Social Science Research 56 (2016) 133–143

Contents lists available at ScienceDirect

Social Science Research

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

Discovering complex interrelationships between socioeconomic status and health in Europe: A case study applying Bayesian Networks

Javier Alvarez-Galvez [a, b, *]

[a] Loyola University Andalusia, Department of International Studies, Campus de Palmas Altas, Faculty of Political Sciences and Law, Seville 41014, Spain
[b] Complutense University of Madrid, Department of Sociology IV (Research Methodology and Communication Theory), Campus de Somosaguas, Faculty of Political Sciences and Sociology, Pozuelo de Alarcón, Madrid 28223, Spain

ARTICLE INFO

Article history:
Received 30 November 2013
Received in revised form 7 November 2015
Accepted 31 December 2015
Available online 8 January 2016

Keywords:
Socioeconomic status
Self-rated health
Structure discovery
Additive Bayesian networks
Europe

ABSTRACT

Studies assume that socioeconomic status determines individuals' states of health, but how does health determine socioeconomic status? And how does this association vary depending on contextual differences? To answer this question, our study uses an additive Bayesian Networks model to explain the interrelationships between health and socioeconomic determinants using complex and messy data. This model has been used to find the most probable structure in a network to describe the interdependence of these factors in five European welfare state regimes. The advantage of this study is that it offers a specific picture to describe the complex interrelationship between socioeconomic determinants and health, producing a network that is controlled by socio-demographic factors such as gender and age. The present work provides a general framework to describe and understand the complex association between socioeconomic determinants and health.

1. Introduction

Socioeconomic circumstances influence our physical and mental health. Evidence has revealed that the higher the socioeconomic status (SES), the lower the prevalence of health problems, illness, disease and death (Wilkinson and Marmot, 2003; Bartley et al., 2004). The connection between SES and self-rated health (SRH) varies in degree in different countries (Adams et al., 2003a; Frijters et al., 2005). Income, education, occupational and class inequalities vary from one country to another and these socioeconomic determinants could lead to different health outcomes (Wilkinson, 1996, 1997; Nummela et al., 2007; Costa-Font and Hernández-Quevedo, 2012).

Literature on this topic has mainly focused on studying the impact of various socioeconomic factors such as the direct determinants of individuals' states of health (Alvarez-Galvez et al., 2013; Eikemo et al., 2008a,b; Kunst et al., 2005). Studies demonstrate that SES affects health outcomes and thus, traditionally, it is assumed that health is determined by socioeconomic conditions. However, this process of direct influence relating socioeconomic factors to health outcomes is difficult to sustain from a theoretical point of view. On the one hand, this relationship might disappear or even be inverted if additional

* Loyola University Andalusia, Department of International Studies, Campus de Palmas Altas, Faculty of Political Sciences and Law, Seville 41014, Spain.
E-mail addresses: jalvarez@uloyola.es, javalvar@ucm.es.

http://dx.doi.org/10.1016/j.ssresearch.2015.12.011
0049-089X/

variables are taken into account, and on the other hand, the initial assumption could be theoretically inverted. That is, studies assume that socioeconomic status determines individuals' states of health, but how does health determine socioeconomic status? The association between socioeconomic status and health is a well-grounded research finding (Costa-Font and Hernández-Quevedo, 2012). Individuals with higher income, education, social class or occupational status tend to report better health, and thus lower mortality, compared to those individuals with lower SES (Wilkinson, 1996, 1997; Nummela et al., 2007; Costa-Font and Hernández-Quevedo, 2012; Alvarez-Galvez et al., 2013; Eikemo et al., 2008a,b; Kunst et al., 2005).

Some attention has been dedicated to the effect of health on education (Bleakley, 2003). In fact, studies have shown that poor health can lead to lower academic achievement, and also variations in this association across genders. In this sense, if health can affect the academic success of students, following the same line of reasoning, it is possible to assume that health can affect professional achievement at work, and therefore our income level and social class. In practice, health is an evident source of human welfare and also an instrument for raising individuals' socioeconomic conditions (Bloom and Canning, 2008). However, there is a major difficulty in measuring the effect of health on our socioeconomic status, especially when attempting to explain the two-way causality between a population's wealth and health (Smith, 1999). To solve this problem of measuring the effect of health on individuals' SES, scientists usually establish the causal paths between these factors through the use of panel data. This strategy correlates the prevalence of health shocks with personal socioeconomic growth over time (e.g. an increase in income level, worker productivity, student academic achievement, etc.) (Adams et al., 2003b).

Although panel data represent a pertinent alternative for dealing with the problem of causality, this approach presents an additional weakness. Again, a certain kind of (bivariate) directionality between the two factors is assumed. That is, from this point of view we find two possibilities: (1) SES affects health (SES → Health); or (2) Health affects SES (Health → SES). Therefore, we are stuck in a vicious circle. Although these theoretical approaches or SES influence health or, to the contrary, health influences SES, in practice these relationships might be dynamic, multiple and, possibly, context-dependent. In other words, the directionality and relevance of these factors may vary in different contexts. In fact, literature provides evidence that the effect of the SES determinant varies depending on welfare state regimes (for example, depending on the level of public spending or the GDP in specific countries) (Eikemo et al., 2008a,b). Subsequently, standard multivariate regression models are not useful to explain the complexity of these relationships, especially when multiple associations (i.e. potentially interdependent variables) and contextual dependency could (and even should) be considered in the analysis. Exploratory analysis of data guided by automated regression analysis (e.g. stepwise regression models) may be useful if we want to explain the isolated effects of different predictors on the outcome variable, but not for studying their interrelationships and how these complex associations might produce multiple outcomes for the variables we are trying to describe (Lewis and McCormick, 2012).

To solve these difficulties, this exploratory study uses a different strategy, based on Bayesian Networks (BN). BN analysis is a type of statistical modeling that uses empirical data to perform a graphical network describing the dependency structure existing between a set of variables. Although the origin of Bayesian Networks modeling is related to the field of computer science (e.g. machine learning and data mining) (Heckerman et al., 1995; Needham et al., 2007), these models have also been used recently in health sciences and epidemiology (Lewis and McCormick, 2012; Lewis and Ward, 2013; Caillet et al., 2015).

The present study has two basic objectives: (1) to identify alternative interrelationships between SES and health in different welfare state regimes; and (2) to use this information to develop new hypotheses and a better theoretical understanding, using a data-driven approach. Thus, in order to explain the complex interrelationship between health and SES determinants, an Additive Bayesian Networks (ABN) model (Koivisto and Sood, 2004) has been performed to find the most probable structure in a network that, being formally depicted as a directed acyclic graph (DAG), identifies the most probable ties between these factors. The resulting networks have been contextualized in five welfare state regimes, selecting the European context as a case study. Some peculiarities of the different welfare regimes have been set out in Table 1 (Alvarez-Galvez et al., 2013).

2. Methods

2.1. Data and variables

A cumulative data set containing the four waves of the European Social Survey (2002–2008) is used (ESS Round 1-4, 2002) to explain how the association between SES and health varies in European welfare state regimes. The ESS dataset contains a total of 185,154 units at individual level, and 29 countries at aggregated level: Austria, Belgium, Bulgaria, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Israel, Italy, Luxembourg, Netherlands, Norway, Poland, Portugal, the Russian Federation, Slovenia, Slovakia, Spain, Sweden, Switzerland, Turkey, Ukraine and the United Kingdom.

In this case, contrary to the previously mentioned approach based on standard regression analysis (SES → Health vs. Health → SES), in the present model there is not one specific dependent variable but multiple ones whose associations may vary depending on the internal interdependency structure in the dataset and also on context peculiarities.

With respect to previous studies (Alvarez-Galvez et al., 2013; Hanibuchi et al., 2012), different variables have been included in the BN model. The variables under study are the following: