Health & Place 16 (2010) 1129–1135

Contents lists available at ScienceDirect

Health & Place

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

A framework for identifying similarities among countries to improve cross-national comparisons of health systems

David T. Bauer [a,*], Carl F. Ameringer [b]

[a] Department of Systems and Information Engineering, University of Virginia, 151 Engineer's Way, P.O. Box 400747, Charlottesville, VA 22904, United States
[b] L. Douglas Wilder School of Government and Public Affairs, Virginia Commonwealth University, Virginia Commonwealth University 919 W Franklin Street, P.O. Box 843061, Richmond, VA 23284, United States

ARTICLE INFO

Article history:
Received 24 March 2010
Received in revised form
15 July 2010
Accepted 16 July 2010

Keywords:
World heath
Health care evaluation
Cluster analysis
Principal component analysis

ABSTRACT

Cross-national research on health system performance can yield important findings for public policy purposes. We seek to further this research by examining the problem of selection bias, an important methodological issue that investigators initially should consider. Because of the logistical difficulties and enormous expense involved in collecting voluminous data from many countries, researchers often must rely on information contained in data sets of international organizations, such as the World Health Organization (WHO) and the Organization for Economic Cooperation and Development. Under the circumstances, the comparisons that researchers can make will depend to a great extent on the availability and richness of data for certain measures. This situation raises the potential for selection or experimenter bias. We use multivariate statistics to group countries with similar characteristics, an approach that we believe will mitigate the problem. We perform a cluster analysis of 186 countries using principal components derived from 7 demographic variables and 27 mortality and burden of disease variables. Our analysis produced six clusters that we believe represent suitable groupings for comparative purposes.

1. Introduction

Cross-national research on health system performance can yield important findings for public policy purposes (Murray and Frenk 2010). Thorpe et al. (2007), for example, showed that the spending gap between the United States and European countries reflected differences in diagnosis and treatment rates for certain chronic conditions. Their findings indicated that policy makers who seek to reduce costs and improve quality in the US should pay closer attention to risk factors for chronic disease and the way that physicians practice medicine (Thorpe, w685). Other examples include those of Schoen et al. (2009) (the connection between universal health insurance and health system performance), Yach et al. (2004) (the connection between chronic illness and the prevalence of key risk factors), and Anderson and Hussey (2001) (the connection between spending and outcomes).

In each of the above examples, researchers based their findings on data obtained from certain countries. Thorpe and associates, for instance, compared the United States to ten European countries; Schoen and colleagues based their findings on data obtained from the Australia, Canada, Germany, New Zealand, United Kingdom, and the US; Yach and associates compared information collected by the World Health Organization (WHO) for "developing countries;" Anderson and Hussey compared data for member countries of the Organization for Economic Cooperation and Development (OECD).

The selection of countries should be an important first step in any analysis that seeks to draw conclusions based on cross-national comparisons of health system performance. Because of the logistical difficulties and enormous expense involved in collecting data, researchers often must rely on information contained in data sets of international organizations, such as WHO or OECD. Under the circumstances, the comparisons that researchers can make will depend to a great extent on the availability and richness of data for certain measures (life expectancy, for example). This situation raises the potential for selection or experimenter bias. It also raises the concern that comparisons may not be valid because the countries selected are not sufficiently similar.

Researchers should seek to refine their methods for selecting countries for comparison to the extent possible. Doing so will help to validate their findings and to give policy makers good reasons to rely upon them. The problem particularly resonates in the US where there is a strong perception on the part of "many in the health policy community that international comparison is not useful because of the uniqueness of the United States" (Murray

* Corresponding author. Tel.: +1 434 806 5336; fax: +1 434 982 2972.
E-mail address: dtb5f@virginia.edu (D.T. Bauer).

1353-8292/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.healthplace.2010.07.004

and Frenk, 2010: p. 1). This perception gained traction when the World Health Organization in 2000 ranked the US 37th out of 191 countries in terms of "overall efficiency," its measure for health system performance. Though most critics of the WHO study did not focus on selection bias, the placement of the US behind the countries of San Marino (3rd), Oman (8th), Monaco (13th), and Morocco (29th), among others, created much controversy.[1] Did it make sense, they asked, to compare health system performance in the US with countries that appeared so dissimilar along several dimensions?

One way that researchers can enhance their findings is to provide objective and credible reasons for the countries they select. Writing on the problem of selection bias in the *American Sociological Review*, Berk (1983:396) stressed the need "to formulate a theoretical model of the selection process." "One needs a *theory of selection*," he said. "Without a theory, it is difficult to draw even preliminary inferences about the nature of the problem and impossible to choose how best to implement sample selection corrections."

Though we do not presume in this article to formulate an advanced theoretical model to resolve the problems we have identified, we do offer an approach that we believe will mitigate selection or experimenter bias and thereby enhance the findings of researchers for public policy purposes. We have chosen to cluster countries using five principal components derived from a number of health outcome and demographic measures. Clustering has several potential benefits. First, this method simplifies relatively large data sets such that researchers can more easily recognize patterns. At a high level, these patterns show which countries are similar (i.e., in the same cluster), and the factors that make clusters of countries significantly different from each other. Second, clustering is a well-established technique that has been employed in a variety of applications, including analyzing consumer purchasing behavior (Punj and Stewart, 1983), crime data (Chen et al., 2004), academic programs (Brown and Scherer, 2000), and health care systems (Day et al., 2008; Shelton et al., 2006; Shouls et al., 1996; Wendt, 2009). Third, cluster analysis is a flexible technique that is useful at multiple scales; thus, it provides consistency or uniformity if future studies use the same or a similar approach. Consequently, it provides a sound rationale for selecting countries to compare and reduces the perception among some policy makers that studies are biased by the inclusion or exclusion of certain countries.

2. Methods

Two multivariate statistical methods serve as the basis of our analysis. We use principal component analysis to determine the sources of variance among countries and to provide visual tools for understanding the data structure. We then identify sets of countries with similar characteristics by using the principal components as inputs to a cluster analysis.

2.1. Data

Comparative assessments of health care performance often focus on only a few variables despite the wealth of data available. The WHO Statistical Information System (WHOSIS; World Health Organization, 2009), for instance, contains 164 variables for 193 countries, and the OECD tracks over 1200 health related variables for its 30 member countries. Each variable contains potentially valuable information, but the challenge of interpreting high-dimensional data is a primary motive for limiting the number of variables in an analysis. Multivariate statistical analysis provides the means to interpret such data by revealing relationships and identifying sources of variance.

For our analysis, we selected 34 variables from two of the categories contained in WHOSIS—7 from the "demographics" category and 27 from "mortality and burden of disease." These variables are shown in Table 1. Seven of the 193 WHO member nations were excluded due to insufficient data: Monaco, Montenegro, San Marino, Serbia, Nauru, Niue, and Andorra. We employed a cluster-based data imputation method (Zhang et al., 2008) to estimate the 33 missing values (0.3% of the data) in the remaining 186 countries.

2.2. Statistical analysis

The goal of principal component analysis is to replace a group of variables with a set of linear combinations of those variables, called principal components. The main benefits of this approach are that a small number of principal components often sufficiently describe the variation in the data and that the components are uncorrelated, whereas strong correlations may exist among the original variables. Two key steps in interpreting the results of a principal component analysis are determining (1) the minimum number of components that sufficiently summarize the data and (2) those aspects of the data are represented by each component. We selected the group of components that accounted for at least 80% of the variance, provided that each of those components has a standard deviation greater than one.

Principal component analysis is often used in conjunction with other statistical methods, such as the cluster analysis. Cluster analysis refers to a set of techniques for dividing entities into groups, such that the entities in each group are more similar to each other than to entities in other groups. The results of a cluster analysis that used principal components are easier to interpret because fewer variables are included and they can be visualized graphically.

In agglomerative hierarchical cluster analysis, all countries (or other entities) are initially in their own cluster. During each stage one calculates the similarity between every pair of clusters and joins the most similar pair. After the first stage, one cluster will contain two countries, and the remaining clusters will still have only one country. The process continues until all countries are in a single cluster. The similarity between clusters is determined by the choice of distance measure and the linkage method. The distance measure indicates how to calculate the distance between two countries, and the linkage method determines how to use those distances to calculate the distance between clusters. For example, in single linkage clustering, the distance between two clusters is equal to the distance between the most similar pair of countries, while in complete linkage, it is equal to the distance between the least similar pair. Gan et al. (2007) provide details of these and other linkage methods.

Our distance measure was Euclidean distance using the first five principal components. This paper focuses on the complete linkage results, but we also performed analyses using average linkage, single linkage, Ward's method, as well as the *K*-means algorithm, a non-hierarchical method. These results are briefly discussed in a section on cluster validity. The results of a cluster analysis can vary drastically based on the specification of the linkage method and distance measure (Punj and Stewart, 1983).

[1] Much of the criticism centered on the measures that the authors of the WHO study used to determine a country's overall health system performance, coupled with the asserted lack of data to support those measures (Almeida et al., 2001). Our purpose is not to credit these claims or to argue the merits of any particular position. Rather, our purpose is to propose a means for determining which countries are the best candidates for comparison.