Applied Geography 95 (2018) 9–18

Contents lists available at ScienceDirect

Applied Geography

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

Geographic and statistic stability of deprivation aggregated measures at different spatial units in health research

Alba Cebrecos[a,b,c], M. Felícitas Domínguez-Berjón[c,d], Ignacio Duque[c,e], Manuel Franco[a,f], Francisco Escobar[a,b,*]

[a] Social and Cardiovascular Research Group, University of Alcala, Spain
[b] Dept. of Geology, Geography and Environmental Sciences, University of Alcala, Spain
[c] Group of Social Determinants of Health. Spanish Society of Epidemiology, Spain
[d] Dirección General de Salud Pública, Consejería de Sanidad, Madrid, Spain
[e] Subdirección General de Tecnologías de la Información y las Comunicaciones, Instituto Nacional de Estadística, Madrid, Spain
[f] Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, USA

ARTICLE INFO

Keywords:
Modifiable areal unit problem
Socioeconomic factors
Cardiovascular disease
Geographic scale

ABSTRACT

Deprivation indices constitute a valuable tool for assessing health inequalities. A key issue when analyzing deprivation is the choice of the geographical scale and spatial unit of analysis. Our objective was to evaluate statistical and geographical stability of an Area Based Deprivation Index (ABDI) computed at different spatial scales and to study their relation with cardiovascular disease.

The present study has been conducted in the city of Madrid, Spain. Madrid divides its territory in three different administrative units nested within each other: census section, neighborhoods and districts. For each unit a deprivation index was calculated through Principal Component Analysis (PCA). The data source was the 2011 national census from where a range of socioeconomic and demographic indicators were selected. To study statistical and geographical stability of deprivation we used an Exploratory Spatial Data Analysis and bivariate Local Indicators of Spatial Association analysis. We also conducted Pearson correlation analyses to study the change in the relationship between deprivation and the prevalence of cardiovascular disease (CVD) across the three scales.

At census section and neighborhood level, first component showed four and five factors loading higher than 0.6, respectively. These factors loading related to occupancy/labor market and education. However at district level, first component showed seven factors loading higher than 0.6 and related to occupancy/labor market, education and immigration. With indicators of these factors loading, deprivation indices were calculated for each administrative unit by extracting a single PCA axis. Variance explained for each index was 65%, 86% and 79%, respectively. Bivariate local autocorrelation analyses showed aggregated areas of low and high stability with variable degree of significance in the three scales. The ABDIs calculated at census section level, neighborhood level and district level presented different significant correlations with CVD prevalence ($r = 0.328$; $r = 0.635$; and $r = 0.739$ respectively). These results show that the deprivation index did not remain stable across the three scales, neither were the correlations between deprivation and age-adjusted CVD prevalence.

Understanding the stability of a spatial phenomenon across different scales is essential to determine the best unit of aggregation of data when studying an important process such as socioeconomic deprivation and its possible health impacts.

1. Introduction

Mortality and morbidity increase as the social position of a population decreases. This progressive and lineal phenomenon was defined by Marmot as *social gradient* (Marmot, 2005). This is a universal phenomenon, although the magnitude and extent may vary between countries and change overtime (Marmot, 2005; Sir Michael, 2006). In Nova Scotia (Canada) 35,266 premature deaths over a 11 year-period were studied and concluded that about 40% were attributable to socioeconomic inequalities (Saint-Jacques, Dewar, Cui, Parker, &

* Corresponding author. Dpt of Geology, Geography and Environmental Sciences, University of Alcala, Calle Colegios 2, 28801, Alcala de Henares, Madrid, Spain.
E-mail addresses: alba.cebrecos@uah.es (A. Cebrecos), felicitas.dominguez@salud.madrid.org (M.F. Domínguez-Berjón), ignacio.duque.rodriguez@ine.es (I. Duque), manuel.franco@uah.es (M. Franco), francisco.escobar@uah.es (F. Escobar).

https://doi.org/10.1016/j.apgeog.2018.04.001
Received 22 June 2017; Received in revised form 30 March 2018; Accepted 2 April 2018

Dummer, 2014). In Europe, researchers studied 26,229,104 European inhabitants from 16 cities and found that up to 30% of excess deaths were attributable to socioeconomic disparities (Borrell et al., 2014). In Spain, mortality excess related to deprivation was 59,445 deaths among men and 23,292 among women (Borrell, Marí-Dell'Olmo, Serral, Martínez-Beneito, & Gotsens, 2010).

A given area may be socioeconomically deprived as a result of multiple interrelated factors. Deprived indices were first developed in the U.K (Carstairs & Morris, 1989; Townsend, 1987) as a multivariate tool which allows to study the level of deprivation in an area, comparing deprivation effects across a variety of geographic regions and as a proxy of individual-level socioeconomic status (Smith, Hart, Watt, Hole, & Hawthorne, 1998). These indices provide a synthetic measure of different aspects of deprivation.

In Spain, the project "Socio-economic and environmental inequalities in mortality in small areas of Spanish cities", or MEDEA project (http://www.proyectomedea.org/), describes the geographical patterns of mortality and their relationship with the socioeconomic and environmental characteristics. To assess the socioeconomic status of areas in Spain, the MEDEA team developed an deprivation index based on 2001 national census (MF. Domínguez-Berjón et al., 2008). Results of this work have been used in a variety of research areas such as the study of the association between air pollution, Socio-Economic Status (SES) and health (Barceló, Saez, & Saurina, 2009; Cambra et al., 2013; García-Pérez et al., 2009; Ramis et al., 2009); health inequalities and mortality (M. F. Domínguez-Berjón et al., 2010; Gandarillas et al., 2011; Gotsens et al., 2011; Nolasco et al., 2009; Segura del Pozo et al., 2010) and deprivation and cancer incidence (Saurina et al., 2010) among others. In addition to their scientific impact, these results have been used as complementary information in the prioritization of interventions with an equity approach, published in the State Health Reports of Madrid Region (Primaria, 2007–2014). These reports are the base-document for public health planning in Madrid.

Given the dramatic social and economic changes occurred in Spain since late 2007, a redeveloped deprivation index with 2011 census was necessary. The general criticism of the deprivation indices focuses on the selection of the indicators that build the index, however less attention has been given to its geographical variability, a problem known as the Modifiable Areal Unit Problem (MAUP) (Schuurman, Bell, Dunn, & Oliver, 2007).

MAUP is an inherent problem of overlapping artificial spatial units (e.g. administrative areas) over a continuous geographical phenomenon (Openshaw, 1984). That implies a potential measurement error due to the aggregation of statistical data into these artificial units. Boundaries of these areas are defined by historical, political or operational reasons. There are not designed to define homogeneous zones with respect to social, economic or health characteristics of population. Social, economic or environmental phenomenon which promote or restrict health risks, are not limited by the boundaries of these artificial units.

The MAUP is composed by two interrelated effects (Openshaw, 1984). First, the scale effect which is the variation in results that can be obtained when data for one set of areal units are progressively aggregated into fewer and larger units of analysis (Openshaw, 1984). As an example, when census data are aggregated in neighborhoods, districts, and municipalities, results may change with increasing scale. Second, the zoning effect which refers to the different configurations of zones of the same size that may generate different results (Houston, 2014). Another example, when results using a 100-m continuous grid system differed from results using a 100-m grid system oriented in different ways (Houston, 2014).

In the international public health research field, there have been pioneering works that studied the implications of this problem in health research. The Geocoding project, studied which level of geography would be most apt for monitoring US socioeconomic inequalities in health, overall and within diverse racial/ethnic-gender groups (Krieger et al., 2003a; Krieger, Waterman, Chen, Soobader, & Subramanian, 2003b; Krieger et al., 2002). Another Canadian study highlighted the effect of scale on indices by mapping ABDIs at multiple census scales in an urban area and compared self-rated health data with ABDIs at two census scales (Schuurman et al., 2007).

Selecting the spatial scale involves setting the spatial resolution of the study, this means, the ability to distinguish objects on the earth surface. On a larger scale, statistical aggregation of data on smaller surface units (e.g. census sections) will be required, but these areas contain fewer cases and therefore less stable rates. By contrast, at smaller scales, aggregation occurs in larger areas (e.g. districts), blurring significant variability and can sometimes lead to interpretations that are contradictory to those derived from finer-resolution data (Nelson & Brewer, 2015). Thus, it is necessary to consider that MAUP affects most statistics and has an impact on variance, standard deviation, correlation, regression analysis and any other statistical result (Flowerdew, Manley, & Sabel, 2008).

Any public health phenomenon under analysis is inseparably related to the scale because the scale provides its meaning. However, Geographic Information Systems (GIS) have been underused in health studies and wherever used, limited analyses selecting the relevant scale area have been conducted. The present study assesses the statistic and geographic (in)stability that arises when using deprivation data aggregated at different scales when understanding social determinants of health with a geographical perspective.

The aims of the study are twofold: first, to build an Area Based Deprivation Index (ABDI) at three different spatial scales (census section, neighborhoods and districts) for the city of Madrid. Second, to study the statistical and geographical stability of the ABDI throughout the three scales and its relationship with cardiovascular disease (CVD) prevalence.

2. Methods

2.1. Study setting and sample

This study is framed within a larger project, the Hearth Healthy Hoods (https://hhhproject.eu/). The HHH general objective is to understand the urban physical and social environment of the city of Madrid (Spain) and its relation with residents cardiovascular health (Bilal et al., 2016; Cebrecos et al., 2016; Julia; Díez et al., 2016; J Díez et al., 2017; Gullón et al., 2015). The HHH project includes the sociodemographic characterization of Madrid administrative spatial units. In this task, HHH researchers collaborate closely with the MEDEA project, with extensive experience in study of social deprivation and its association with health in Spain (Barceló et al., 2009; Cambra et al., 2013; MF.; Domínguez-Berjón et al., 2008; Gandarillas et al., 2011; García-Pérez et al., 2009; Nolasco et al., 2009; Ramis et al., 2009; Saurina et al., 2010; Segura del Pozo et al., 2010).

Madrid municipality, with an extension of 605 sqkm, consists of 21 districts, which are in turn divided into 128 neighborhoods and in 2409 basic spatial units called census sections (based on the 2011 Census data of National Statistical Institute -INE-). Madrid municipality had a total population of 3,186,595 in 2016 (INE, 2016). The administrative configuration of Madrid began at XII and XIII centuries and increasing its size as the capital of the Spanish Kingdom during the second half of XVIII century with different administrative divisions ("cuarteles", "barrios" and others). It was in the second half of XIX century, due to important migration flows from rural to urban areas that appeared for the first time the denomination of "district". Population increases and restructuring of administrative division continued with subsequent urban plans and developments until 1988 with the final and current number of district and neighborhoods (Prado Martínez, 2004). According to the 2011 census, neighborhood average population is 24,895 (max. 69,300 and min. 770 inhabitants) and average surface is 4.7 sqkm (max. 187.5 and min. 0.25 sqkm). Districts have an average population of 151,743 inhabitants (max. 246,020 and min. 45,625