FEP WORKING PAPERS
FEP WORKING PAPERS

Research Work in Progress

n. 563 July 2015

ISSN: 0870-8541

Comparing Comparables: An Approach to Accurate Cross-country Comparisons of Health Systems for Effective Healthcare Planning and Policy Guidance

Mário Amorim Lopes[1,2]
Carlos Soares[1,2]
Álvaro Almeida[3,4]
Bernardo Almada-Lobo[1,2]

[1] INESC-TEC
[2] FEUP, Faculty of Engineering, University of Porto
[3] CEF.UP, Research Center in Economics and Finance, University of Porto
[4] FEP-UP, School of Economics and Management, University of Porto

COMPARING COMPARABLES: AN APPROACH TO ACCURATE CROSS-COUNTRY COMPARISONS OF HEALTH SYSTEMS FOR EFFECTIVE HEALTHCARE PLANNING AND POLICY GUIDANCE

Mário Amorim Lopes *

INESC-TEC and DEIG, Faculdade de Engenharia, Universidade do Porto

Carlos Soares

INESC-TEC and DEI, Faculdade de Engenharia, Universidade do Porto

Álvaro Almeida †

CEF.UP and Faculdade de Economia, Universidade do Porto

Bernardo Almada-Lobo

INESC-TEC and DEIG; Faculdade de Engenharia, Universidade do Porto

May 8, 2015

* Email: mario.lopes@fe.up.pt

† Email: almeida@fep.u ... 00-464 Porto. CEF.UP, Centre for Economics and Fina ... y the Foundation for Science and Technology (FCT),

ABSTRACT

With rising healthcare costs, using health personnel and resources efficiently and effectively is critical. International cross-country and simple worker-to-population ratio comparisons are frequently used for improving the efficiency of health systems, planning of health human resources, and guiding policy changes. These comparisons are made between countries typically of the same continental region. However, if used imprudently, inconsistencies arising from frail comparisons of health systems may outweigh the positive benefits brought by new policy insights. In this work, we propose a different approach to international health system comparisons. We present a methodology to group similar countries in terms of mortality, morbidity, utilization levels, and human and physical resources, which are all factors that influence health gains. Instead of constructing an absolute rank or comparing against the average, the method finds countries that share similar ground, upon which more reliable comparisons can then be conducted, including performance analysis. We apply this methodology using data from WHO's HFA-DB, and we present some interesting empirical relationships between indicators that may provide new insights into how such information can be used to promote better healthcare planning and policy guidance.

Keywords: health systems, health indicators, health policy

JEL Classification: I18, I19

1. INTRODUCTION

Healthcare costs have increased sharply over the years, well above the average growth rate of the GDP (Chandra *et. al*, 2013). The immediate impact is a decrease in the real income of the population, reducing the disposable income available for other living expenses (Auerbach and Kellermann, 2011). Wealthier countries may be able to accommodate such increases but others may be confronted with the sensitive situation of having to opt between drugs and treatments based on their cost, and not on their clinical merits alone. In the worst-case scenario, no treatment is offered at all, or rationing is imposed through waiting lists. In order to avoid facing such dilemma, policy-makers must strive to manage healthcare resources efficiently, both physical and human.

Health human resources (HHR) planning has been identified as a fundamental tool for mitigating rampant healthcare costs while preserving the quantity and quality of service provided (Dreesch, 2005). Briefly defined, HHR consists in assessing the right number of people with the right skills, in the right place at the right time, to provide the right services to the right people (Birch, 2002). There are multiple approaches to HHR planning, each with its advantages and drawbacks (Amorim Lopes *et. al*, 2015). Benchmarking, simple worker-to-population comparisons and other comparative analysis techniques are approaches frequently used to draw international comparisons between healthcare systems, including assessing HHR needs. The techniques consist of identifying similar regions or countries in terms of demographic and epidemiological profiles but differing sharply in the cost structure and resource allocation (Roberfroid *et. al*, 2009).

Although benchmarking and other comparative analysis may be useful tools to assess and compare health systems, including HHR resources, imprudent use may take its toll. For example, consider the decision by the British National Health Service to increase the intake to medical schools by 60%, a resolution motivated by the observation that the physician-to-population ratio was low in comparison with other OECD countries (Bloor *et. al*, 2006). The policy was adopted without first evaluating for other criteria that may affect the performance of the medical staff, namely the skill mix and the productivity of HHR. Subsequent research justified the lower ratio of physicians with a better distribution of the skill mix and increased productivity resulting from a more efficient task delegation (Bloor and Maynard, 2003). Health system performance comparisons and composite indexes have also been subject to heavy criticism, in part due to some methodological fragilities (Richardson *et. al*, 2003).

Provided that a valid methodology is used to support cross-country comparisons, comparative analyses may continue to be an easy and straightforward way of gaining quick insight into the health system performance of a country by observing the best practices, especially when few data exist to conduct more advanced analyses (Amorim Lopes *et. al*, 2015). There are performance profiles and typological classifications of health systems in the literature that identify countries with similar systems (for a review of typologies of health systems, see Burau and Blank, 2006). Previous studies have mapped European healthcare systems according to a subset of indicators on healthcare expenditure, healthcare financing, healthcare provision and institutional characteristics, and then establishing a relative performance index between groups (Wendt, 2009). In other studies, several other dimensions, such as acquisition of human, financial, technological or material resources, health outcomes, risk factors or equity on access to healthcare were used to elaborate an absolute performance index, which is then used to group countries according to their performance profiles (Tchouaket *et. al*, 2012). International comparisons have also been conducted using nonparametric techniques like Data Envelopment Analysis (Bhat, 2005), techniques typically used for micro-level service efficiency measurements, such as hospital units. DEA measures the efficiency of health systems by calculating the ratio between health outcomes and healthcare spending.

In this work, we propose a new methodology to perform cross-country comparisons. As an alternative to constructing composite indices or absolute performance rankings, we start by creating clusters of countries that have similar results in several reference indicators, including those usually associated with demand for healthcare services (mortality and morbidity-based indicators, and utilisation statistics), and with the supply of healthcare services (physical and human resources available). This allows for intra and inter-group local comparisons, avoiding attractive and yet inconclusive global performance rankings of substantially different health systems that have generated discord (Richardson *et al.*, 2003)**,** if not outright criticism (Bronnum-Hansen, 2014).

We then apply the method to data from World Health Organization's (WHO) Health for All database (HFA-DB). Clusters are generated for each indicator and then intersected against each other to obtain groups of countries with similar features in more than one dimension. With this bottom-up methodology, future research using benchmarking or DEA can build upon a reliable basis of countries exhibiting similar trends in parts of their health systems. These indicators can then be used within HHR planning models, or any other area of healthcare planning and health policy, to improve the forecasts and projections to assist decision and policy makers.

The remainder of this paper is organised as follows: in Section 2 we describe the methodology proposed, and in Section 3 we apply that methodology to WHO's HFA-DB database and present the results. A discussion of the results and some empirical insights is provided in Section 4. Finally, Section 5 concludes this paper with a brief summary and future research topics.

2. METHODOLOGY

Our methodology consists of grouping countries that are similar to each other in different dimensions of a health system. In this particular case, we consider the following: mortality-based indicators, morbidity-based indicators, utilisation indicators, physical resources and human resources. We then analyse which countries feature in the same cluster in more than one dimension, which makes it possible to generate a similarity matrix. Comparisons and cross-country performance analyses can then be conducted within the cluster, comparing countries with the local benchmark serving as a reference, or between clusters. By deliberately narrowing the scope, we ensure that cross-country comparisons are performed between countries with similar characteristics.

We use a sample of the WHOs's European Health for All database (HFA-DB), targeting countries belonging to the European Union, as European countries share common ground and have health systems that derive from either Bismarckian or Beveredgian models. Notwithstanding, this study can be applied interchangeably to any set of countries provided that data are available.

2.1. Data sources

The main source of data was the HFA-DB, last updated in April 2014. This database contains a selection of core health statistics covering basic demographics, health status, health determinants and risk factors, healthcare resources, utilisation and expenditure in the 53 countries in the WHO European Region (Europe WROF, 2015). The data are compiled from various sources, including WHO's and United Nations' European delegation offices, OECD and Eurostat.

2.2 Selection of indicators

For this study, we have selected indicators that characterise not only the supply of and demand for healthcare services, but are also used to assess input/output and outcome efficiency. On the demand side, we have selected mortality-based, morbidity-based, and healthcare utilisation indicators. These indicators are sometimes also used as indicators to measure output efficiency (see Varabyova and Schreyögg, 2013). For the supply side, we resorted to healthcare resources, both physical and human, also commonly used as proxies to measure input efficiency. These indicators were selected due to the availability of data, and also because they share a list of desirable features: valid, communicable, effective, reliable, objective, available, contextual, attributable, interpretable, comparable, remediable and repeatable (Pringle *et. al*, 2002). This is critical for any subsequent expert validation of the clusters formed.

On the demand-side, mortality-based indicators are provided as Standardised Death Rates (SDRs) by group of disease defined by International Classification of Diseases (ICD) codes. Morbidity-based indicators describe the general health status of a population and incidence of diseases. Due to the significant amount of unreported data regarding these indicators, we resorted to hospital discharges by disease type as a proxy to health status (incidence and prevalence data was incomplete for a large number of countries). Finally, we also include the available sample of healthcare utilisation statistics, measured in terms of bed occupancy rate, inpatient care discharges and average length of stay.

On the supply-side, the selection includes indicators that describe both the physical and human resources available. Physical resources are both the hospitals and their capacity, measured in terms of number of beds available. HHR are accounted for by the number of health professionals (we consider only physicians, nurses, dentists and midwives).

2.3. Data treatment, standardisation and aggregation

The HFA-DB reports values to 2013. In some instances, data from 2013 were unavailable. In such cases, one of three procedures was followed: (1) if a clear linear trend could be found, a linear regression was run to estimate the missing value; (2) if no trend line exists, data from previous years were used up to three years back; (3) if a considerable amount of countries did not report the data, the indicator was excluded from the analysis.

Thereafter, we run a multicollinearity bivariate analysis to identify correlated indicators. If no action is taken to fix multicollinearity, an overrepresentation of a particular dimension may occur, thereby potentially biasing the cluster formation (Ketchen and Shook, 1996). The literature reports significantly different upper bounds for the correlation coefficient (Hair *et. al,* 2013). Since we do not have a large enough sample to accurately estimate the correlation coefficients, we define the cutting point for indicators exhibiting a correlation coefficient equal to or greater than 0.85. Statistical significance was set to a global level of $p \leq 0.05$ (2-tailed distribution). To select the indicator to retain, we take one of two possible actions: if one indicator encompasses the other, we select the most complete one; otherwise, we select the indicator that best differentiates countries by choosing the one with the highest coefficient of variation.

Data standardisation was applied on a case-by-case basis. In some instances, maintaining the absolute difference between indicators was intentional and useful. For example, it is relevant to retain the difference between the numbers of SDRs caused by different diseases, as the impact on the healthcare system will be notoriously different (although it may not be in a linear way as different diseases put different levels of stress on the system). In contrast, indicators using different measurement units require scaling to remove the effect of different scales. For instance, the number of hospitals and the number of hospital beds cannot be compared directly. To remove this effect a scale change is used.

Finally, and whenever appropriate, data were aggregated by summing the indicators. For instance, the SDRs caused by each group of diseases were summed to a grand total. Similarly, the number of hospital dispatches by group of diseases was also added up. This allows for a direct comparison between mortality and morbidity levels between groups of countries. The methodology adopted is summarised in Figure 1.

2.4. Clustering algorithm

We employ a two-stage clustering algorithm to group similar countries in each indicator. The first stage is exploratory. We apply an agglomerative Hierarchical Clustering Algorithm (HCA) with Ward's method (see Everitt *et. al*, 2001) with a squared Euclidean distance to analyse possible cluster formations and then decide on the number of clusters. The second stage is explanatory. We use the cluster centres defined previously using HCA with Ward's method, and run a k-means algorithm to obtain descriptive statistics about the clusters, including an ANOVA table detailing which indicators were more relevant in grouping the countries.

In more detail, the first stage is an iterative process. Initially, each country is part of one single cluster. An agglomerative HCA procedure is then applied for merging clusters. Very similar clusters are combined. Similarity is measured in terms of distance, and there are several ways of calculating this distance. Single linkage calculates the shortest distance between any two members in the two clusters. Complete linkage looks for the longest distance between any two members. Average linkage and centroid are also common approaches, but since the dataset does not include outliers, we resort to Ward's method. In this procedure, the single clusters are merged if the merger results in the minimum merging cost, with merging cost being defined as the increase in the sum of squared errors. This procedure favours the formation of a cluster with very similar members. The result of this process is a distance matrix between clusters that can be used to draw a dendrogram. Dendrograms are tree diagrams used to illustrate the arrangement of clusters. The size of the branches is a meaningful measure, depicting the distance between clusters, and the further apart, the greater the heterogeneity between clusters. A decisive step in the clustering process is defining the *right* number of clusters. Again, the dendrogram provides relevant information to assist in this decision. Optionally, a scree plot can be used, which is in fact another way of representing the same information, exposing the distance within clusters as a new cluster is added. We are interested in a distinct break (elbow), after which the creation of one additional cluster does not create a significant distinction.

The second stage consists of running a k-means partitioning method based on the number of clusters and cluster centres previously obtained. K-means is a non-hierarchical procedure that does not require calculating distances. The aim of the procedure is to obtain descriptive statistics to assist in explaining the cluster formations. K-means will form k clusters using the cluster centres defined previously using the scree plot, by finding the point where the inclusion of an additional cluster does not significantly increase heterogeneity. This procedure will then originate the final cluster formations, and also an ANOVA table with F-tests for each indicator.

2.5. Validation and interpretation

The final step consists of validating and interpreting the clusters. External validation comprises of comparing the clusters with the true partition, which in this case cannot be known a priori. Internal validation, on the other hand, deals with the intrinsic properties of the dataset regardless of any external information, and is captured through the descriptive statistical information provided by k-means. To make this process more robust, we also visually inspect the correlations of the two most significant indicators.

To interpret the results we need to understand what binds members together but also what separates them. To do so, we examine the cluster centroids, which are the clustering variables' average values for all countries in a given cluster. To help understand which indicators maximise intra-cluster similarity, we resort to a one-way ANOVA table that calculates F-tests for each variable. We want to test if the clustering variables' means differ significantly across at least two of the *k* segments (where *k* is the number of clusters selected). Unless the null hypothesis is rejected, the indicator was relevant in the cluster formation. Besides validating the procedure, this is also a way of understanding the results obtained.

3. RESULTS

In this section we apply the two-stage clustering algorithm to the five groups of indicators, namely: mortality-based, morbidity-based, utilisation, physical and human resources indicators. With the clusters formed based on each set of indicators, we validate and analyse the results obtained. Each step of the methodology followed is described thoroughly.

3.1. Demand-based indicators

Demand for healthcare services can be assessed in two conceptually different ways (Amorim Lopes *et al.*, 2015). Effective demand measures demand for care effectively observed. Utilisation indicators are commonly used as proxies to estimate effective demand. Alternatively, medical needs are assessed based on the epidemiological conditions of the population, and then translated into a given quantity of healthcare services necessary to meet those needs (for an example see Harper *et al.*, 2010). Since not all the needs may turn into actual demand due to intentional or unintentional waiting lists, this is also referred to as potential demand. These concepts need not to be used separately, and can actually be combined. Utilisation indicators provide current usage levels, whereas health and disease patterns of the population may reflect unmet or future care needs. This is especially relevant in countries with extensive waiting lists, where healthcare services are being delayed due to lack of capacity.

HFA-DB provides both indicators for measuring potential demand (morbidity-based indicators) and for measuring effective demand (mortality-based and utilisation indicators). Our analysis targets both types of indicators. Note that these indicators are also commonly used to measure the level of output of health systems (Varabyova and Schreyögg, 2013).

3.1.1. Mortality

3.1.1.1. Selection of indicators

Concerning mortality-based indicators, HFA-DB contains both actual death rates per age cohort, and SDRs that cross-reference the medical cause of death. We have focused exclusively on the disease-specific mortality indicators rather than on the crude death statistics, as the former carry more explanatory power. There were no missing values in the data collected, therefore no estimation technique had to be applied.

3.1.1.2. Data treatment, standardisation and aggregation

Regarding collinearity, and considering that the list of mortality-based indicators contains both groups and subgroups of diseases, and groups of diseases aggregate the data of subgroups, it follows that indicators belonging to the same taxonomical group will naturally exhibit a strong correlation. For instance, the indicator that reports deaths caused by diabetes, "SDR, diabetes", will most probably correlate to the indicator that encompasses this type of disease, "SDR, endocrine, nutritional and metabolic diseases. Likewise, "SDR, cancer of the cervix uteri" will probably correlate to with a general indicator on the incidence of cancer, "SDR, malignant neoplasms." We have addressed this potential source of multicollinearity by choosing only the top-level indicators that represent the groups and already account for the subgroups.

Notwithstanding this case, multicollinearity may still occur. To mitigate this, we also ran a bivariate analysis on the pre-selected list of SDRs. As expected, risk-factor indicators such as "SDR, selected alcohol-related causes" and "SDR, selected smoking-related causes" are highly correlated to other SDRs, thus were removed. For the remaining variables, and given that a degree of correlation between diseases is in fact a standard medical occurrence (Jeon *et al.*, 2005; Li *et al.*, 2014), we will set the absolute correlation threshold to a value close to the upper bound typically found in the literature (Dormann *et al.*, 2013). For this case, none were removed. After applying the aforementioned procedure we obtain the list of indicators shown in Table 1.

Since SDRs are already partly standardised by age distribution and by one hundred thousand people, no standardisation techniques had to be applied. Moreover, scaling also will not be applied, since the SDR indicators all report the number of deaths, and absolute differences between indicators are relevant, representing the relative impact each disease may put on the health system.

3.1.1.3. Clustering algorithm

Next, we apply HCA using Ward's method. In Figure 2 we present the resulting dendrogram. In this particular case, it is immediate to see a notorious separation between two large groups. One side features Western Europe countries; the other features countries from Eastern Europe, and the data translate this significant difference. On average, Eastern European countries perform worse in all mortality indicators, an observation that is very significant.

Of the two initial groups, further separation can be found one level down the tree. Although not as significant as before (note the horizontal distance between the branches), it is still relevant. The two major clusters composed of Eastern European and Western European countries exhibit the largest distance. Nevertheless, given the high level of heterogeneity within these two clusters, a better grouping can be obtained if we further split into two more clusters, thus increasing the homogeneity within the cluster and increasing the distance between clusters. Henceforth, introducing additional clusters marginally increases the distance, implying that most heterogeneity has already been explored. The scree plot, represented in Figure 3, reinforces this indication.

With a pre-established number of clusters, we can now apply K-means clustering with the centroids (cluster centres) obtained with the HCA, and use the ANOVA statistical information to understand which indicators were most relevant in maximising intra-group similarity while minimising inter-group similarity. Running k-means for k=4, we obtain four final clusters (Table 2).

3.1.1.4. Validation and interpretation

Table 3 provides detailed information on the statistical significance tests for each variable. With a significance level below 0.001, "SDR, diseases of circulatory systems" had the largest impact on the definition of cluster centroids, followed by "SDR, diseases of digestive system", and "SDR, malignant neoplasms". With a smaller significance but still below the significance threshold, "SDR, motor vehicle traffic accidents" and "SDR, mental disorders, diseases of nervous system and sense organs" were also relevant when generating the clusters.

After determining the indicators that most contributed to form the cluster centres, it is possible to obtain several indicators in a single scatter plot, such as the most relevant SDRs, clusters and countries. Figure 4 is then useful to understand with visual guidance how the clusters were formed. Distance between points reflects intra-member dissimilarity, which is common to most clustering algorithms, including Ward's. In fact, the more homogeneous the members are, the most likely is the chance of them featuring together in a single cluster.

The cleavage between the two main groups featuring Western and Eastern European countries, quite evident in the dendrogram of Figure 2, is then explained by "SDR, diseases of circulatory system," for which the absolute difference is blatant. The other indicators were subsequently used to extract further heterogeneity within these two main groups. In fact, Eastern European countries feature, on average, twice the mean value in the number of deaths caused by this group of diseases. These countries are also the worst performers in the number of deaths due to diseases of the digestive system and to malignant neoplasms. In contrast, the Southern and some Nordic countries in cluster 3 have the best performance, registering a considerably lower number of deaths, in some cases half the total average.

Finally, we resort to a simple visual tool for quickly inspecting the quality and validity of the results obtained. We plot the two most significant indicators against each other and use the cluster cases as labels. Heterogeneous (badly formed) clusters are featured in a dispersed, uncorrelated way. Homogeneous groups, on the other hand, can be clearly identified. Figure 5 depicts this.

3.1.2. Morbidity

3.1.2.1. Selection of indicators

The HFA-DB contains three distinct types of morbidity indicators: incidence of diseases, prevalence of diseases, and hospital discharges per disease. Each group of indicators portrays different information. Since a lot of missing blanks exist for incidence and prevalence of diseases, we will only consider the hospital discharges as this type of indicator is commonly used to measure system output (Varabyova and Schreyögg, 2013). The list of hospital discharge indicators can be found in Table 1. Similarly to the case of mortality-based indicators, only groups of diseases were considered.

3.1.2.2. Data treatment, standardisation and aggregation

In this particular case, it was verified that some indicators for 2013 were missing, which needs to be handled adequately, since cluster analysis cannot be conducted if data are missing. We have decided to use the last available year as not enough data was available to run a statistically significant regression.

Subsequently, the process followed to obtain the clusters using the morbidity-based indicators is very similar to the one previously applied to mortality rates. A multicollinearity analysis needs to be conducted in order to identify correlations between indicators that could potentially boost the importance of a particular category of indicators, thereby biasing the cluster formation. None have surpassed the threshold of 0.85 below the significance threshold, and therefore no indicators have been removed. As before, no standardisation or transformation technique will be applied as hospital discharges by disease are reported in ratios of 100 000 people, and the idea is to preserve the relative differences between diseases.

3.1.2.3. Clustering algorithm

Running an HCA on the morbidity-indicators we obtain a dendrogram to assist with formation of the clusters. The tree obtained is more complex to interpret than the previous one (see Figure Figure 6), suggesting that more indicators were used to calculate dissimilarity distances within and between clusters. Firstly, there is a clear separation between the first group of 8 countries and all the remaining ones. A quick look at the data reveals that these countries have, on average, the lowest number of hospital discharges registered. All the remaining groups exhibit higher hospital discharges, with Austria and Germany leading the chart with the highest number. The further expansion of the nodes suggests that further differences can be found within countries.

Before going deeper into the analysis, we again resort to a scree plot for determining the cutting point and defining the number of clusters (Figure 6). Contrarily to the previous case, there is no sharp elbow, but it is noticeable that after the sixth cluster the distance that can be increased is significantly reduced. We therefore define the number of clusters to six. It should be kept in mind that there is a trade-off in the aggregation process, where the distance is at its lowest value only when each country belongs to a single and unique cluster, but defeating altogether the purpose of the aggregation.

After fixing the number of clusters to six, we run the k-means algorithm to obtain a detailed description of the most relevant indicators to form the clusters (Table 2). The first thing we notice is that the algorithm was unable to group France in any cluster, implying that its values are so unique that the country is better classified as an outlier. Secondly, a clear geographical segregation between Eastern and Western countries is not as evident as with the mortality-based indicators, although it exists to a certain degree.

3.1.2.4. Validation and interpretation

The analysis of the F-tests helps clarify on the relative importance of each variable when forming the clusters (cf. Table 3). Contrarily to the clusters generated with the mortality-based indicators, all the variables were used to form the groups, as they all were statistically significant at a level below 0.001. With the highest F-test score, “2520 Hospital discharges, digestive system diseases” was the most important indicator, followed by hospital discharges related to diseases in the circulatory system. Expectably, these indicators exert a large influence since they represent the largest amount of hospital discharges. In contrast, infectious and parasitic diseases put less stress on the healthcare system when measured only in terms of number of persons discharged, and so it was less critical when generating the clusters. Note that treatments to diseases differ in the amount of human and physical resources required, and so the number of hospital discharges is an incomplete proxy to the stress put on the healthcare system. Although this has no impact on the methodological approach, it may have on conclusions drawn from the results. For that reason, it would be beneficial to have a proper balancing taking this into account.

With this information at hand, we can now proceed with a more precise analysis of the clusters. Firstly, cluster 4 has, on average, the lowest number of hospital discharges for all diseases. In sharp contrast, cluster 1, composed of Austria and Germany, has the highest number of discharges, followed by cluster 3, which is mostly composed of Eastern European countries. France has the highest record in some of the indicators (e.g. “2520 Hospital discharges, digestive system diseases”, where it has twice the average of discharges), but average or below average values in others. For this reason, it did not fit any of the clusters and was thus put into a single cluster (cluster 6). Finally, clusters 2 and 5 have similar number of dispatches, on average, although countries from cluster 2 have a lower number of dispatches. Overall, cluster 4 has the smallest number of hospital discharges and cluster 1 has the highest. Note that using only this indicator makes it impossible to ascertain whether this is a result of inefficiency or low incidence. This information is summarised in Table 4.

3.1.3. Utilisation

While mortality- and morbidity-based indicators provide us with an insight of the general healthcare *needs* of a population, which may or may not translate into effectively observed demand (Amorim Lopes *et al.*, 2015)), utilisation-based indicators translate actual utilisation ratios of the healthcare facilities, such as hospitals or primary care centres. Health needs may not always translate into actual demand due to several impediments such as financial constraints arising from expensive out-of-pocket treatments, or due to long waiting lists resulting from inefficient healthcare systems. Either way, there may exist unmet needs that do not translate into actual demand.

3.1.3.1. Selection of indicators

The procedure followed does not differ significantly from the method previously used. We start by selecting the variables to be used. The set of available variables is limited, essentially reporting overall inpatient care discharges, bed occupancy rates, average length of stay and number of outpatient contacts per year. Table 1 lists all the indicators used.

3.1.3.2. Data treatment, standardisation and aggregation

Note that the indicators are reported in different units, and so variable standardisation is required to remove this potentially misleading effect. We map the indicators to a 0 to 1 range, preserving relative differences between countries. We also run a multicollinearity analysis, in the end removing "Acute care hospital discharges" as these discharges are already accounted for in the more general inpatient care discharges.

3.1.3.3. Clustering algorithm

It is then possible to run the HCA and obtain a preliminary grouping. Complementing the dendrogram with a scree plot analysis (see Figure 8 and 9), we find no clear elbow where to define the cutting point. Nevertheless, and considering that decreases in the distance coefficient get smaller between the seventh and the ninth cluster, after which they are almost marginal, we fix the number of clusters to seven. Running a k-means cluster analysis with k=7, we obtain the final cluster formation reported in Table 2.

3.1.3.4. Validation and interpretation

Again, the ANOVA table helps to explain and validate the cluster formations (Table 3). According to the F-tests, both "Outpatient contacts" and "Inpatient care discharges" were the most critical indicators for grouping the countries. Following the results already obtained for the morbidity indicators, Austria and Germany, and also Lithuania, exhibit the highest number of inpatient care discharges. As for the number of outpatient contacts per year, cluster 3, composed of Czech Republic, Hungary and Slovakia, tops the list. On the opposite side, countries from cluster 4 have the lowest number of outpatient contacts per person, and, therefore, a small percentage of ambulatory care. In contrast, this group of countries exhibits, on average, the highest rate of bed occupancy (in acute care hospitals). Cluster 5, on the other hand, has both a relatively low bed occupancy rate, and a low number of outpatient contacts. However, on average, it has the highest average length of stay. The Netherlands did not fit any of the pre-existing clusters due to its extremely low bed occupancy rate and low inpatient care discharges. Interestingly, Southern countries from cluster 6 (Portugal, Italy and Spain) exhibit the lowest number of inpatient care discharges and below-average bed occupancy rates.

In summary, these results help to divide the clusters as follows: countries with health systems more orientated towards inpatient care and with average outpatient contacts (clusters 1 and, to some extent, 2); and countries with a model devised towards outpatient contacts (cluster 3). We can also identify countries with high occupancy rates, average inpatient care discharges and low outpatient contacts (cluster 4), and countries with average to low inpatient care discharges but long stays at the hospital and high bed occupancy rates (clusters 5 and 6).

3.2. Supply-based indicators

3.2.1. Physical resources

Indicators reporting the available physical resources can be used in several ways. Firstly, when used together with the utilisation-based indicators, these indicators help to identify countries in a situation of under or overcapacity of its physical resources. Secondly, they can shed a light on the capital intensity of the health system, that is, whether it is labour-intensive or capital-intensive. Finally, they can be used to identify countries with a similar infrastructure. This is highly relevant since infrastructure investment is a long-term decision that cannot be taken lightly, and therefore policy proposals that compare asymmetric countries without taking this into account will fail to provide realistic (applicable) suggestions.

3.2.1.1. Selection of indicators

The indicators used to perform the analysis are identified in Table 1. Data reported either relates to the number of hospitals or hospital beds by specialty.

3.2.1.2. Data treatment, standardisation and aggregation

The multicollinearity analysis identified a strong correlation between “Hospital beds” and “Acute care hospital beds”, which is expectable since the first indicator already encompasses the second. Therefore, it has been removed. Also, rescaling was applied to remove the effect introduced by the usage of different scales.

3.2.1.3. Clustering algorithm

Running HCA followed by a scree plot analysis points to a cutting-point at 4, optionally at 5 clusters (cf. Figure 10 and 11). Fixing the number of clusters to 5 and running a k-means cluster analysis originates the cluster formations reported in Table 2.

Exhibiting high statistical significance, the number of hospitals, reported both through the indicators "Hospitals" and "Acute (short-stay) hospitals", played a significant role in forming the clusters. In fact, cluster 3 includes the countries with the highest number of hospitals, both for short and long stays. Cluster 4, composed of countries like Estonia, Latvia and Switzerland (cf. Table 2), is similar, but with a lower number of short-stay hospitals and hospital beds. Cluster 5, composed of Southern and some Nordic countries, exhibits some of the lowest numbers in terms of hospitals and hospital beds. Note that the capacity to deliver does not imply more health gains, as more resources do not necessarily amount to more healthcare services provided. In fact, it may actually mean that there is a suboptimal resource allocation, with too many (or too few) expensive physical resources. To clarify, we again resort to a brief analysis of the capital-intensity to identify low and high capital intensity in the health systems. Note that the indicators are already normalised by population size, hence removing the difference in that size. Results are reported in Table 4.

3.2.1.4. Validation and interpretation

According to the ANOVA F-test results provided in Table 3, all indicators bar the number of psychiatric hospital beds were highly significant in determining differences between health systems, and therefore in defining the clusters. The clusters formed reflect four different healthcare architectures: a low number of hospitals and hospital beds (cluster 5); a large number of hospitals but with few hospital beds (cluster 4); a small to average number of hospitals but with large amount of hospital beds (clusters 1 and 2); and health systems with both a high number of hospitals and hospital beds (cluster 3). Geographical distance does not explain these clusters, as both eastern and western countries, Nordic and southern countries are intermingled in different clusters. This fact is suggestive of an explicit choice in terms of the organizational model.

3.2.2. Human resources

3.2.2.1. Selection of indicators

The indicators used to perform the analysis are identified in Table 1. Human resources are measured in terms of number of physicians, nurses, dentists and midwives employed, regardless of their specialisation, if any.

3.2.2.2.Data treatment, standardisation and aggregation

Since "General Practitioners" is a subcategory of "Physicians", we anticipate a multicollinearity issue, which is confirmed by running the tests. We then discarded this indicator from the analysis as it was already incorporated in the parent set.

Since all indicators are reported in a pre-standardised way (number of professionals per 100 000 inhabitants), no further standardisation was required.

3.2.2.3.Clustering algorithm

We repeat the exact same procedure, this time applied to the indicators portraying the availability of human resources (cf. Table 2). We ran HCA followed by a coefficient analysis through a scree plot to decide on the number of clusters. With the help of both the dendrogram and the scree plot (Figure 10 and 11) we fix this number to seven in order to run k-means and obtain the final cluster composition.

3.2.2.4.Validation and interpretation

The resulting clusters are reported in Table 2, and the ANOVA F-test statistics are reported in Table 3. The statistical significance tests seem to suggest that no particular indicator had a distinctive influence in the cluster formations. Despite this, the number of nurses was the most significant. Regarding how the countries were grouped, of the seven clusters formed, it is immediate to see that Greece is a clear outlier. It tops the chart in both the number of physicians and dentists, having twice the average number, while at the same time it is the country with the lowest number of nurses. With also a significant amount of physicians but a not so high number of nurses, although well below Greece, is cluster 1, composed of countries like Austria, Germany (again, these two countries are part of the same group), Lithuania, Italy or Portugal. In contrast, cluster 2, composed of countries like Luxembourg, Belgium, Finland and Norway, has the exact opposite characteristics: a relatively small number of physicians and a quite large number of nurses and midwives. Cluster 4, composed of Switzerland and Denmark, also demonstrates an extremely high nurse-to-physician ratio, but with a considerably higher ratio of physicians, therefore not featuring on cluster 2.

We can then characterize some of the clusters according to their healthcare delivery model: countries with a low number of physicians and a low number of nurses (cluster 7); countries featuring a high number of physicians but a low number of nurses (cluster 1);

countries featuring a low number of physicians but a high number of nurses (clusters 2 and 4); countries with both a high number of physicians and nurses (cluster 3).

4. Discussion

In this article, we had several objectives: (i) warn against the methodological flaws potentially posed by the usage of benchmarks or simple ratio comparisons drawn between severely different health systems, subject to disparate health patterns, asymmetric healthcare infrastructure, and several other contrasting characteristics; (ii) propose a methodology to group countries according to their similarities in each health indicator, providing clusters of countries that share similar characteristics so that comparisons can be made in a (methodologically) safer way; (iii) discern the profiles of the groups of countries for each indicator, providing preliminary insights and a similarity matrix ranking the most and the least similar countries. With this theoretical and empirical contribution, we aim to provide the underpinnings for more accurate comparisons between health systems, upon which performance analyses or benchmarking can be conducted. To some degree, the proposed approach avoids error-prone comparisons with the global average.

4.1. Relative performance analyses

Applying the methodology to the indicators provided by the OECD's HFA-DB provides some interesting insights that benefit from further discussion. After the clusters are generated, we perform an inter-cluster qualitative comparison and characterise the groups obtained in terms of relative performance. To do so, we start by calculating the average of the cluster for each variable considered, and then label countries as below average exhibiting a low performance, or above average exhibiting a high performance (note that *low* does not mean "bad" or "inefficient", but only that the measurement is lower comparatively to the average, and vice-versa). In terms of mortality, we then establish that, in relative terms, clusters 1 and 3 exhibit low mortality rates, while clusters 2 and 4 exhibit high mortality rates. Unsurprisingly, this reflects the Western/Eastern asymmetries mentioned previously, with former Soviet countries still trying to catch up in terms of healthcare gains in comparison with other health systems. In fact, this sharp difference is quite noticeable, with the worst performing clusters registering twice the average of SDRs.

Contrarily to mortality-based indicators, it would be fallacious to draw conclusions from the qualitative assessment on the relative performance of the clusters solely based on the number of dispatches, the proxy selected to characterise morbidity. Exhibiting a high or low number of hospital discharges does not characterise the system. The health system can be either highly efficient and effective because it can treat a large number of patients, or simply because the incidence and prevalence rates of those diseases are lower in some countries, which would naturally lead to less hospital entries, and hence less discharges. Nevertheless, this analysis provides a first insight into the comparative performance of the groups. We see that high numbers of hospital discharges are not a characteristic unique to less developed healthcare systems. In fact, the cluster composed of Austria and Germany report, on average, the highest number of hospital discharges.

4.2. Characterising demand

As previously discussed, a simple quantitative analysis of the morbidity rates measured in terms of hospital discharges would be misleading. High hospital discharges may not necessarily suggest that the health system is efficient. But if we cross the morbidity-based hospital discharges with the mortality standardised death rates, we can better characterise the potential demand (*needs*) for healthcare. A priori, it can be conjectured that countries exhibiting high mortality rates and low hospital discharges are not providing appropriate care. Likewise, countries exhibiting low mortality rates and high hospital discharges would suggest a high amount of care. Such information can be useful for characterising the stress that different healthcare systems have to endure, and therefore enhance the robustness of future comparisons.

In Figure 12 shows how countries perform when both mortality-based indicators and morbidity-based indicators are considered. We also group those countries that were part of the same group both in the mortality- and morbidity-based cluster analysis. Despite reducing the number of countries in the clusters, this meta-clustering technique strengthens the cluster formations bringing together countries that are similar in more than one set of indicators. The plot is divided into four quadrants. Countries exhibiting both low (high) hospital discharges and low (high) mortality rates are of no particular interest, as it is expectable that if more people carry a disease, more hospital discharges and more mortality will follow, and vice-versa. The revealing cases are those countries that have either a high number of hospital discharges but a low number of death rates, suggesting that these countries are making efforts for containing the diseases and treating patients, or a low number of hospital discharges and high mortality rates, implying that a considerable number of people are dying without proper treatment or inadequate care is being provided. The former is the case of Austria and Germany, which have a comparatively low mortality rate despite the large number of hospital discharges. Latvia and Slovakia are on the opposite side, as they have an extremely high mortality rates with a low number of hospital discharges, which can be interpreted as incapacity to deliver effective healthcare, or an healthcare system based on an outpatient contact approach that is not generating satisfactory results in terms of health gains.

With this information, we can form clusters of countries featuring similar morbidity trends, when measured in terms of outputs (hospital discharges), and effectiveness of the care services delivered, when measured in terms of mortality rates (Table 5).

With these super-clusters, it is possible to make comparisons at two different levels, between and within groups. For instance, if Romania would like to improve its death accruals, they should probably look to Hungary for guidance, as it has the same number of hospital discharges, and yet lower death records. Similarly, if Portugal wants to improve the death records, it should look to Spain or Italy, rather than the EU average or any other country outside its cluster. Comparisons between clusters are also possible using a relative performance index. In particular, clusters 1 and 5 in Table 5 both exhibit low levels of mortality. However, contrarily to Austria or Germany, countries from cluster 5 register significantly lower hospital discharges, implying that for each person discharged, holding everything else constant, more are dying due to a disease. This may be indicative of lack of proper or timely treatment.

4.3. Characterising supply

Output in terms of healthcare services is the result of the production using input factors needed to deliver care, both physical (capital) and human (labour). Physical resources encompass hospitals, hospital beds, screening and treatment technology, or drugs. We have focused only on the indicators provided in the HFA-DB, which include the number of hospitals and hospital beds. Human resources include physicians, nurses, midwives, dentists, medical assistants, etc. Only those present in the HFA-DB and with no significant missing information were included in the analysis.

The data available on the health human resources in each country seem to suggest a very different approach to healthcare delivery, focused essentially in the role of the physician in the first case, and on a higher delegation of tasks to the nursing profession in the second case[1]. To illustrate this, we draw a two-dimensional graph depicting health human resources (Figure 13). As mentioned before, Greece is a clear outlier, with a number of physicians well above the EU average, while having the lowest record of nurses. This is also visible in countries such as Austria, Italy or Lithuania. On the other extreme in terms of healthcare model are countries such Belgium, Denmark, Norway, Luxembourg, Finland or Switzerland, which register a high number of nurses but a low number of physicians. It is interesting to note that this trend is common to Northern European countries, apparently favouring a healthcare workforce with a high prevalence of nurses.

Another conclusive analysis consists of plotting both human and physical resources, which makes it possible to obtain a degree of capital and labour intensity of the healthcare system in each country. To assess labour intensity, we considered an unweighted sum of the number of physicians and nurses. Without disregarding the importance of other clinical actors, these are the core human resources of any healthcare system. As for capital intensity, we consider only the number of hospital beds. The scatter plot in Figure 14 provides visual guidance to understand these health models.

The most interesting cases are those countries exhibiting asymmetries between capital and labour, implying a radically different approach to healthcare delivery. Interestingly, rich countries such as Denmark, Norway or Switzerland appear to be using plenty of human resources, especially nurses, while keeping the number of hospital beds relatively low. A tentative explanation could lie in the fact that these countries are able to treat patients more quickly, and hence free physical resources (hospital beds). Considering that in our case capital intensity portrays the number of hospital beds and not the entire set of technologies used to treat patients, this is the only hypothesis standing. On the opposite side are the countries employing a significant amount of capital with a low labour intensity. Almost all Eastern European countries exhibit this trend. This apparently high availability of

[1] Alternatively, this may be the result of disparities in the way the number of nurses are reported to WHO, since in some countries some auxiliary professions count as nursing.

resources may help with a surge in demand, but it says nothing about the capacity to deliver effective healthcare services. While a physician or a nurse may always deliver services in better or worse conditions regardless of the infrastructure available, that is not the case with an empty bed. Either way, validating any of the hypotheses would require a thorough and methodologically rigorous approach, outside the scope of this work.

4.4. Similarity matrix

Finally, and gathering all the clusters obtained for each of the indicators, we can construct a similarity/dissimilarity matrix, pinpointing countries that are part of the same group, and hence similar to each other, in a given number of indicators. Figure 15 shows that while some countries are similar to each other in several dimensions, others differ significantly. For instance, Portugal and Italy are part of the same cluster for five out of five indicators. With a similarity of four, countries like Portugal and Spain, Hungary and Romania, Belgium and Luxembourg, or Austria and Germany, feature in the same cluster in four out of the five indicators. Also important to international comparisons may be to identify countries that share no common ground at all. Although this may be due to different stages of development, which appears to be the case, for instance, of Austria and Hungary, it may also reflect substantially different approaches to healthcare delivery. For instance, Austria and Ireland, countries with an almost identical GDP per capita, are extremely different, and so they do not feature in the same cluster in any of the indicators. Note that due to the lack of data, it was not possible to include all countries for all the indicators, which means that this matrix may be underrepresented.

5. Conclusion

In this paper we have proposed a new approach to cross-country comparisons, which avoids the pitfalls of benchmarking against the average, or of absolute rankings that sometimes lead to very arguable conclusions, or even erroneous policy choices. Instead of finding winners and losers, this methodology establishes similarities between countries. Performance and efficiency analyses can still be conducted, but at a different granularity level. From a policy perspective, it is easier to suggest a reform based on the experience of a country that shares common ground, but still excels in comparison with the benchmarked, rather than on a stalwart and yet structurally different country.

Our methodology consists in applying state-of-the-art clustering techniques to group similar countries according to several indicators. We applied this methodology to WHO's HFA-DB, a comprehensive database containing indicators related to mortality, morbidity, utilisation, and physical and human resource indicators. We were able to group countries in each of the five dimensions, and explain the cluster formations. Moreover, and based on these results, we established a similarity matrix. The results obtained and discussed serve the primary purpose of demonstrating the methodology, but also allow for a preliminary comparative analysis of the health systems. In the future, we expect this methodology to be applied to a more comprehensive set of data so that more insightful and robust conclusions can be drawn to guide policy reforms.

Another interesting application of this methodology is crossing groups of indicators. From the demand analysis, we highlighted that Hungary and Romania were part of the same cluster, but Hungary featured a considerably smaller mortality rate within the group when compared to Romania. If we add the dimension of supply, characterized by both labour and capital, we notice both a higher capital and labour intensity in comparison to Romania. Although a more rigorous econometric analysis that controls for other factors would be needed, this seems to suggest that improved health results may be due to more physical and human resources. This result can also be seen between Portugal and Italy, also part of the same group in both dimensions. Italy has a smaller mortality rate than Portugal, but also more resources to deliver healthcare, both in terms of physicians and nurses, but also in hospital beds. But in comparison with Spain, Portugal exhibits a higher mortality rate compared with Spain, despite employing more capital and more labour. Given the similarities and the results achieved, Spain may be reference to guide future policy decisions in Portugal.

Finally, it would be also interesting to apply a nonparametric tool like DEA to obtain the efficiency frontier within clusters. Across the board DEA is also possible, but it would defeat altogether the purpose of this methodology, which is to benchmark against countries with similar health systems, in this way ensuring that emerging policy actions are applicable.

6. References

Amorim Lopes M, Almeida A, Almada-Lobo B. (2015), “Handling healthcare workforce planning with care: where do we stand?”, *Human Resources for Health*, 13:38. DOI 10.1186/s12960-015-0028-0.

Auerbach DI, Kellermann AL (2011), “A Decade Of Health Care Cost Growth Has Wiped Out Real Income Gains For An Average US Family”, *Health Affairs*, 30:1630–6.

Bhat VN (2005) “Institutional arrangements and efficiency of health care delivery systems”, *Eur J Health Econ*, 6:215–22.

Birch S. (2002), “Health human resource planning for the new millennium: inputs in the production of health, illness, and recovery in populations”, *Can J Nurs Res*, 33:109–14.

Bloor K, Hendry V, Maynard A. (2006), “Do we need more doctors?” *J R Soc Med*, 99:281–7.

Bloor K, Maynard A. (2003) “Approaches to human resources in health: an international comparative review”, *Canadian Health Services Research Foundation*, 1–34.

Bronnum-Hansen H. (2014) “Ranking health between countries in international comparisons”, *Scandinavian Journal of Public Health*, 42:242–4.

Burau V, Blank RH. (2006), “Comparing health policy: An assessment of typologies of health systems”, *Journal of Comparative Policy Analysis: Research and Practice*, 8:63–76.

Chandra A, Holmes J, Skinner J. (2013), *Is This Time Different? The Slowdown in Healthcare Spending*.

Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G, et al. (2013) “Collinearity: a review of methods to deal with it and a simulation study evaluating their performance”, *Ecography*, 36:27–46.

Dreesch N. (2005), “An approach to estimating human resource requirements to achieve the Millennium Development Goals”, *Health Policy and Planning*, 20:267–76.

Europe WROF (2015) *European health for all database*. Copenhagen, Denmark: WHO Regional Office.

Everitt BS, Landau S, Leese M. (2001) *Cluster Analysis*, Wiley.

Hair JF, Black WC, Babin BJ, Anderson RE. (2013) *Multivariate Data Analysis. 7 ed.* Pearson New International Edition.

Harper PR, Powell NH, Williams JE. (2010) “Modelling the size and skill-mix of hospital nursing teams”, *Journal of the Operational Research Society*, 61:768–79.

Jeon Y-D, Lee J-I, Cho B-O, Rotaru H, Kim J-C, Kim S-G. (2005) “Statistical correlation between pharyngitis and temporomandibular joint disease”, *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology*, 99:677–81.

Ketchen DJ, Shook CL. (1996) *The application of cluster analysis in strategic management research: an analysis and critique*. *vol. 17*. John Wiley & Sons, Ltd.

Li J-M, Liu C, Hu X, Cai Y, Ma C, Luo X-G, et al. (2014) “Inverse correlation between Alzheimer’s disease and cancer: implication for a strong impact of regenerative propensity on neurodegeneration?”, *BMC Neurology*, 14:211.

Pringle M, Wilson T, Grol R. (2002) “Measuring “goodness” in individuals and healthcare systems”, *BMJ*, 325:704–7.

Richardson J, Wildman J, Robertson IK. (2003) “A critique of the World Health Organisation's evaluation of health system performance”, *Health Economics*, 12:355–66.

Roberfroid D, Leonard C, Stordeur S. (2009) “Physician supply forecast: better than peering in a crystal ball?”, *Hum Resour Health*, 7:10.

Tchouaket ÉN, Lamarche PA, Goulet L, Contandriopoulos AP (2012) “Health care system performance of 27 OECD countries”, *Int J Health Plann Manage*, 27:104–29.

Varabyova Y, Schreyögg J. (2013) “International comparisons of the technical efficiency of the hospital sector: Panel data analysis of OECD countries using parametric and non-parametric approaches”, *Health Policy*, 112:70–9.

Wendt C. (2009), “Mapping European healthcare systems: a comparative analysis of financing, service provision and access to healthcare” *RszarfIpsUwEduPl*, 19:432–45.

Acknowledgements

This work is financed by the FCT – Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) within project UID/EEA/50014/2013.

7. Tables

MORTALITY-BASED INDICATORS			
Included			
Code	**Name**	**Target**	**Sample**
1320	SDR, diseases of circulatory system	all ages	per 100k
1520	SDR, malignant neoplasms	all ages	per 100k
1740	SDR, motor vehicle traffic accidents	all ages	per 100k
1820	SDR, infectious and parasitic diseases	all ages	per 100k
1830	SDR, diseases of respiratory system	all ages	per 100k
1850	SDR, diseases of digestive system	all ages	per 100k
1870	SDR, endocrine, nutritional and metabolic diseases	all ages	per 100k
1900	SDR, mental disorders, diseases of nervous system and sense organs	all ages	per 100k
1910	SDR, disease of genitourinary system	all ages	per 100k
1920	SDR, symptoms, signs and ill-defined conditions	all ages	per 100k
1960	SDR, acute respiratory infections, pneumonia and influenza	< 5 years	per 100k
Excluded			
1340	SDR, ischaemic heart disease	all ages	per 100k
1360	SDR, cerebrovascular diseases	all ages	per 100k
1540	SDR, trachea/bronchus/lung cancer	all ages	per 100k
1560	SDR, cancer of the cervix uteri	all ages	per 100k
1590	SDR, malignant neoplasm female breast	all ages	per 100k
1840	SDR, bronchitis/emphysema/asthma	all ages	per 100k
1860	SDR, chronic liver diseases and cirrhosis	all ages	per 100k
1880	SDR, diabetes	all ages	per 100k
1890	SDR, diseases of the blood, blood forming organs and certain immunity disorders	all ages	per 100k
1930	SDR, tuberculosis	all ages	per 100k
1940	SDR, diarrhoeal diseases	< 5 years	per 100k
1970	SDR, selected alcohol-related causes	all ages	per 100k
1980	SDR, selected smoking-related causes	all ages	per 100k

MORBIDITY-BASED INDICATORS			
Included			
2300	Hospital discharges, infectious and parasitic diseases	all ages	per 100k
2310	Hospital discharges, all neoplasms	all ages	per 100k
2450	Hospital discharges, circulatory system diseases	all ages	per 100k
2500	Hospital discharges, respiratory system diseases	all ages	per 100k
2520	Hospital discharges, digestive system diseases	all ages	per 100k
2530	Hospital discharges, musculoskeletal system and connective tissue diseases	all ages	per 100k
2540	Hospital discharges, injury and poisoning	all ages	per 100k
Excluded			
2460	Hospital discharges, ischaemic heart disease	all ages	per 100k
2480	Hospital discharges, cerebrovascular diseases	all ages	per 100k

UTILISATION INDICATORS			
Included			
6010	Inpatient care discharges	all ages	per 100
6100	Average length of stay, all hospitals	all ages	per 100k
6210	Bed occupancy rate (%), acute care hospitals only	all ages	% of total
6300	Outpatient contacts	all ages	per person per year
Excluded			
6020	Acute care hospital discharges	all ages	per 100

PHYSICAL RESOURCES INDICATORS			
Included			
Code	**Name**	**Target**	**Sample**
5010	Hospitals	-	per 100k
5050	Hospital beds	-	per 100k
5020	Acute (short-stay) hospitals	-	per 100k
5070	Psychiatric hospital beds	-	per 100k
5100	Nursing and elderly home beds	-	per 100k
Excluded			
5060	Acute care hospital beds hospitals	-	per 100k

HUMAN RESOURCES INDICATORS			
Included			
Code	**Name**	**Target**	**Sample**
5250	Physicians	-	per 100k
5300	Dentists (PP)		
5320	Nurses (PP)	-	per 100k
5350	Midwives (PP)	-	per 100k
Excluded			
5290	General practitioners (PP)	-	per 100k

Table 1 - List of pre-selected variables for each indicator, and variables excluded for failing to meet all of the inclusion criteria.

Indicators	Cluster						
	1	2	3	4	5	6	7
Mortality-based	Austria, Finland, Germany, Greece, Malta, Slovenia, Sweden	Croatia, Czech Republic, Estonia, Poland	Belgium, Denmark, France, Ireland, Italy, Luxembourg, Netherlands, Norway, Portugal, Spain, Switzerland, United Kingdom	Hungary, Latvia, Lithuania, Slovakia, Romania			
Morbidity-based	Austria, Germany	Belgium, Luxembourg, Norway, Slovenia, Sweden, Switzerland	Hungary, Lithuania, Romania	Croatia, Ireland, Italy, Malta, Netherlands, Portugal, Spain, United Kingdom	Czech Republic, Denmark, Estonia, Finland, Greece, Latvia, Poland, Slovakia	France	
Utilization	Austria, Germany, Lithuania	Belgium, Croatia, Estonia, Latvia, Luxembourg, Slovenia	Czech Republic, Hungary, Slovakia	Denmark, Greece, Ireland, Norway, Sweden, Switzerland, United Kingdom	Finland, France	Italy, Portugal, Spain	Netherlands
Physical resources	Austria, Czech Republic, Greece, Hungary, Luxembourg, Poland, Romania, Slovakia	Belgium, Croatia, Malta, Netherlands, Slovenia	Bulgaria, Finland, France, Germany, Lithuania	Estonia, Latvia, Switzerland	Italy, Norway, Portugal, Spain, Ireland, Sweden		
Human resources	Austria, Germany, Lithuania, Italy, Portugal	Luxembourg, Belgium, Finland, Norway	Czech Republic, Croatia, Bulgaria, France, Estonia	Denmark, Switzerland	Hungary, Romania, Netherlands, Slovenia, Latvia, Spain	Poland, Slovakia, Malta, United Kingdom	Greece

Table 2 - Cluster compositions obtained for each of the group of indicators.

ANOVA F-test results		
	F-test	Sig.
MORTALITY-BASED INDICATORS		
1320 SDR, diseases of circulatory system	183.844	.000
1520 SDR, malignant neoplasms	10.404	.000
1740 SDR, motor vehicle traffic accidents	5.073	.007
1820 SDR, infectious and parasitic diseases	1.959	.147
1830 SDR, diseases of respiratory system	2.836	.059
1850 SDR, diseases of digestive system	20.126	.000
1870 SDR, endocrine, nutritional and metabolic diseases	.331	.803
1900 SDR, mental disorders, diseases of nervous system and sense organs	3.699	.026
1910 SDR, disease of genitourinary system	.832	.490
1920 SDR, symptoms, signs and ill-defined conditions	.710	.555
1960 SDR, acute respiratory infections, pneumonia and influenza	2.163	.119
MORBIDITY-BASED INDICATORS		
2520 Hospital discharges, digestive system diseases	31.863	.000
2450 Hospital discharges, circulatory system diseases	28.524	.000
2310 Hospital discharges, all neoplasms	19.504	.000
2530 Hospital discharges, musculoskeletal system and connective tissue diseases	17.991	.000
2540 Hospital discharges, injury and poisoning	12.204	.000
2500 Hospital discharges, respiratory system diseases	11.550	.000
2300 Hospital discharges, infectious and parasitic diseases	6.841	.001
UTILISATION INDICATORS		
6010 Inpatient care discharges	17.747	.000
6100 Average length of stay	7.895	.000
6210 Bed occupancy rate (%), acute care hospitals only	8.855	.000
6300 Outpatient contacts	17.968	.000
HEALTHCARE PHYSICAL RESOURCES		
5010 Hospitals per 100 000	29.748	.000
5050 Hospital beds per 100 000	21.278	.000
5020 Acute (short-stay) hospitals per 100 000	24.331	.000
5070 Psychiatric hospital beds per 100 000	5.082	.005
HEALTHCARE HUMAN RESOURCES		
5250 Physicians per 100 000	11.458	.000
5300 Dentists (PP) per 100 000	11.566	.000
5320 Nurses (PP) per 100 000	14.521	.000
5350 Midwives (PP) per 100 000	12.516	.000

Table 3 - ANOVA F tests for the variables used to form the clusters. For descriptive purposes only as the clusters have been chosen to maximize the differences among cases in different clusters and the significance levels are not corrected for this.

Relative Status	Demand			Supply	
	Mortality-based indicators	Morbidity-based indicators (hospital dispatches)	Utilization	Human resources	Physical resources
Low	1 Austria, Finland, Germany, Greece, Malta, Slovenia, Sweden	2 Belgium, Luxembourg, Norway, Slovenia, Sweden, Switzerland	2 Belgium, Croatia, Estonia, Latvia, Luxembourg, Slovenia	3 Czech Republic, Croatia, Bulgaria, France, Estonia	2 Belgium, Croatia, Malta, Netherlands, Slovenia
	3 Belgium, Denmark, France, Ireland, Italy, Luxembourg, Netherlands, Norway, Portugal, Spain, Switzerland, United Kingdom	4 Croatia, Ireland, Italy, Malta, Netherlands, Portugal, Spain, United Kingdom	3 Czech Republic, Hungary, Slovakia	5 Hungary, Romania, Netherlands, Slovenia, Latvia, Spain	5 Italy, Norway, Portugal, Spain, Ireland, Sweden
			5 Finland, France	6 Poland, Slovakia, Malta, United Kingdom	
			7 Netherlands	7 Greece	
High	2 Croatia, Czech Republic, Estonia, Poland	1 Austria, Germany	1 Austria, Germany, Lithuania	1 Austria, Germany, Lithuania, Italy, Portugal	1 Austria, Czech Republic, Greece, Hungary, Luxembourg, Poland, Romania, Slovakia
	4 Hungary, Latvia, Lithuania, Slovakia, Romania	3 Hungary, Lithuania, Romania	4 Denmark, Greece, Ireland, Norway, Sweden, Switzerland, United Kingdom	2 Luxembourg, Belgium, Finland, Norway	3 Bulgaria, Finland, France, Germany, Lithuania
		5 Czech Republic, Denmark, Estonia, Finland, Greece, Latvia, Poland, Slovakia	6 Italy, Portugal, Spain	4 Denmark, Switzerland	4 Estonia, Latvia, Switzerland
		6 France			

Table 1 - Cluster and country relative status by group of indicators.

Super Cluster	Countries
1	Austria, Germany
2	Finland, Greece
3	Estonia, Poland
4	Hungary, Lithuania, Romania
5	Ireland, Italy, Netherlands, Portugal, Spain, and United Kingdom
6	Latvia, Slovakia

Table 5 - Clusters formed by the intersection of the clusters obtained using mortality- and morbidity-based indicators.

8. Figures

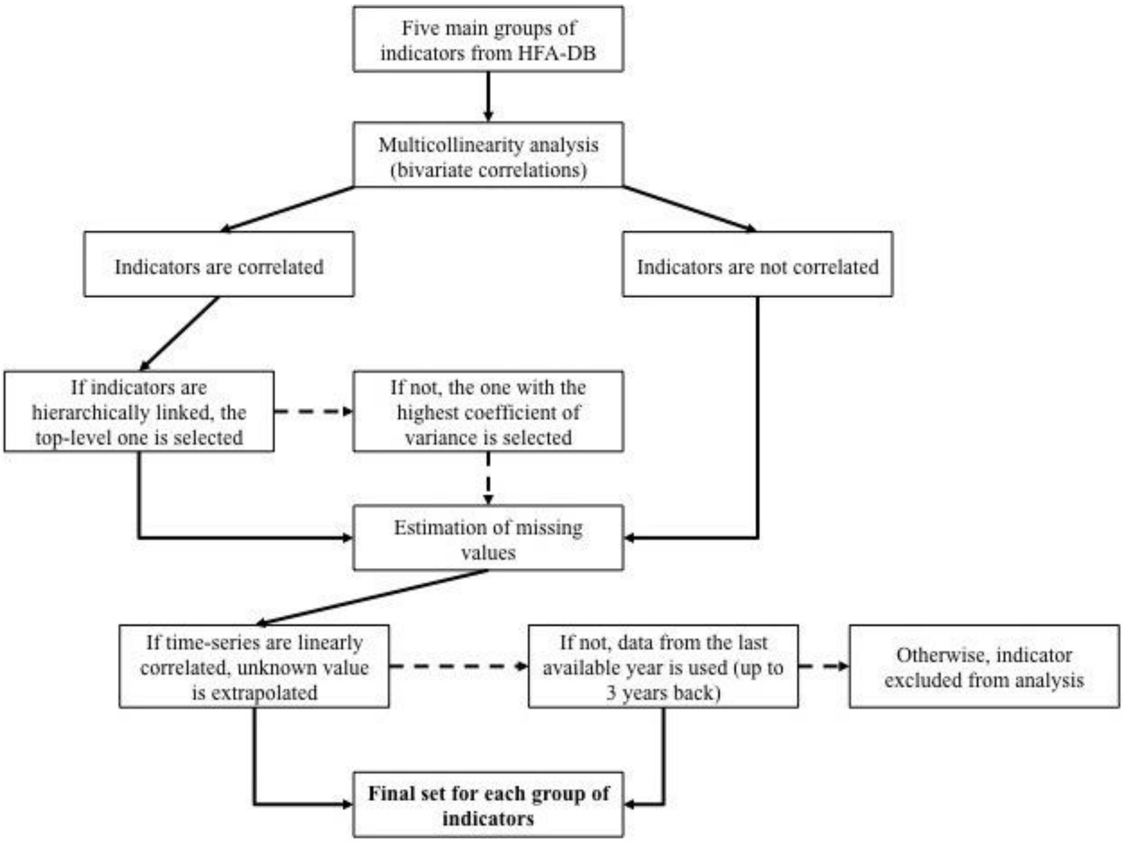

Figure 1 - Methodology adopted for obtaining the final selection of indicators to be used for clustering countries.

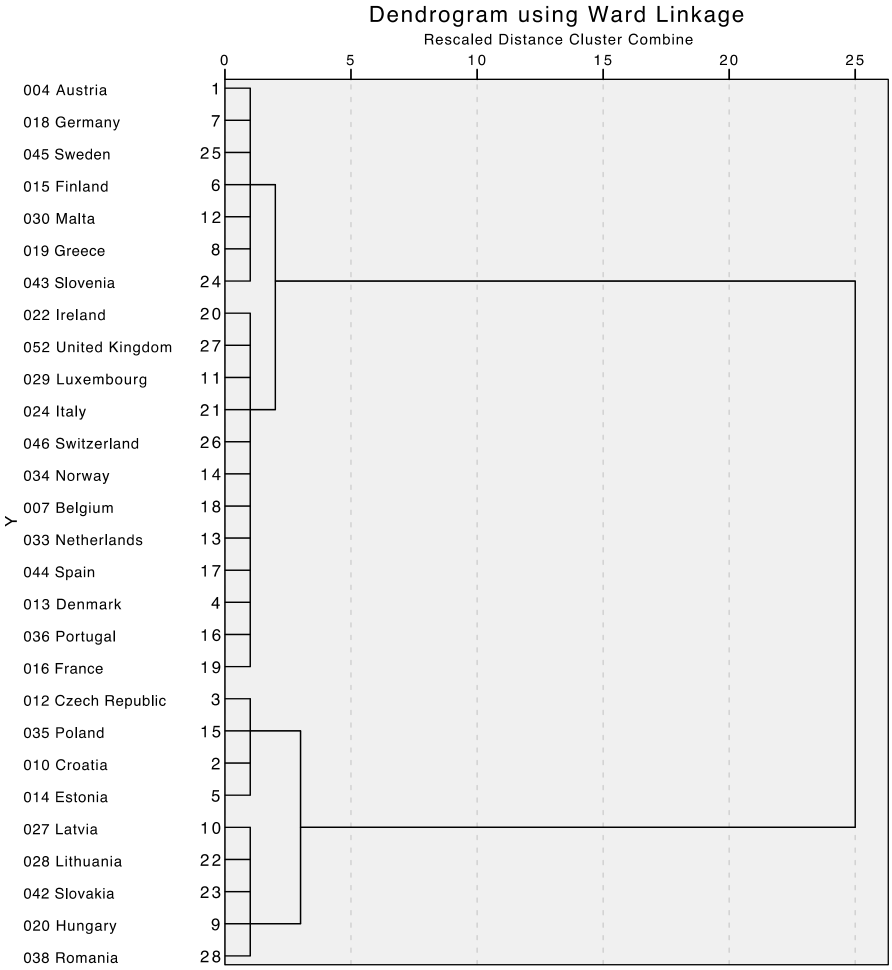

Figure 2 - Dendrogram obtained from applying HCA with Ward's method to mortality-based indicators.

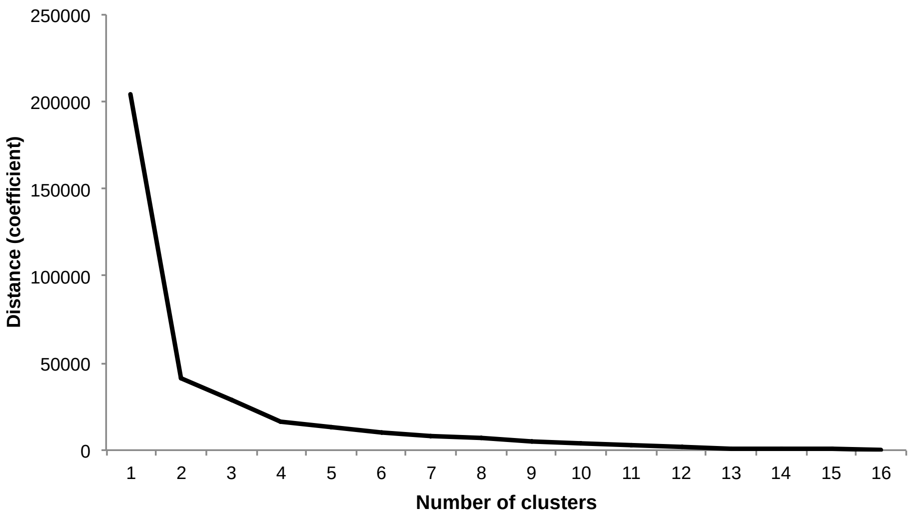

Figure 3 - Scree plot for defining the elbow, i.e., the cutting point in the number of clusters. The distance measured is within the cluster.

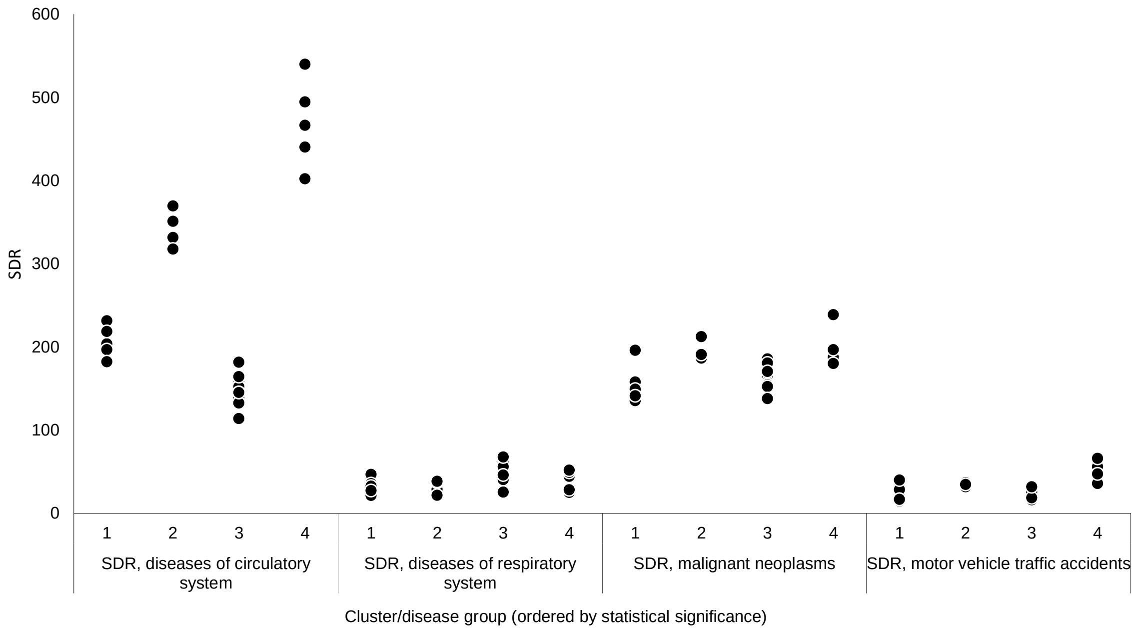

Figure 4 - A scatter plot depicting the most significant SDRs used to form the clusters along with the distances between indicators for each country.

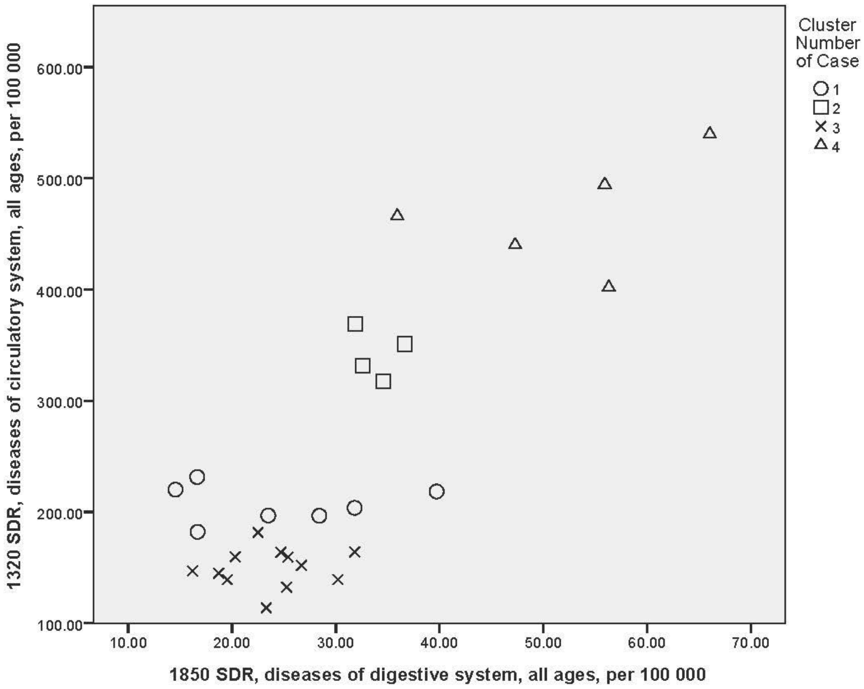

Figure 5 - Scatter plot depicting the correlation between the two most significant indicators. Clusters can be clearly identified, suggesting an acceptable partitioning.

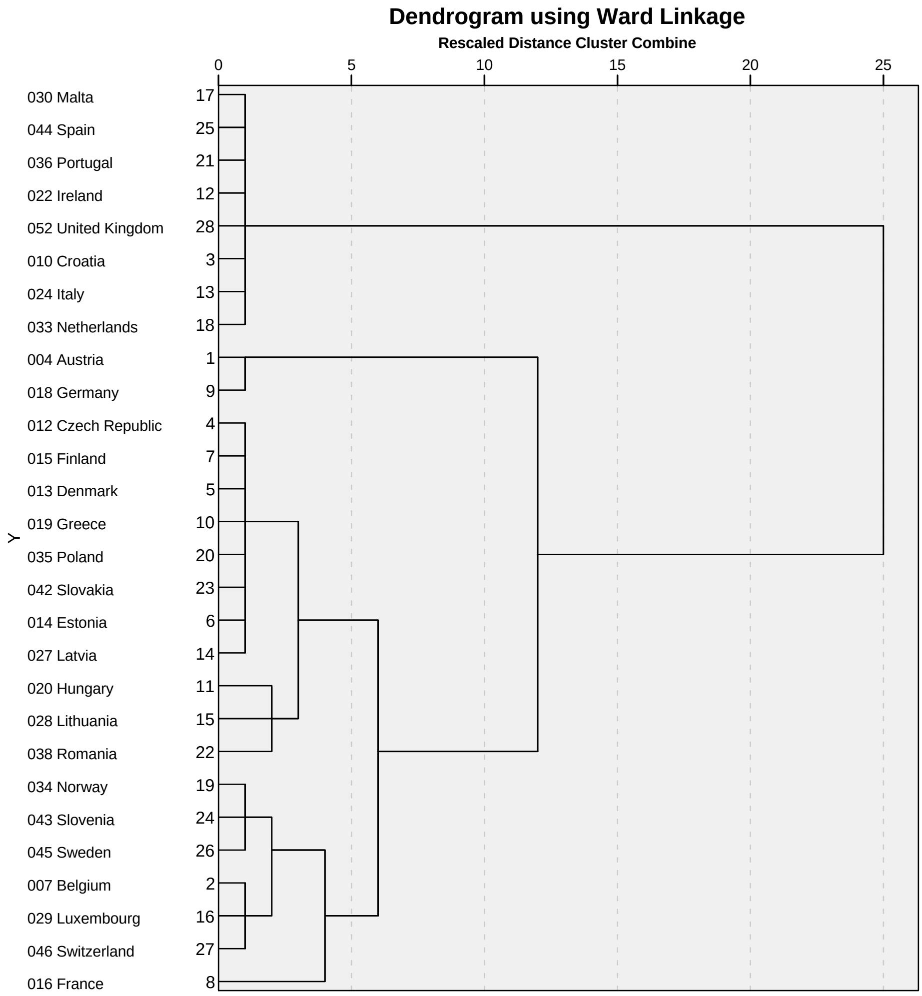

Figure 6 - Dendrogram obtained from applying HCA with Ward’s method to morbidity-based indicators.

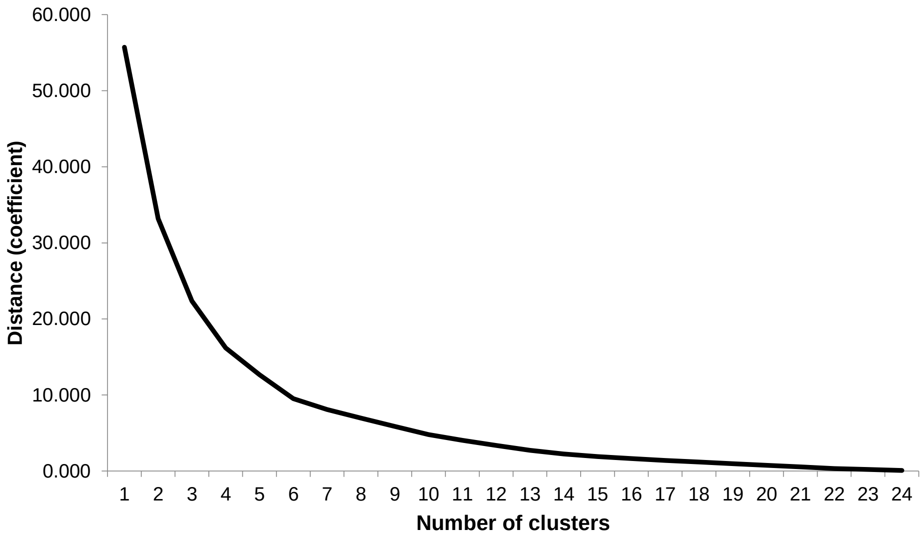

Figure 7 - Scree plot for defining the elbow, i.e., the cutting point in the number of clusters. The distance measured is within the cluster.

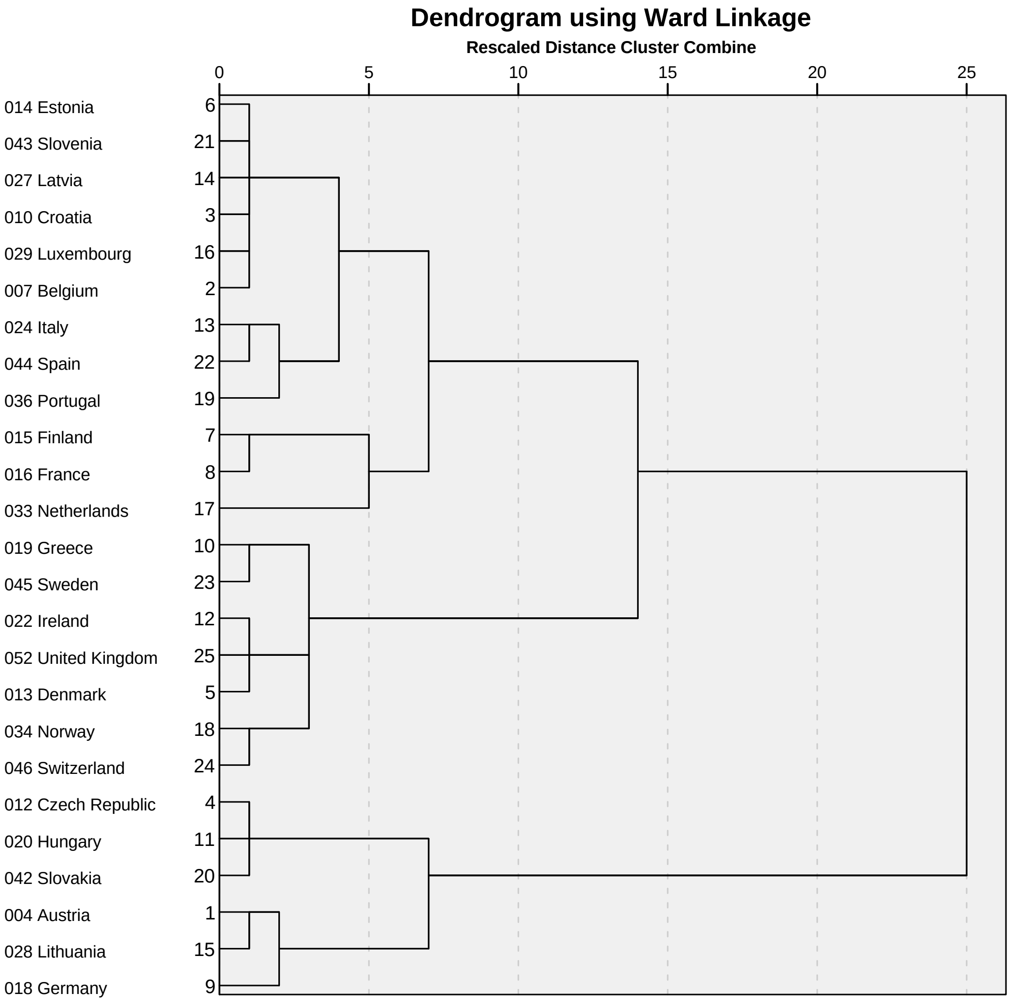

Figure 8 - Dendrogram obtained from applying HCA with Ward's method to utilization indicators.

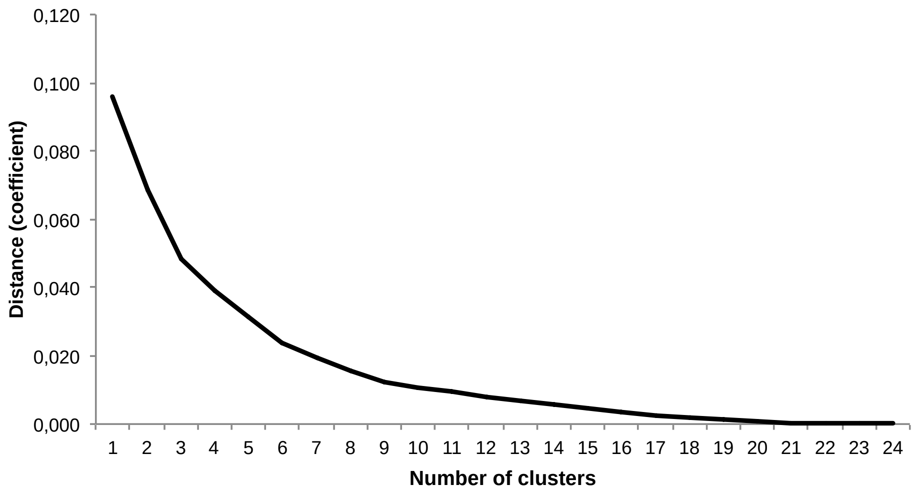

Figure 9 - Scree plot for defining the elbow, i.e., the cutting point in the number of clusters. The distance measured is within the cluster.

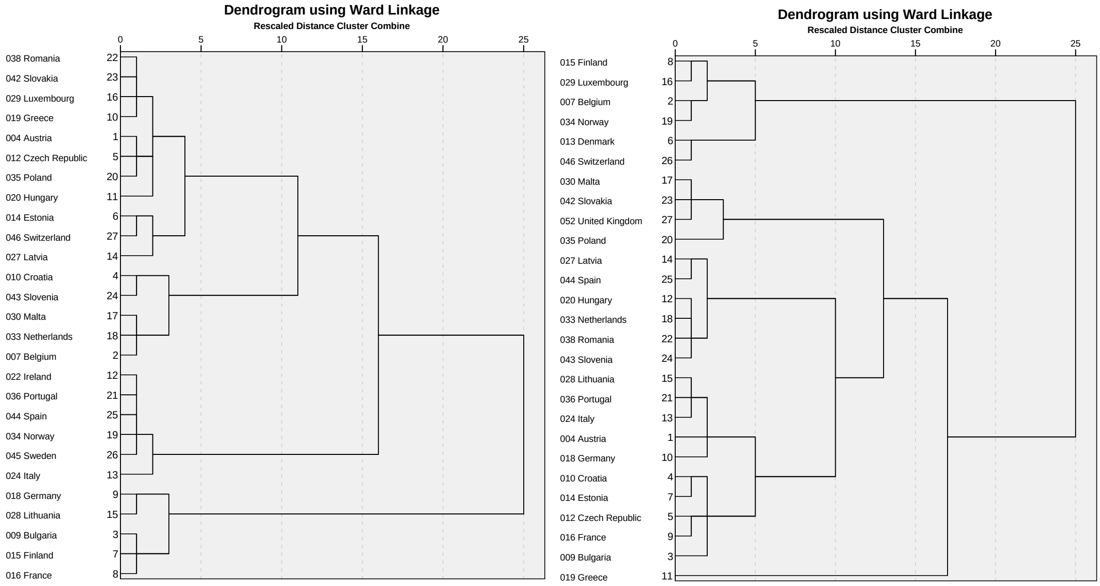

Figure 10 - Dendrogram obtained from applying HCA with Ward's method to physical resources (left) and human resources (right) indicators.

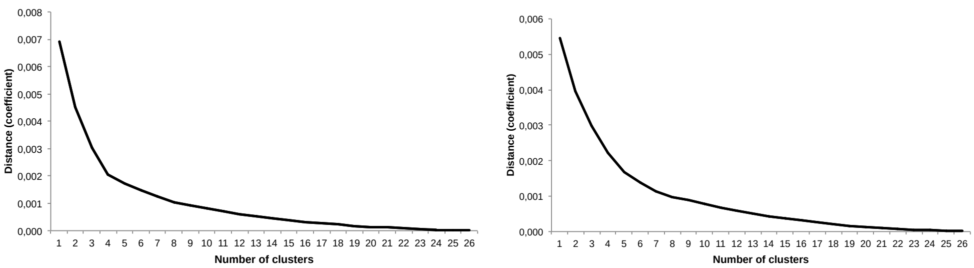

Figure 11 - Scree plot for defining the elbow, i.e., the cutting point in the number of clusters in healthcare physical (left) and human (right) resources.

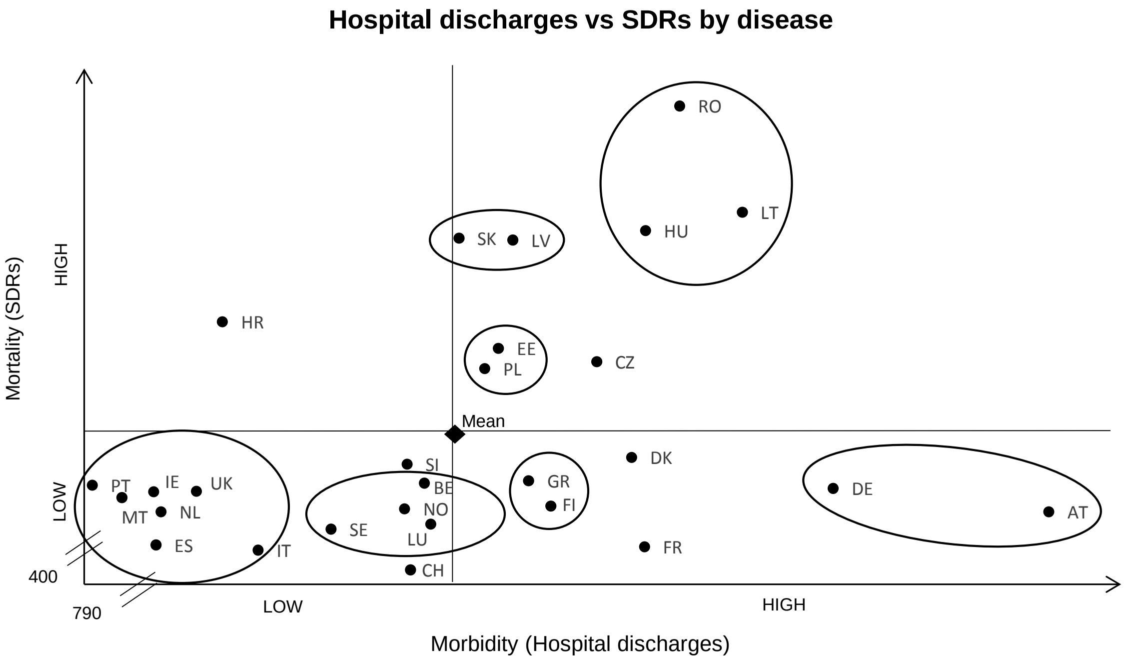

Figure 12 - Hospital discharges versus mortality SDRs per country. The groups formed include countries that are featured in the same cluster in both criteria, mortality and morbidity rates.

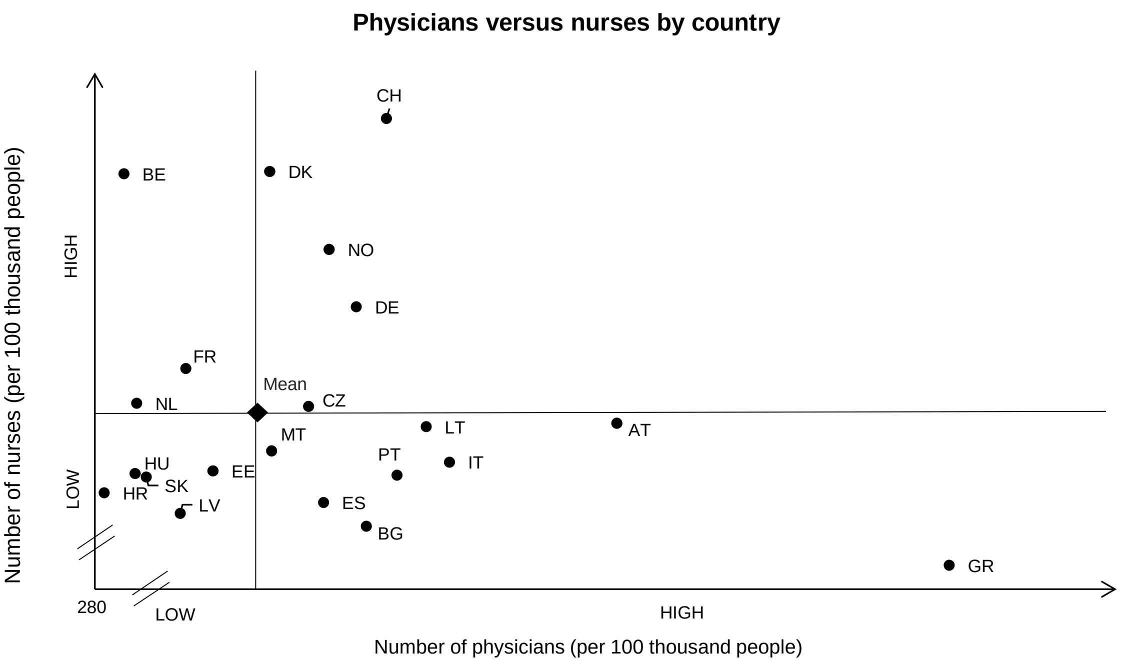

Figure 13 - A two-dimensional graph depicting physician and nursing intensity in each country.

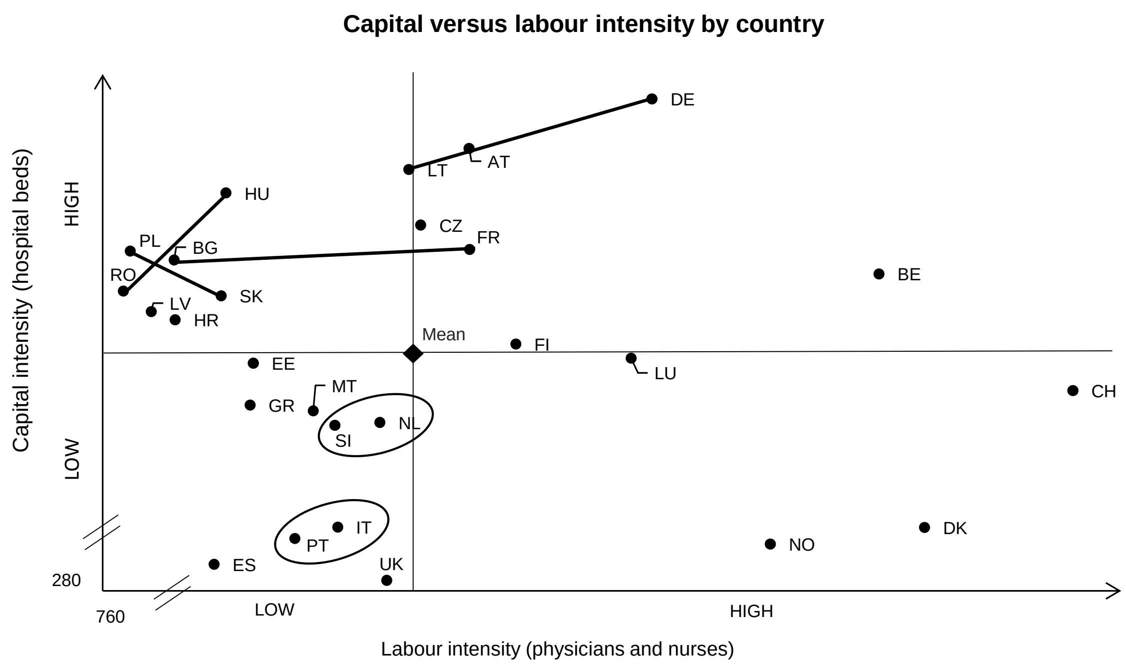

Figure 14 - Capital versus labour intensity by country. Lines and circles group countries that were part of the same cluster both in the human and physical resources clustering.

	AT	BE	BG	HR	CZ	DK	EE	FI	FR	DE	GR	HU	IE	IT	LV	LT	LU	MT	NL	NO	PL	PT	RO	SK	SI	ES	SE	CH	UK
AT																													
BE																													
BG																													
HR		2	1																										
CZ	1		1	2																									
DK		1			1																								
EE		1	1	**3**	**3**	1																							
FI	1	1	1		1	1	1																						
FR		1	2	1	1	1	1	2																					
DE	**4**		1					2	1																				
GR	2				2	2	1	2		1																			
HU	1				2						1																		
IE		1		1		2			1		1																		
IT	1	1		1		1			1	1			**3**																
LV		1		1	1	1	**3**	1			1	2																	
LT	2		1					1	1	**3**		2		1	1														
LU	1	**4**		1	1	1	1	1	1		1	1	1	1	1														
MT	1	1		2				1		1	1		1	1															
NL		2		2		1			1			1	2	2	1		1	2											
NO		**3**				2		1	1		1		**3**	2			**3**		1										
PL	1			1	**3**	1	2	1			2	1			1		1	1											
PT	1	1		1		1			1	1			**3**	**5**		1	1	1	2	2									
RO	1				1						1	**4**			2	2	1		1		1								
SK	1				**3**	1	1	1			2	**3**			2	1	1	1			**3**		2						
SI	1	**3**		2			1	1		1	1	1			2		2	2	2	1			1						
ES		1		1		1			1			1	**3**	**4**	1		1	1	**3**	2		**4**	1		1				
SE	1	1				1		1		1	2		2	1			1	1		**3**		1			2	1			
CH		2				**3**	1		1		1		2	1	1		2		1	**3**		1			1	1	2		
UK		1		1		2			1		1		**3**	2			1	2	2	2	1	2		1		2	1	2	

Figure 15 - Similarity matrix obtained by counting the intersections between clusters for the five indicators. High numbers indicate high similarity, and vice-versa.

Editorial Board (wps@fep.up.pt)
Download available at: http://wps.fep.up.pt/wplist.php
also in http://ideas.repec.org/PaperSeries.html

www.fep.up.pt

U.PORTO
FEP FACULDADE DE ECONOMIA UNIVERSIDADE DO PORTO
EFMD
School of Economics and Management, University of Porto
Rua Dr. Roberto Frias | 4200-464 Porto | Portugal
Telephone: +351 225 571 100