · Max Chacón[1] · Pedro Zitko[2] · Mario Inostroza-Ponta[1]

Received: 7 October 2015 / Accepted: 29 January 2016 / Published online: 15 February 2016

Abstract The public health system has restricted economic resources. Because of that, it is necessary to know how the resources are being used and if they are properly distributed. Several works have applied classical approaches based in Data Envelopment Analysis (DEA) and Stochastic Frontier Analysis (SFA) for this purpose. However, if we have hospitals with different casemix, this is not the best approach. In order to avoid biases in the comparisons, other works have recommended the use of hospital production data corrected by the weights from Diagnosis Related Groups (DRGs), to adjust the casemix of hospitals. However, not all countries have this tool fully implemented, which limits the efficiency evaluation. This paper proposes a new approach for evaluating the efficiency of hospitals. It uses a graph-based clustering algorithm to find groups of hospitals that have similar production profiles. Then, DEA is used to evaluate the technical efficiency of each group. The proposed approach is tested using the production data from 2014 of 193 Chilean public hospitals. The results allowed to identify different performance profiles of each group, that differs from other studies that employs data from partially implemented DRGs. Our results are able to deliver a better description of the resource management of the different groups of hospitals. We have created a website with the results (bioinformatic.diinf.usach.cl/publichealth). Data can be requested to the authors.

Keywords Data Envelopment Analysis · Hospital · Technical efficiency · Casemix · Diagnosis-related groups

This article is part of the Topical Collection on *Systems-Level Quality Improvement*.

Electronic supplementary material The online version of this article (doi:10.1007/s10916-016-0458-9) contains supplementary material, which is available to authorized users.

✉ Mario Inostroza-Ponta
mario.inostroza@usach.cl

1 Departamento de Ingeniería Informática, Facultad de Ingeniería, Universidad de Santiago de Chile, Santiago, Chile

2 Unidad de Estudios Asistenciales, Hospital Barros Luco Trudeau, Facultad de Medicina, Universidad Diego Portales, Santiago, Chile

Introduction

Given the fact that hospitals have limited resources, it is necessary to know the efficiency in the use of their resources. In order to perform this task, several works have used well-known non-parametric and parametric methodologies like Data Envelopment Analysis (DEA) and Stochastic Frontier Analysis (SFA) [1–6]. To compare technical efficiency among hospitals with different casemix, studies have used production data of hospitals adjusted by weights from Diagnosis Related Groups (DRGs) [8–13]. Recently, Rezaee and Karimdadi [7] propose a new methodology to evaluate hospitals efficiency based in DEA multi-group analysis. Authors also performed an updated literature review of hospitals evaluation methods that can be found in their work and references therein.

Since its development in the 70s, DRGs have become one of the most used hospital's system classification in the world, supporting prospective payment systems in developed countries such as United States, Germany and Australia [14]. By grouping patients with same features, DRGs allow to characterise the casemix of hospitals, perform benchmarking to manage costs and reduce the length of stay of patients. However, in developing countries, DRGs may have not being implemented or with a low coverage, with just a small number of hospitals using this system in a very early stage if implementation (low proportion of patients registered, including reliability constraints due to lack of informatics tools to record medical history of patient, overcoding and coding errors [15–17]). In other cases, DRGs only include partial data from inpatients, excluding information about ambulatory activities, emergency care, morbidity or outpatient surgery. Moreover, studies have presented differential considerations about hospitals located in remote geographical areas, or specialised hospitals, such as mental or geriatric centres [18–20]. All of these conditions, in addition to differences in versions of the DRGs, avoid a fair comparison between hospitals.

In particular in Chile, the number of hospitals that have implemented DRGs has rised in public hospitals (currently 61), and has been used for performance comparison. However, this number represents only 32 % of coverage of all public hospitals in the country, making hard to perform a comprehensive country-level comparison.

This study presents a new strategy for benchmarking among hospitals, grouping hospitals with same casemix and measuring technical efficiency regardless of the DRGs weights. This strategy allows the evaluation of the performance among hospitals with equal casemix when the DRGs are not implemented or they have an incomplete register.

In order to evaluate our proposal, financial and production data from Chilean public health facilities were used to evaluate the performance among 193 public hospitals using available public data.

The rest of the paper is organised as follows: "Materials and methods" shows the proposal description including data selection and preparation and the algorithms used. Next, "Results" shows the computational results of applying the proposal on a dataset of the Chilean public health system, highlighting the main findings and characterising the clusters found. The last section presents the discussion of the results and the main conclusions reached in this work.

Materials and methods

Scenario of application

We applied the proposed methodology to the Chilean public health system (Proposed approach). We use data of 193 public hospitals related with production and finance behaviour.

The production data was collected from the website of the National Health Statistics Department (www.deis.cl, Departamento de Estadísticas e Información, DEIS) of the Chilean Minister of Health (Ministerio de Salud de Chile, MINSAL). Hospital production variables were selected balancing inpatient and outpatient data using clinical expert knowledge. The data was collected during February 2015 and corresponds to the 2014 execution year. In particular we use series A, BS and 20 defined in REM (Registro Estadístico Mensual).

In Table 1 we show the description of variables used. The first column shows the category of the variables. Then some categories are further divided in two or more subcategories (column 2), and for each of them we show the number of variables and the percentage that they represent from total number of variables (257). For example, category discharged patients is further divided into three subcategories (discharges, occupied-bed days and length of stay) which in turn have 26 variables according to the clinical service from which they were collected. The variables defined are the main variables used in the DEIS database. Selection process did not consider variables that have an incomplete record or are type-specific according to the MINSAL recommendations. The detailed list of variables is available at on-line resource and as a supplementary material.

The information used to perform the technical efficiency analysis of each the results was collected from the annual budget execution report of the public insurance system, *Fondo Nacional de Salud* (FONASA). This data involves human resources expenses (Subtitle 21) and service and goods expenses (Subtitle 22).

Proposed approach

The proposed methodology is divided in five steps: "Data selection and preparation", "Clustering", "Sensibility analysis", "Clusters processing" and Comprehensive results analysis. Cluster processing considers three sections: Characterisation of clusters (Characterisation of clusters), Technical efficiency analysis (Technical efficiency evaluation) and Layout of clusters (Layout of clusters). These steps are shown in Fig. 1.

This proposal uses state of the art algorithms for performing each of the step. Also, it uses post hoc techniques for the result analysis that allow to explain and characterise each of the groups found. This methodology mainly differs from classical approaches [9] that our proposal applies technical efficiency measurements to compare all hospitals at once. The benefits of this are later discussed.

Data selection and preparation

The first step corresponds to the selection of a set of n variables to describe the profile of a hospital, incorporating information from inpatient, outpatient and emergency care. Variables are normalised using feature scaling normalisation obtaining values between 0 and 1 ($Max - Min$ scaling,

Table 1 Categories of production variables

Category	Subcategory	Number of variables	
Discharged patient	Discharges	26	(10 %)
	Occupied-bed days	26	(10 %)
	Length of stay	26	(10 %)
	Total	**78**	(30 %)
CUDYR[a]	Levels: A1-D3	12	(5 %)
Appointments of medical and other clinical professionals	Medical professional	46	(18 %)
	Other clinical professional	10	(4 %)
	Total	**56**	(22 %)
Emergency care attention	Clinical professional	7	(3 %)
Dental care attention	Procedures	26	(10 %)
Exams production	Clinical laboratory	11	(4 %)
	Imagenology	6	(2 %)
	Pathological anatomy	1	(0 %)
	Total	**18**	(7 %)
Surgeries and childbirth production	Surgeries	19	(7 %)
	Childbirth	2	(1 %)
	Total	**21**	**(8 %)**
Hemodialysis and other procedures	Hemodialysis	8	(3 %)
	Other procedures	31	(12 %)
	Total	**39**	**(15 %)**
Total		257	(100 %)

For each category we show the subcategories in which it is divided, and for each of them, the number of variables that it contains. We also show the % it represents from the total number of variables (257)

[a]CUDYR: Categorizing Users According to Care Dependency and Risk [32]

Total numbers for each category are marked in bold text

Eq. 1). A variance filter is applied to remove variables with variance less than the 10th percentile, since these variables do not contribute to the differentiation of hospitals.

$$x' = \frac{x - min(x)}{max(x) - min(x)} \tag{1}$$

In order to know the distribution of data, we apply Kolmogorov-Smirnov and Shapiro-Wilk tests with a 1 % of significance level. This result helps to decide the statistical methods that will be used during the clustering processing step.

Clustering

The next step corresponds to the search of groups of hospitals with similar performance profiles. In health context, there are different approaches for carrying out grouping of observations based on a set of variables: classically hierarchical approaches (including dendrograms and trees), centroids models (k-means), graph-based, optimisation and meta-heuristic algorithms [21]. Also, it is necessary to define a distance metric to compute the distance between hospitals according to their performance.

A hospital x is represented as a vector H^x of m values (one for each variable). Each value H_i^x represents the performance of the hospital x according to variable i. Then, the distance between a pair of hospitals is computed using a Pearson correlation based distance $(1 - \rho(H^x, H^y))$, which produces a distance matrix D with the distance between all hospitals with values between 0 and 2. With this matrix D, a clustering algorithm is applied to find groups of hospitals with similar profiles. In particular we propose the use of the *MST-kNN* graph-based clustering algorithm [22]. Briefly, the algorithm computes a complete graph $G(V, E)$ with vertex $(v_i \in V)$ representing hospitals, and an edge $(e_{ij} \in E)$ between each pair of hospitals with a weight representing the distance between profiles of the hospitals according to the distance metric used. From the complete graph, it computes two proximity graphs: minimum spanning tree (G_{MST}) and k nearest neighbour graph (G_{kNN}). Then it intersects both graphs producing $c \geq 1$ groups of objects. In each group the algorithm is recursively applied until no

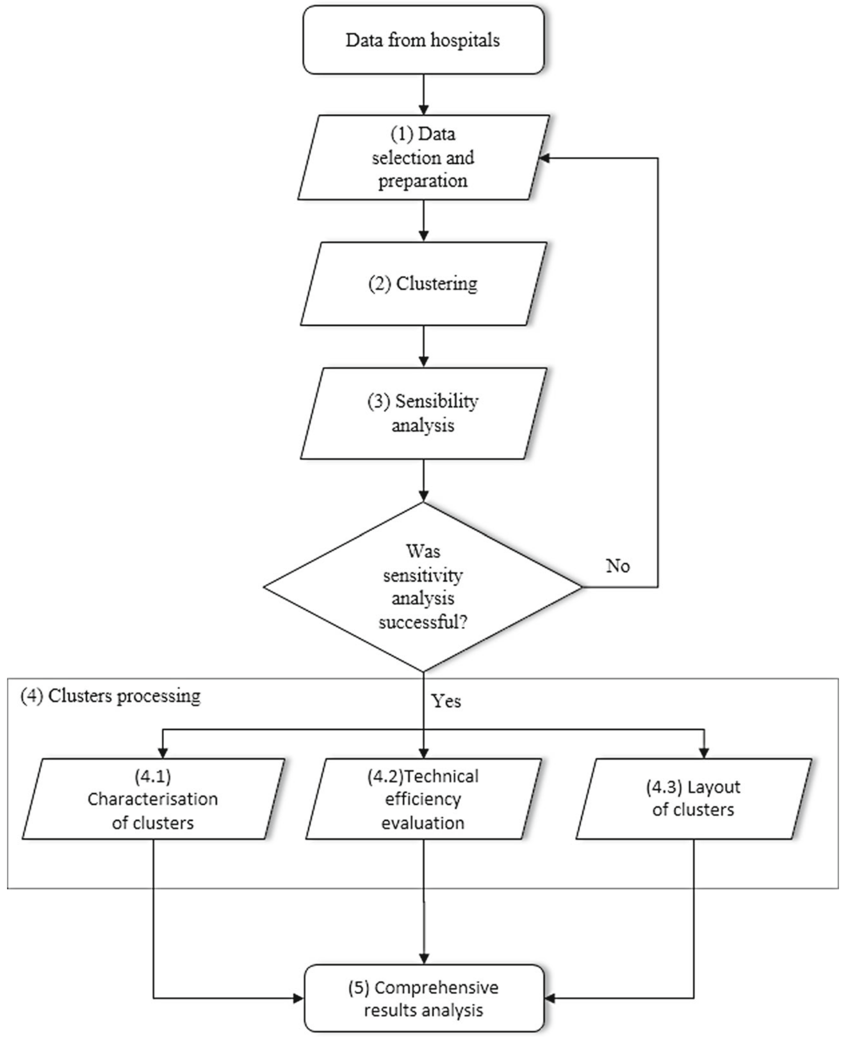

Fig. 1 Schema of the proposed methodology

further partition can be found. More details of the algorithm and its implementation can be found in Inostroza-Ponta [22].

Sensibility analysis

The next step tests the outcome of the previous two sections. The goal is to know the performance of the clustering algorithm selected, the effect of using categories of variables and to measure possible biases from the variables selected.

We implemented three tests of sensibility. The first test was designed to evaluate the quality of the MST-kNN algorithm. For this purpose, we ran two other clustering algorithms widely used, namely k-means and isodata [24], and compared their performance in terms of homogeneity and separation [26]. The k value used for the k-means algorithm considers the number of clusters obtained by the MST-kNN algorithm. The second test is designed to evaluate the selection and distribution of variables. For this purpose, the MST-kNN algorithm was applied using categories of variables (Column 1 in Table 1). After that, we compute the Jaccard Index [23] to determine the similarity of the results of using each individual category, compared with the result obtained using all the categories. If there are categories that obtain a Jaccard Index close to 1, then it is possible to remove other categories with lower Jaccard Index, since the overall result is mainly obtained because of that category. On the contrary, if all categories show a similar Jaccard Index and far from 1, then no category will be removed. The last test aims to determine the influence (possible bias) of each variable over clustering step (Clustering). For this purpose, we apply Kruskal–Wallis one-way analysis of variance by ranks test. The post hoc analysis was performed using Dunn-Sidák multiple comparison tests. The significance level applied was 1 %. Then, we rank the variables according to their p-values. Top 50, 100, 150, 250 variables are considered to apply the clustering algorithm and compared with the global model using Jaccard Index. The number of variables to be used will depend on the Jaccard index obtained. The decision to use a parametric or nonparametric test depends of the distributions of the data. If data shows a normal distribution, the statistical method applied could be replaced by a one-way ANOVA.

Clusters processing

The next step corresponds to the processing of the results found by the clustering algorithm. It is composed by three sections: (a) Characterisation of clusters, (b) Technical efficiency evaluation and (c) Layout of clusters.

a. *Characterisation of clusters*

In order to understand the specific differences in casemix or similarities among groups of hospitals, we identified variables that best differentiate each cluster from the rest. To do it, we apply the same statistical test performed in the previous step. Variables with lower p-values are selected as the ones that better explains the differences between the members of the clusters and the rest.

b. *Technical efficiency evaluation*

Technical efficiency aims to measure the relation between outcomes (hospital products) and economic resources. A hospital is technically efficient if it maximises its production consuming the minimum quantity of resources. Two well-known methods used for technical efficiency analysis are parametric stochastic frontier and non-parametric data envelopment analysis (DEA). We applied DEA since it does not depend on normally distributed data. Briefly, DEA is a non-parametric method which using linear programming techniques identifies an efficiency frontier on which efficient hospitals are placed [4]. These elements are called Decision Making Units (DMU). Each DMU can be approximated according to Eq. 2.

$$
\begin{aligned}
&\min_{\theta,\lambda} \theta,\\
&s.t. \qquad -y_i + Y\lambda \geq 0,\\
&\theta x_i - X\lambda \geq 0,\\
&N_1\lambda \leq 1,\\
&\lambda \geq 0
\end{aligned} \tag{2}
$$

Where X is a $(k \times n)$ input matrix, Y is the $(m \times n)$ output matrix, and $\theta \in [0, \infty]$. λ is a $(n \times 1)$ vector of constants that measures the weights used to compute the location of an inefficient DMU if it was to become efficient. If the restriction $N_1\lambda \leq 1$ does not exist, then the model corresponds to a constant return to scale (CRTS) model. It means that changes in the input produce proportional changes in outputs. On the contrary, the model corresponds to a variable return to scale (VRTS) model. A detailed explanation of DEA can be found in Sodani and Madnani [27] and Barnum et al. [28].

Index of CRTS and VRTS models are numbers between 0 and 1, where 1 is the maximum value of technical efficiency. For example, if the value of a specific hospital is 0.8, it means that this hospital is producing a 20 % less than a hospital with an index of 1, using the same input resources.

We implemented an output-oriented model using MaxDEA Basic 6.4 [25]. The input used was: human resources expenses and service and goods expenses and the output variables were: number of discharged patients, occupied bed day and clinical appointments.

c. *Layout of clusters*

The majority of the clustering algorithms will produce a set of disjoint groups with no or little information about how the groups are related. In order to tackle this situation we use a layout procedure to locate in a two dimensional grid the members of a cluster according to their relationships and also to locate the clusters considering the profiles between their members. As a result a layout of hospitals is produced with hospitals with similar profiles in nearby locations. In the same way, clusters that have member with similar profiles are located in nearby locations.

We used the QAPgrid algorithm [29] to produce the layout. The QAPgrid has been also used to the analysis of non-coding RNA data [30] and wine yeast identification [31]. One of the main characteristics of the algorithm is the ability to incorporate a proximity graph and a cluster result in the generation of the layout. In this case, the algorithm receives as input the distance matrix between hospitals including clustering results from "Clustering". For the algorithmic and implementation details we refer the reader to Inostroza-Ponta et al. [29].

Results

Data selection and preparation

As a result of the application of the variance filter (Data selection and preparation), six variables were removed: **four** variables from hemodialysis and other procedures category and **two** variables from discharged patient data category, remaining 251 variables. Kolmogorov-Smirnov and the Shapiro-Wilk tests showed non-normal distribution for all variables.

Table 2 Sensibility analysis of groups: comparison of clustering algorithms

Clustering method	Jaccard Index	Homogeneity Index	Separation Index	Number of clusters
MST-kNN (k = 2)	1.000	0.268	0.801	10
k-Means k = 8	0.296	0.326	0.787	8
k-Means k = 9	0.252	0.336	0.777	9
k-Means k = 10	0.301	0.309	0.789	10
k-Means k = 11	0.263	0.330	0.796	11
k-Means k = 12	0.267	0.294	0.783	12
Isodata	0.302	0.317	0.567	6

Clustering

We applied the MST-kNN algorithm on the set of 193 public hospitals. The distance between two hospital's profile is computed using the Pearson correlation based metric mentioned before. The MST-kNN algorithm produces 10 clusters of hospitals with different cardinality. On these groups we apply the sensibility analysis and the Clustering Processing according to the three sections.

Sensibility analysis of groups

Table 2 shows a comparison between the clustering obtained by MST-kNN, *k*-means and isodata. The MST-kNN obtains the best scores of Homogeneity Index, 0.268 (column 3) and Separation Index, 0.801 (column 4). Isodata has the worst quality results, finding only six groups. None of the methods resulted in similar clustering compared to MST-kNN approach (column 2) which means that the MST-kNN algorithm finds clusters of hospitals that are not found the other methods.

The sensibility analysis applied by category of variables is shown in Table 3. Column 2 shows results of the Jaccard Index when compared to the use of all variables. It is possible to see that the results obtained are not similar with a maximum Jaccard Index of 0.280 for category hemodialysis and other procedures. Column 3 and 4 show the results of Homogeneity and Separation index respectively. The best index are found when we used all categories of variables. These results show a good performance of the proposal compared to the clustering results using individual categories of variables.

Table 4 shows the results of applying the proposal using the Top 50, 100, 150, 200 and 250 variables sorted by smallest p-value obtained from the Kruskal-Wallis procedure (Sensibility analysis. Column 2 shows the Jaccard index of the clustering obtained compared to the use of variables. It is possible to see that by eliminating the 50 worst variables, the Jaccard Index immediately decreases from 1 to 0.272. This result validates the importance of each variable selected in data selection and preparation step (Data selection and preparation) and confirms the diversity of variables determined by expert knowledge. Columns 3 and 4 show a good performance in terms of Homogeneity (0.268) and Separation (0.801) using 250 variables. Only when we use the top 100 variables we obtain a slightly better Separation Index (0.816), but the Homogeneity index is much worst (0.323).

Table 3 Sensibility analysis of groups: variable selection

Categories of variables	Jaccard Index	Homogeneity Index	Separation Index
All variables	1.000	0.268	0.801
Discharged patient data	0.154	0.371	0.767
Categorizing Users According to Care Dependency and Risk	0.217	0.443	0.672
Appointments of medical and other clinical professionals	0.255	0.354	0.571
Emergency care attention	0.219	0.358	0.644
Dental care attention	0.223	0.487	0.630
Exams production	0.141	0.428	0.676
Surgeries production and childbirth	0.258	0.396	0.755
Hemodialysis and other procedures	0.280	0.510	0.681

Table 4 Sensibility analysis of groups: main variables sorted by p-value

Number of variables	Jaccard Index	Homogeneity Index	Separation Index	Number of clusters
250 variables	1.000	0.268	0.801	10
200 variables	0.272	0.349	0.718	7
150 variables	0.297	0.363	0.789	7
100 variables	0.200	0.323	0.816	10
50 variables	0.196	0.285	0.795	12

In addition to previous evaluations, sensitivity of decimals in numerical data was also measured demonstrating that the model remains constant with a minimum of 3 numbers after the decimal point. This result considers feature scaling normalisation (Eq. 1).

Clusters analysis

a. *Characterisation of clusters*

After applying the process described in "Clusters processing", we are able to characterise each cluster. In this step we use extra data such as location of hospitals and clinical complexity according MINSAL categorisation to characterise the clusters found. These data were not consider during the application of the proposal.

Cluster 1 and **Cluster 2** include high complexity hospitals and referral centres located in main cities with a large target population. Hospitals in Cluster 1 have a higher production in obstetric, neonatology and paediatrics services than Cluster 2. **Cluster 8** contains the paediatrics hospitals.

Cluster 0 includes medium and high complexity hospitals that are not located in large cities. These have a lower production in relation to high complexity hospitals. **Cluster 4** has all psychiatric hospitals and **Cluster 9** has private hospitals economically supported by government (e.g. Hospital Parroquial de San Bernardo and Hospital de Puerto Varas).

Clusters 3, 5, 6 and 7 include low complexity hospitals. These hospitals are located in small cities or rural areas and their production is lower in comparison with medium complexity hospitals. Medical care is mainly performed by general practitioners. In spite of these similarities, features in these clusters differ from each other. **Cluster 3** includes mainly hospitals with complex patients. **Cluster 5**

Table 5 Details of average of technical efficiency and cost by cluster

Hospitals			Technical efficiency			HR expenses			S & G expenses		
Cluster	N (%)	CMPLX	CRTS	VRTS	Scale	$/DIS	$/OBD	$/ACP	$/DIS	$/OBD	$/ACP
0	23 (11.9 %)	7-11-5 (1.9)	0.747	0.879	0.851	2,303	474	128	951	195	57
1	17 (8.8 %)	0-0-17 (3.0)	0.955	0.989	0.965	2,474	342	197	1,904	259	146
2	38 (19.7 %)	0-4-34 (2.9)	0.826	0.877	0.937	2,300	359	178	1,565	224	111
3	71 (36.8 %)	67-4-0 (1.1)	0.683	0.757	0.907	3,541	615	85	1,063	185	26
4	4 (2.1 %)	0-3-1 (2.3)	0.895	1.000	0.895	17,631	116	206	6,366	46	78
5	7 (3.6 %)	7-0-0 (1.0)	0.836	1.000	0.836	2,094	375	139	591	115	44
6	23 (11.9 %)	20-2-1(1.2)	0.636	0.836	0.772	2,480	388	92	960	152	33
7	3 (1.6 %)	3-0-0 (1.0)	1.000	1.000	1.000	6,973	1,983	101	949	270	17
8	4 (2.1 %)	0-0-4 (3.0)	0.962	1.000	0.962	2,617	546	107	1,731	351	79
9	3 (1.6 %)	0-2-1 (2.3)	1.000	1.000	1.000	1,611	117	41	235	51	18

First it shows the size of the cluster (with the percenatge from the total), CMPLX shows the complexity of the members according to MINSAL classification: Low, Medium, High, and an average complexity score, CRTS shows constant return to scale, VRTS shows variable return to scale, Scale: CRTS/VRTS, HR expenses: human resources expenses, S & G expenses: service and goods expenses, DIS: discharges, OBD: occupied bed days and ACP: appointment of clinical professionals, $:US dollar

Table 6 VRTS (variable return to scale) results for high complexity hospitals

Hospitals	No DRGs	Santelices et al. [9]	MST-kNN	
	VRST	VRST	Cluster	VRST
HOSP BASE DE OSORNO	0.594	–	0	1.000
HOSP DR JUAN NOE CREVANNI (ARICA)	0.680	0.837	0	1.000
HOSP SAN JOSE (CORONEL)	1.000	–	0	1.000
HOSP DE TOME	0.894	–	0	0.914
HOSP DE LOTA	0.848	–	0	0.875
INST DE NEUROCIRUGIA DR ALFONSO ASENJO	0.423	–	1	1.000
HOSP DR LAUTARO NAVARRO AVARIA (PUNTA ARENAS)	0.511	–	1	1.000
HOSP DR GUSTAVO FRICKE (VIA DEL MAR)	0.660	0.708	1	1.000
HOSP DR FELIX BULNES CERDA (QUINTA NORMAL)	0.614	0.478	1	1.000
INST NAC DE ENF RESP Y CIRUGIA TORACICA	0.589	1.000	1	1.000
HOSP CARLOS VAN BUREN (VALPARAISO)	0.684	1.000	1	1.000
HOSP DR LEONARDO GUZMAN (ANTOFAGASTA)	0.588	–	1	1.000
HOSP SAN JUAN DE DIOS (SANTIAGO)	0.650	0.557	1	1.000
HOSP LAS HIGUERAS (TALCAHUANO)	0.680	–	1	1.000
HOSP DR CESAR GARAVAGNO BUROTTO (TALCA)	0.635	–	1	1.000
HOSP BARROS LUCO TRUDEAU (SAN MIGUEL)	0.601	0.528	1	1.000
HOSP CLI REGIONAL (VALDIVIA)	0.664	0.636	1	1.000
COMP HOSP DR SOTERO DEL RIO (PUENTE ALTO)	0.695	1.000	1	1.000
HOSP DR HERNAN HENRIQUEZ ARAVENA (TEMUCO)	0.712	0.648	1	1.000
HOSP CLI REG DR GUILLERMO GRANT BENAV. (CONCE)	0.610	–	1	1.000
HOSP DR ERNESTO TORRES GALDAMES (IQUIQUE)	0.543	0.694	1	0.984
HOSP CLI SAN BORJA-ARRIARAN (SANTIAGO)	0.550	–	1	0.833
COMP ASIS DR VICTOR RIOS RUIZ (LOS ANGELES)	0.712	0.736	2	1.000
HOSP CLI HERMINDA MARTIN (CHILLAN)	0.700	0.667	2	1.000
HOSP SAN MARTIN (QUILLOTA)	0.976	–	2	1.000
HOSP DR LUIS TISNE B (PEALOLEN)	1.000	0.998	2	1.000
COMP HOSPARIO SAN JOSE (INDEP)	1.000	0.913	2	1.000
HOSP PRESIDENTE CARLOS IBAEZ DEL CAMPO (LINARES)	0.954	–	2	1.000
INST NAC GER PRESIDENTE EDUARDO FREI MONTALVA	0.897	–	2	1.000
HOSP DR CARLOS CISTERNAS (CALAMA)	0.719	–	2	1.000
HOSP DE SAN CARLOS	0.788	–	2	1.000
HOSP DR EDUARDO PEREIRA RAMIREZ (VALPARAISO)	0.743	–	2	1.000
HOSP SAN JUAN DE DIOS (LOS ANDES)	0.885	–	2	1.000
HOSP CLAUDIO VICUA (SAN ANTONIO)	0.893	–	2	0.993
HOSP EL PINO (SAN BERNARDO)	0.812	1.000	2	0.986
HOSP DR MAURICIO HEYERMANN (ANGOL)	0.802	1.000	2	0.975
HOSP SAN JOSE (VICTORIA)	0.741	–	2	0.968
HOSP DE SAN CAMILO (SAN FELIPE)	0.886	–	2	0.966
HOSP PADRE ALBERTO HURTADO (SAN RAMON)	0.686	–	2	0.933
HOSP SAN JUAN DE DIOS (SAN FERNANDO)	0.804	–	2	0.910
HOSP SAN JUAN DE DIOS (CURICO)	0.764	0.787	2	0.876
HOSP REGIONAL DE RANCAGUA	0.612	0.730	2	0.870
HOSP SAN JOSE (MELIPILLA)	0.755	–	2	0.869
HOSP DE QUILPUE	0.693	–	2	0.864
HOSP DE PUERTO MONTT	0.566	–	2	0.840
HOSP SAN PABLO (COQUIMBO)	0.683	0.720	2	0.830
HOSP DE CASTRO	0.585	1.000	2	0.826

Table 6 (continued)

Hospitals	No DRGs	Santelices et al. [9]	MST-kNN	
	VRST	VRST	Cluster	VRST
HOSP DR ANTONIO TIRADO LANAS (OVALLE)	0.731	1.000	2	0.808
INST TRAUMATOLOGICO DR TEODORO GEBAUER	0.709	–	2	0.806
HOSP DEL SALVADOR (PROVIDENCIA)	0.653	0.689	2	0.797
HOSP SAN JOSE DEL CARMEN (COPIAPO)	0.634	–	2	0.781
HOSP SAN JU AN DE DIOS (LA SERENA)	0.725	–	2	0.767
HOSP DE URG ASIS PUB DR ALEJANDRO DEL RIO (SANT.)	0.297	0.607	2	0.572
HOSP REGIONAL DE COIHAIQUE	0.396	0.792	2	0.522
HOSP CLI MET EL CARMEN DR LUIS VALENTIN FERRADA	0.370	–	2	0.502
HOSP CLI MET LA FLORIDA DRA. ELOISA DIAZ INSUNZA	0.341	–	2	0.407
INST PSI DR JOSE HORWITZ BARAK (RECOLETA)	1.000	–	4	1.000
INST NAC DEL CANCER DR CAUP PARDO CORREA (REC)	0.677	–	6	1.000
HOSP DE NIOS DR LUIS CALVO MACK (PROVI)	0.494	1.000	8	1.000
INST NAC DE REHA INF PRESI PEDRO AGUIRRE CERDA	1.000	–	8	1.000
HOSP DR EXEQUIEL GONZALEZ CORTES (SAN MIGUEL)	0.555	1.000	8	1.000
HOSP CLI DE NIOS DR ROBERTO DEL RIO (INDEP)	1.000	0.630	8	1.000

First column shows the technical efficiency without DRGs information. Second column shows results from previous work, incorporating data from DRGs partially integrated. Third column shows results using the proposed approach

and **Cluster 6** have hospitals with a larger number of emergency and paediatrics activities in comparison to the other low complexity hospitals, respectively. Finally, **Cluster 7** has hospitals with the lowest level of activities. Details of all these comparisons can be found in on-line resource. It also provides the ability to compare particular clusters of interest.

b. *Technical efficiency analysis*

Table 5 shows the average technical efficiency of each group and their expenses according to medical discharged, occupied bed day and clinical appointments.

Column 1 shows the Cluster ID and column 2 shows the number and percentage of hospitals in each cluster. Using complexity in health care determined by MINSAL, hospitals are classified in low, medium and high complexity. In order to compute a complexity score of the cluster, we assign a value of 1, 2 and 3, for low, medium and high complexity respectively. The complexity profile and the average score of each cluster is shown in column 3. This score allows to see if there is a correlation between complexity and resources consumption.

Columns 4, 5 and 6 show technical efficiency results. In column 4, **Clusters 0, 2, 3 and 6** have values close to 0.7 and 0.8 for CRTS. It means that hospitals in these groups produce 20–30 % less using the same resources. This tendency is repeated in column 5 for VRTS.

Columns 7, 8 and 9 show human resources expenses in each cluster according hospitals production. In column 7, the higher spending corresponds to **Cluster 4**, because psychiatric hospitals in this cluster do not produce discharges. However, when we consider the spending per occupied bed day, this cluster has the lowest value (column 8). **Cluster 5** in column 9 has the highest spending in relation to clinical appointments. Human resource expenses in a country-level are US$3.149, US$447 and US$126 according to medical discharges, occupied bed day and clinical appointments, respectively. There is no correlation between hospital complexity and human resource expenses.

Finally, columns 10, 11 and 12 show service and goods expenses in relation to hospitals production. Excepting **Cluster 4**, column 10 shows high expenses for **Clusters 1, 2 and 8** according to hospitals discharges.This is because services and goods expenses consider important spending as pharmacy and imagenology for referral hospitals. This is repeated for spending per occupied bed day and clinical appointments shown in column 11 and 12. Service and goods expenses in a country-level are US$1.333, US$196 and US$62 according to medical discharges, occupied bed day and clinical appointments, respectively. Hospital complexity and service and goods expenses related to clinical appointments shows a correlation of 0.776. It means that high complexity hospital tend to have a larger expenditure in this item.

In Table 6 we show a comparison of technical efficiency (VRTS) for high complexity hospitals using three alternatives: all hospitals compared at once (column 2), applying incomplete data adjusted for DRGs partially implemented [9] (column 3), and the results obtained by MST-kNN algorithm (columns 5). It is possible to see that results of each alternative differ, which means that they are showing a different perspective of the technical efficiency. For example, **Hospital Dr. Ernesto Torres Galdames** in Table 6 has values for technical efficiency (VRTS) of 0.543, 0.694 and 0.984 using each alternative. A score of 0.694 indicates that this hospital produces about 30 % less according to DRGs data [9]. However, when this hospital is compared with hospitals with similar size and casemix, **Cluster 1**, it does not differ greatly with the rest of the cluster, producing only 1.6 % less. Technical efficiency increases because MST-kNN approach considers outpatient production like medical appointments. For example, products and services expenses by clinical appointments are US$112. This value is lower in relation to the average cluster expenses (US$146) increasing technical efficiency.

Another example is **Hospital Clínico de Niños Dr. Roberto del Río**. In this case, the result using DRGs data shows a techincal efficiency of 0.630, a significantly lower value compared with score 1.00 obtained using MTS-kNN approach. The technical efficiency increases because this hospital is now being compared only with paediatrics hospitals in **Cluster 8**, that have similar characteristics according to the clustering algorithm.

Details of each cluster can be seen in the supplementary material (Supplementary Tables file).

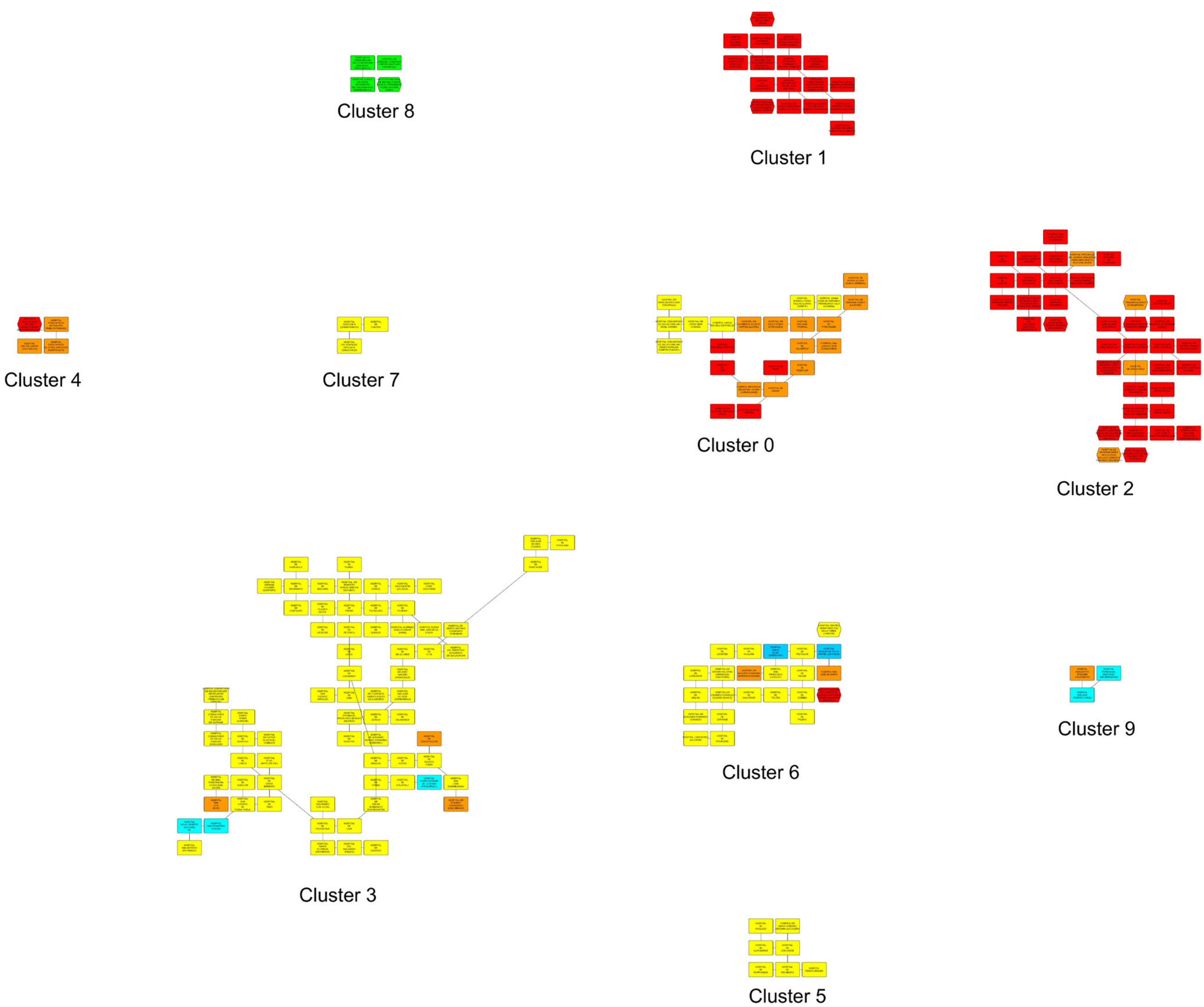

Fig. 2 Final layout of clusters using QAPgrid. The clustering algorithm has provided 10 clusters. In *red colour* are shown hospitals with high health care complexity, orange colour hospitals with medium complexity, *yellow colour* hospitals with low complexity, *green colour* paediatrics hospital and cyan other hospitals. Classification according to MINSAL. http://bioinformatic.diinf.usach.cl/publichealth

c. *Layout of clusters*

The layout of clusters is shown in Fig. 2. Clusters are located according to the similarity of members that represents their clinical complexity. The right top side of the figure shows hospitals with high clinical complexity **(Clusters 1, 2 and 8)** and the left bottom side shows hospitals with low clinical complexity **(Clusters 3, 5, 6 and 7)**. **Cluster 8** with paediatrics hospitals is near to **Cluster 1** which includes high clinical complexity hospitals with obstetric, neonatology and paediatrics services. **Cluster 4** is isolated at the right side of the layout because it has all psychiatric hospitals with high and medium clinical complexity.

Discussion and conclusions

When comparing the technical efficiency of hospital it is a common mistake to compare hospital with different casemix as one. Because of the above, DRGs are used to overcome this situation, which are not always available. We propose a novel approach for comparing technical efficiency of hospitals with similar casemix.

The proposal generates groups of hospitals with similar casemix and different sizes. When analysing the technical efficiency of these groups it is possible to see that they also have a high technical efficiency among them. It shows that grouping hospitals based on their production profiles, allows to better compare the efficiency of hospitals. Furthermore, in each group it is possible to identify hospitals with similar production profiles that are not as efficient as other with the same profile. The above allows to generate policies to increase the expenditure of resources in the health administration.

The results obtained differ from studies with partially implemented DRG, increasing the efficiency of some hospitals. This difference can be explained because of two factors: it compares the technical efficiency of hospitals with similar casemix and it considers variables like medical and other professional appointment that are not currently consider in the DRG.

In the Chilean health system there are some hospitals that are considered by the experts as particular hospitals. It means that experts avoid to compare them with the rest of hospitals. In this work we consider all hospitals and we only used their production profile as a criteria for the clustering conformation. In order to highlight those hospitals we show them as hexagons on Fig. 2, like Instituto Nacional del Cáncer Dr. Caupolicán Pardo Correa, Instituto Nacional de Rehabilitación Infantil Presidente Pedro Aguirre Cerda and Hospital de Enfermedades Infecciosas Dr. Lucio Córdova among others. It is worth to mention that these hospitals correspond to outliers in the cluster when we applied the inter-quartil range method over distances between hospitals.

The results obtained by the strategy show groups of hospitals which have the same complexity MINSAL classification in the majority of its members (third column in Table 5). The only two clusters that do not agree with this are Cluster 0 and Cluster 4. In the case of Cluster 0 it has a mixture of low, medium and high complexity hospitals, with non of the categories representing more than the 50 %. Some of the hospitals that belong to the cluster are of high complexity but they show a production profile similar to those of medium complexity. In the same way the low complexity hospitals that belong to the cluster have production profiles more similar to the medium complexity hospitals of the cluster. In the case of Cluster 4, it is composed by all of the psychiatric hospitals considered. This result allows to compare the technical efficiency of these specific type of hospitals.

The strategy proposed in this work has well defined steps and sensitivity analysis of the results is performed as part of the strategy. The above aims to guarantee significant results. Also, it uses a non-parametric method to calculate technical efficiency: Data Envelopment Analysis. This involves some limitations: results are sensitive to inputs and outputs selection and the number of efficient hospitals on the frontier tends to increase proportionally to the number inputs and outputs [33]. DEA can be modified applying other variables, using an input-based orientation or be replaced for a more recent technique without alter the overall strategy. Other variables can be considered during the application of the strategy. For example, we may incorporate variables from geographical, political or demographics data. Also, the model will be able to include data from DRGs when the progress in their implementation is completed.

Further technical efficiency evaluation with strategy can be incorporated to support hospital budgets distribution, compare hospitals from various countries and evaluate different periods of time, allowing the visualisation of the evolution of each hospital inside or among groups.

Acknowledgments The authors would like to thank the Chilean Ministry of Health (Ministerio de Salud, MINSAL) and the South Metropolitan Health Service (Servicio de Salud Metropolitano Sur, SSMS) for their cooperation providing the data needed for this academic research. Special acknowledgements to Managers, Finance, Statistics and DRGs departments. MV-C would like to thank CITIAPS. This article was partially funded by Proyect PMI USA1204.

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Alonso, J., Clifton, J., and Díaz-Fuentes, D., The impact of new public management on efficiency: An analysis of Madrid's hospitals. *Health Policy* 119(3):333–340, 2015.
2. Al-Amin, M., Makarem, S. C., and Rosko, M.: Efficiency and hospital effectiveness in improving hospital consumer assessment of healthcare providers and systems ratings. *Health Care Manage Rev.* 2015 Jun 19. [Epub ahead of print] PubMed PMID: 26099007.
3. Hadji, B., Meyer, R., Melikeche, S., Escalon, S., and Degoulet, P., Assessing the relationships between hospital resources and activities: A systematic review. *J. Med. Syst.* 38(10):127–148, 2014.
4. Asandului, L., Roman, M., and Fatulescu, P., The efficiency of healthcare systems in Europe: A data envelopment analysis approach. *Procedia Econ. Finance* 10:261–268, 2014. International Conference on Applied Statistics (ICAS) 2013.
5. Goudarzi, R., Pourreza, A., Shokoohi, M., Askari, R., Mahdavi, M., and Moghri, J., Technical efficiency of teaching hospitals in Iran: The use of Stochastic Frontier Analysis, 1999–2011. *Int. J. Health Policy Manag.* 3(2):91–97, 2014.
6. O'Neill, L., Rauner, M., Heidenberger, K., and Kraus, M., A cross-national comparison and taxonomy of DEA-based hospital efficiency studies. *Socio Econ. Plan. Sci.* 42(3):158–189, 2008.
7. Rezaee, M. J., and Karimdadi, A., Do geographical locations affect in hospitals performance? A multi-group data envelopment analysis. *J. Med. Syst.* 39(9):278, 2015.
8. Herwartz, H., and Strumann, C., Hospital efficiency under prospective reimbursement schemes: An empirical assessment for the case of Germany. *Eur. J. Health Econ.* 15(2):175–186, 2014.
9. Santelices, E., Ormeño, H., Delgado, M., Christopher, M., Valdés, R., and Durán, L., Análisis de la eficiencia técnica hospitalaria 2011. *Rev. Méd. Chile* 141:332–337, 2013.
10. Linna, M., Häkkinen, U., Peltola, M., Magnussen, J., Anthun, K., Kittelsen, S., Roed, A., Olsen, K., Medin, E., and Rehnberg, C., Measuring cost efficiency in the Nordic Hospitals—a cross-sectional comparison of public hospitals in 2002. *Health Care Manag. Sci.* 13(4):346–357, 2010.
11. Vitikainen, K., Linna, M., and Street, A., Substituting inpatient for outpatient care: What is the impact on hospital costs and efficiency? *Eur. J. Health Econ.* 11(4):395–404, 2010.
12. Magnussen, J., and Nyland, K., Measuring efficiency in clinical departments. *Health Policy* 87(1):1–7, 2008.
13. Gannon, B., Testing for variation in technical efficiency of hospitals in Ireland. Technical report, Economic and Social Research Institute (ESRI), 2005.
14. Wang, Z., Liu, R., Li, P., Jiang, C., and Hao, M., How to make diagnosis related groups payment more feasible in developing countries—A case study in Shanghai, China. *Iran. J. Public Health* 43(5):572–578, 2014.
15. Jurges, H., and Koberlein, J., What explains DRG upcoding in neonatology? The roles of financial incentives and infant health. *J. Health Econ.* 43:13–26, 2015.
16. Jian, W., Lu, M., Han, W., and Hu, M.: Introducing diagnosis-related groups: Is the information system ready? *Int. J. Health Plan. Manage*, 2014 Aug 11. doi:10.1002/hpm.2270.
17. Wardle, G., Wodchis, W. P., Laporte, A., Anderson, G.M., and Ross-Baker, G., The sensitivity of adverse event cost estimates to diagnostic coding error. *Health Serv. Res.* 47:984–1007, 2012.
18. Schousboe, J. T., Paudel, M. L., Taylor, B. C., Mau, B. A., Virnig, L. W., Ensrud, K. E., and Dowd, B. E., Estimation of standardised hospital costs from Medicare claims that reflect resource requirements for care: Impact for cohort studies linked to Medicare claims. *Health Serv. Res.* 49(3):929–949, 2014.
19. Malyon, R., Zhao, Y., and Oates, B., Differences in the cost of admitted patient care for indigenous people and people from remote locations. *Aust. Health Rev.* 37(1):26–31, 2013.
20. Jordan, I., Barry, H., Clancy, M., O'Toole, D., and Machale, S., Financial impact of accurate discharge coding in a liaison psychiatry service. *J. Psychosom. Res.* 73(6):476–478, 2012.
21. Gan, G., Ma, C., and Wu, J., *Data clustering - theory, algorithms, and applications*. SIAM, 2007.
22. Inostroza-Ponta, M., An integrated and scalable approach based on combinatorial optimisation techniques for the analysis of microarray data. University of Newcastle. School of Electrical Engineering and Computer Science, 2008.
23. Jay, J. J., Eblen, J. D., Zhang, Y., Benson, M., Perkins, A. D., Saxton, A. M., Voy, B. H., Chesler, E. J., and Langston, M. A., A systematic comparison of genome-scale clustering algorithms. *BMC Bioinformatics* 13(Suppl 10):S7, 2012.
24. Navulur, K. *Multispectral image analysis using the object-oriented paradigm*. Hoboken: CRC Press, 2006.
25. Gang, C., Data envelopment analysis software and MaxDEA, 2014.
26. Chen, G., Jaradat, S. A., Banerjee, N., Tanaka, T. S., Ko, M. S. H., and Zhang, M. Q., Evaluation and comparison of clustering algorithms in anglyzing es cell gene expression data. *Stat. Sin.* 12:241–262, 2002.
27. Sodani, P. R., and Madnani, G. M. K., Measuring hospital performance through data envelopment analysis: Understanding basic concepts to help novice researchers. *J. Health Manag.* 10(1):129–142, 2008.
28. Barnum, D., Johnson, M., and Gleason, J., Importance of statistical evidence in estimating valid DEA scores. *J. Med. Syst.* 40(3):1, 2015.
29. Inostroza-Ponta, M., Berretta, R., and Moscato, P., QAPgrid: A two level QAP-based approach for large-scale data analysis and visualisation. *PLoS ONE* 6(1):e14468, 2011.
30. Clark, M. B., Johnston, R. L., Inostroza-Ponta, M., Fox, A. H., Fortini, E., Moscato, P., Dinger, M. E., and Mattick, J. S., Genome-wide analysis of long noncoding RNA stability. *Genome Res.* 22(5):885–898, 2012.
31. Ramirez-Castrillon, M., Mendes, S. D., Inostroza-Ponta, M., and Valente, P., (GTG)5 MSP-PCR fingerprinting as a technique for discrimination of wine associated yeasts? *PLoS ONE* 9(8):e105870, 2014.
32. Arechabala-Mantuliz, M., Barrios-Araya, S., Catoni-Salamanca, M., Palma-Castro, E., Ibacache-Celedón, Y., and Richards, J., Validación de un instrumento de medición de cuidados según dependencia y riesgo en pacientes en terapia de hemodiálisis crónica. *Aquichán* 14:473–484, 2014.
33. Kumar, N., Efficiency analysis of banks using DEA: A review. *Int. J. Adv. Res. Innov.* 1:120–126, 2014.