



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.

© 2016 Elsevier Inc. All rights reserved.

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.

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 ($SES \rightarrow Health$); or (2) Health affects SES ($Health \rightarrow SES$). 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 ($SES \rightarrow Health$ vs. $Health \rightarrow SES$), 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:

Table 1
Descriptions of European welfare state regimes (Alvarez-Galvez et al., 2013).

Welfare regime	Countries	Main characteristics
Liberal/Anglo-Saxon	United Kingdom Ireland	Based on the notion of market dominance and private provision; basic and minimal levels of provision and modest social transfers; the state only interferes to ameliorate poverty and provide for basic needs, largely on a means-tested basis; high social stratification.
Social-Democratic/ Scandinavian	Denmark Finland Norway Sweden	Universalism; access to benefits and services based on citizenship; comparatively generous social transfers; commitment to full employment and income protection; a strongly interventionist state; promotion of social equality through a redistributive social security system.
Conservative/ Bismarckian	Germany France Austria Belgium Luxemburg Netherlands Switzerland	Based on the principle of subsidiarity; dominance of social insurance schemes where the redistributive impact is minimal; distinguished by its 'status differentiating' welfare programs in which benefits are often earning-related; the role of the family is also emphasized; and the role of the market is marginalized.
Mediterranean/ Southern	Cyprus Greece Italy Portugal Spain Turkey	Characterized by a fragmented system of welfare provision; income guarantees are linked to work position, so income maintenance schemes vary from being meager to very generous; a high percentage of social expenditure financed through contributions; health care as a right of citizenship; low social expenditure as a percentage of GDP; strong reliance on the family and on the charitable sector
Post-Communist/ Eastern	Bulgaria Czech Republic Estonia Hungary Poland Romania Russian Federation Slovak Republic Slovenia Ukraine	Underdeveloped welfare state; system behind the average European level of social provision; strong labor market institutions and solid industrial economy with well-regulated work relations; compared to other member states of the European Union, they have limited health service provision.

1. Self-rated health (SRH): the individuals' subjective general state of health. In the ESS questionnaire the question was (C7): 'How is your health in general? Would you say it is ...'. The values and respective categories for response were: 1 'Very good'; 2 'Good'; 3 'Fair'; 4 'Bad'; 5 'Very bad'; 7 'Refusal'; 8 'Don't know'; 9 'No answer'. This 5-point Likert scale was inverted to facilitate the logical interpretation: 1 'Very bad'; 2 'Bad'; 3 'Fair'; 4 'Good'; 5 'Very good'.
2. Household total net income (income): this variable is reported in 12 banded categories that are standardized for all countries in the survey (for more information on this variable, visit the following link: <http://ess.nsd.uib.no/>). This variable is treated as continuous by taking the midpoint of each banded category (Yannis et al., 2009).
3. Occupational status: this variable is measured using the International Socioeconomic Index of Occupational Status (ISEI) (Ganzeboom and Donald, 1996). This indicator, derived from the International Standard Classification of Occupations (ISCO88), obtains scores for each occupation category by the optimal scaling of the occupational unit group (ILO (International Labour Office), 1990).
4. Education: this variable was measured as years of completed full-time education.
5. Insider: this variable describes the dual position that individuals occupy in the labor market: outsiders or insiders. Outsiders are understood as individuals with a greater likelihood of being unemployed, atypically employed or having part-time employment than the average member of the workforce in a particular labor market, whereas insiders can be defined as individuals positioned in relatively safe full-time jobs (i.e. the typical workforce in industrial societies). In the model, these individuals are afterwards classified as 0 'outsider' or 1 'insider'. The encoding of this variable is based on the approach of Häusermann and Schwander (Häusermann and Schwander, 2012) (see Table 2), an indicator that defines a precise map of labor market dualization in contemporary welfare state regimes.
6. Self-Perceived discrimination (SPD): this variable was introduced into the model since a relevant part of the literature has highlighted the effect of self-perceived discrimination (based on race or ethnicity, age, gender, sexuality, etc.) on mental and physical health outcomes (Kim and Williams, 2012).
7. Gender: where 0 'male' and 1 'female'. This variable was included in the model as a control for the relationship between health and SES determinants.
8. Age: the age of the respondent, also as a control variable.

In summary, the resulting BN model is constructed with the combination of eight variables: (a) SRH, (b) four SES measures (income, occupational status, education and outsidersness), (c) SPD, and (d) two controls (gender and age). Descriptive statistics for these variables are presented in Table 3.

Table 2
The map of dualization: insiders and outsiders in the five welfare state regimes^a.

	Liberal regime	Nordic regime	Continental regime	Southern regime	Eastern regime
Insiders	Older male LSF	Young male LSF Older male LSF Young female SCP	Young male LSF Older male LSF	Older male LSF	Older male LSF
	Young male SCP	Young male SCP	Young male SCP		
	Older male SCP	Older male SCP	Older male SCP	Older female SCP	Older female SCP
	Young male BC	Young male BC	Young male BC	Older male SCP	Older male SCP
	Older female BC			Young male BC	Young male BC
	Older male BC	Older male BC	Older male BC	Older female BC	Older female BC
	MSF	MSF	MSF	Older male BC	Older male BC
	CA	CA	CA	MSF	MSF
				CA	CA
Outsiders	Young female LSF	Young female LSF	Young female LSF	Young female LSF	Young female LSF
	Young male LSF			Young male LSF	Young male LSF
	Older female LSF	Older female LSF	Older female LSF	Older female LSF	Older female LSF
	Young female SCP		Young female SCP	Young female SCP	Young female SCP
	Older female SCP	Older female SCP	Older female SCP		
	Young female BC	Young female BC	Young female BC	Young female BC	Young female BC
		Older female BC	Older female BC	Older female BC	Older female BC

^a Capital accumulators (CA), mixed service functionaries (MSF), low service functionaries (LSF), blue collar workers (BC) and sociocultural professionals (SCP).

Source: Adapted from Häusermann and Schwander, 2012.

Table 3
Descriptive statistics for variables in the BN model.

	Obs.	Mean	Std. Dev.	Min.	Max.
Self-Rated health (SRH)	184,718	3.76	0.93	1	5
Household net income	129,344	25077.82	25288.78	1350	180,000
Occupational status (ISEI)	129,344	42.06	16.64	16	90
Education (in years)	160,280	11.95	4.11	0	56
Insider (ref. category = outsider)	155,146	0.66	0.47	0	1
Self-Perceived Discrimination (SPD)	181,620	0.06	0.25	0	1
Age of respondent	184,038	46.68	18.47	12	123
Gender (ref. category = male)	184,745	0.54	0.50	0	1

Taking into account that some indicators could be highly correlated due to their internal definition (e.g. gender and insider, or ISEI and education), a multicollinearity diagnosis was performed in order to prevent possible problems in the BN models. This test did not present any problematic values for the variance inflation factor (VIF) and tolerance (see Appendix A).

2.2. Bayesian network modeling

BN modeling is a data analysis technique that combines graph theory and the probability theory; it is ideally suited to find interdependency structures in messy or complex data. BNs are graphical representations for probabilistic relationships among a set of random variables, whose causal relationships are represented by arrows (including bidirectional relationships). The resulting graph (G) is composed of two basic elements: (1) nodes or vertices (i.e. variables); and (2) edges or arcs (i.e. links describing relationships). In BNs, each node represents a random variable $X = [X_1, \dots, X_n]$ where each variable X_i may take values from a finite set (i.e. variable values), whereas the edges represent the dependence relationship (or direct influence) among the variables in the model. For instance, a link between $X_i \rightarrow X_j$ indicates that variable X_i is the parent and X_j the descendant. In addition, variables (nodes) are annotated with a conditional probability distribution (CPD) that represents $p(X_i|Pa(X_i))$, where $Pa(X_i)$ denotes the parents of X_i in G. Thus, the joint probability distribution of G, or in other words, the probability of finding different interrelated events represented in the graph, is the following:

$$p(X_1, \dots, X_n) = \prod_i (p(X_i|Pa(X_i)))$$

In this study, following the Koivisto and Sood approach (Koivisto and Sood, 2004), we use a direct acyclic graph (DAG), a particular type of BN where no cycles are permitted, to find the model with the best goodness of fit to the observed data. The model considered in this study is an additive Bayesian network (ABN), which is essentially different from standard BN for binary or categorical data and based on contingency table parameterization. An ABN model might be described as a multi-dimensional regression model in which all (discrete and continuous) variables are potentially dependent. The method is different from other types of statistical modeling since it aims to discover the optimal graphical model that describes the

interrelationships between multiple and theoretically interdependent variables, identifying the best goodness of fit (based on Bayesian marginal likelihood) of all possible models using an established exact structural algorithm.

Through this method, using the Additive Bayesian network library in R (*abn* package), we attempt to determine the optimal Bayesian network model for the given set of variables (SRH, income, occupational status, education, insider, SPD, gender and age), to identify statistical dependencies that remain hidden in complex or messy data. Therefore, compared to standard BN models based on contingency tables, the additive formulation of these BN can be understood as being equivalent to multivariate generalized linear modeling (GLM), which is specifically aimed at discovering structure, and whose results are computed by determining the most probable (and, therefore, robust) empirical model of causal interdependencies for the variables in the model.

In this model, edges (i.e. links) directed towards the nodes of gender and age have been banned because, theoretically, it is not possible for these socio-demographic variables to be influenced by SES determinants (income, occupational status or education), SPD or SRH. In other words, gender and age can determine other variables in the model, but cannot be determined by these factors.

3. Results

Table 4 shows the bivariate correlation matrix resulting from combining the eight variables in the model (self-rated health -SRH-, income, education, occupational status -ISEI-, insider, self-perceived discrimination -SPD). As can be observed, variables in the model are clearly interrelated and show the expected correlations with each other. Income, education and occupational status are positively associated with SRH. In other words, higher socioeconomic conditions are related to positive health outcomes; in this specific case we could say that upper-class people (i.e. wealthier, more highly skilled or educated) report a positive state of subjective health more often. The same relationship can be applied to the insider variable. Thus, compared to outsiders, being a labor market insider is related to positive health outcomes, although this relationship is weaker than that produced by the previous socioeconomic determinants. Self-perceived discrimination, age and gender (i.e. being female) are negatively correlated with SRH.

Thus these correlations are in line with what could be expected according to the literature. However, our initial question persists: What affects what? Traditionally, studies have presupposed that the socioeconomic situation determines our state of health, but the question that still remains unanswered is: How might health determine socioeconomic status? Despite a certain causality that can be inferred from a vast research corpus on SES determinants of health, correlation cannot be used to infer (inner) interdependencies between variables. In this sense, the existence of correlation between a pair of variables is not a sufficient condition to establish a causal influence in one direction or another. On the other hand, these relationships should be expected to vary between countries, regions and/or welfare state regimes.

Fig. 1 shows five Bayesian networks that show the most probable graphical models to describe the interrelationships between SES and SRH in five European welfare state regimes (see Appendix B for a complete description of coefficients). These welfare state regimes include the following countries: (1) Liberal regime: UK and Ireland; (2) social-democratic regime: Denmark, Finland, Norway and Sweden; (3) Conservative regime: Germany, France, Austria, Belgium, Luxemburg, Netherlands, Switzerland; (4) Mediterranean regime: Cyprus, Greece, Italy, Portugal, Spain and Turkey; and (5) Post-Communist: Bulgaria, Czech Republic, Estonia, Hungary, Poland, Romania, Russian Federation, Slovak Republic, Slovenia and Ukraine (see Table 1). As can be observed in these five models, the relationship between SES variables and SRH is a good deal more complex than standard studies describe. In fact, as we can see in Fig. 1, the most probable associations are susceptible to varying (in either direction) depending on the different characteristics present in welfare state regimes. In every model, we have included two socio-demographic determinants of health that have been included as control variables in the model: gender and age.

The five models show some similarities in the effect of specific variables. Some of these generalities are mentioned in continuation. First, every model must consider the effect of age on education. That is, from this directed relationship we can suppose that age increases the probability of obtaining a higher education (age → education). The second regularity is the effect of age on health (age → SRH) which, in general, is less surprising, since we know that the probability of having good health progressively (and logically) decreases as age increases (i.e. negative correlation between age and health). Finally it is

Table 4
Bivariate correlation matrix.

	SRH	Income	Education	ISEI	Insider	SPD	Gender	Age
SRH	1							
Income	0.158***	1						
Education	0.258***	0.190***	1					
ISEI	0.166***	0.184***	0.535***	1				
Insider	0.075***	0.057***	0.113***	0.249***	1			
SPD	-0.046***	-0.017***	0.023***	-0.012***	-0.008**	1		
Gender	-0.081***	-0.061***	-0.058***	-0.017***	-0.601***	-0.005*	1	
Age	-0.401***	-0.044***	-0.263***	-0.054***	0.001	-0.051***	0.0437***	1

*, **, *** Significant relationship at the level of $p \leq 0.05$, 0.001, and 0.000 respectively.

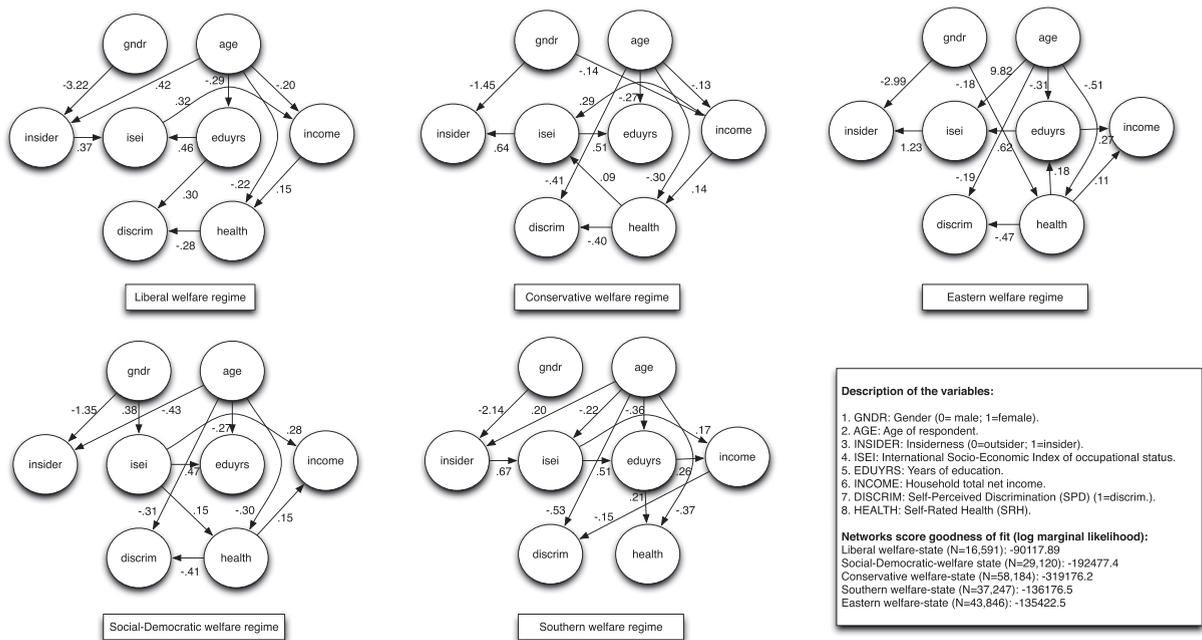


Fig. 1. Bayesian networks describing interrelationships between SES and health in five European welfare states.

relevant to note the negative effect of gender on insiderness. This relationship indicates that gender determines the level of risk that a specific individual undergoes in a concrete labor market (gender → insider), that is, whether the individual may be categorized as an outsider (insecure position in labor market) or an insider (secure position). Specifically, taking into account variable values, this negative relationship indicates that being a woman (1 'female') reduces the probability of being an insider (1 'insider') in European labor markets.

The Liberal model presents clear peculiarities. This model presents an internal structure of data where health seems to be most probably affected by income and age. Of course, the effect of age is less surprising. In fact, as could be expected, the effect of age on health is present in the five models. On the other hand, in this case it is more interesting to perceive that, of the different SES determinants that have been introduced into the model, the effect of income on SRH has stood out (income → SRH), a link that is indirectly defined by age (age → income → SRH; higher age → more income → better health). On the other hand, in this welfare state regime, it is necessary to note the influence of SRH on SPD (SRH → SPD). This finding indicates that in a liberal welfare system, it is more likely for SRH to be affected by economic conditions while SPD is negatively influenced by SRH (i.e. a better state of health reduces the probabilities of perceiving discrimination). In addition, it is interesting to note how occupational status might be an effect (or a product) due to different social determinants (gender → insider → ISEI ← education), while this variable seems to be a direct cause of total net household income (ISEI → income).

The Social-Democratic welfare state regime also presents its own idiosyncrasies. In this model, the variable that is most likely to determine SRH is the ISEI (i.e. the individual occupational status). In other words, subjective general health will probably depend on the position that the individual occupies in the labor market (ISEI → SRH), and not exclusively on economic determinants (e.g. household net income). As in the liberal welfare state system, SPD is found to be affected by SRH (SRH → SPD). In this model, the insider variable is affected both by gender and age, but in this welfare state regime, insiderness does not affect occupational status (i.e. these countries do not present a dualized labor market), while ISEI seems to be a cause of education (ISEI → education).

Thirdly, the Conservative model presents a mixed case between Social-Democratic and Liberal welfare state regimes. As in the Liberal system, SRH seems to be more probably affected by income in the Conservative model (income → SRH) and indirectly by age (age → income → SRH). As in Social-Democratic regimes, SRH is linked to the ISEI and SPD. However, now the association between SRH and ISEI is found to be inverted, that is, subjective general health in this welfare state model seems to be the most likely determinant of individuals' occupational status, whereas the association of health and self-perceived discrimination is the same as that found in the previous model (SRH → ISEI, and SPD), which is a trend that is apparently contrary to what is described in the literature. Finally, in this model, occupational status is found to be more probably associated with insiderness and years of education.

The Mediterranean regime that is composed of Southern European countries presents specific features. In this case, the Southern welfare state regime can be clearly described as a dualized system. Excluding the strong effect of age on other variables (insider, ISEI, education, SPD and SRH), in this model it is interesting to note how gender directly affects (1)

individuals' insidership, and indirectly influences (2) occupational status, (3) education, (4) income, and (5) SRH (through education). In line with previous studies, it is possible to observe that individuals' health in this model is more likely to be related to differences in education (Alvarez-Galvez et al., 2013). Finally, income is found as a determinant of SPD.

Finally, compared to the previously mentioned model, the Eastern welfare regime –corresponding to Post-Communist countries– is probably the least clear. In this welfare system, both age and gender appear to be the main determinants of SRH (age → SRH ← gender), while SRH seems to influence (1) the income level (SRH → income), (2) years of education (SRH → education), and (3) SPD (SRH → SPD). In addition, education is found to be a determinant of occupational status (education → ISEI) and family net income (education → income), while insidership is determined by the ISEI (ISEI → insider).

4. Discussion and conclusion

This study has used a different strategy based on Bayesian Networks (BN) to resolve the gaps in standard approaches when analyzing the relationship between socioeconomic factors and self-rated health. In order to explain the complex association between health and SES determinants, a Bayesian Networks model (Koivisto and Sood, 2004) has been used to find the most probable DAG structure in a network whose interrelationships describe the most probable association links of eight factors in different European regions. Thus, the present study represents a prior step to understand the complex interrelational process between SES determinants and individuals' states of health. To our knowledge, this is the first study that uses a Bayesian Networks approach to offer a self-descriptive and contextualized picture of these complex processes in five European welfare state regimes.

Compared to previous studies, this work presents the advantage of offering a specific picture that describes the complex multi-linkage between social and economic determinants and subjective general health, producing a network that is controlled by socio-demographic factors such as gender and age. Although other studies have focused on describing specific associations between concrete variables and/or the intensity of these relationships in specific countries (Wilkinson and Marmot, 2003; Bartley et al., 2004; Adams et al., 2003a; Frijters et al., 2005; 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), the present work is aimed at providing a general framework to understand the multiple association processes that can emerge from the complex interrelationship of these factors. In fact, compared to standard studies, the biggest advantage of this methodological approach is its strength in graphically clarifying and mapping the most probable paths of relationships from a multi-dependent perspective, where variables can automatically be detected as a cause (independent) and/or an effect (dependent) depending on the internal structure of data. As a result, this method might be used to explore new hypotheses and theoretical constructs that could guide the process of statistical modeling with other predictive techniques for the analysis of dependency (e.g. standard regression models, logistic analysis, or structural equation modeling). Therefore, this exploratory analysis based on a data-driven approach should be seen as a complementary technique that could help us to design and construct subsequent models. Given that 'an apparent relationship between variables may disappear or even be reversed when others are taken into account' (Hand et al., 1997), the present method is a good alternative to avoid incorrect simplifications in the initial analysis of multiple interdependencies within a set of factors.

Obviously, the results obtained are also in line with the findings of previous studies that have used the ESS dataset (Alvarez-Galvez et al., 2013), which certifies the global consistency of these results. For instance, income seems to be the main determinant of health in the Liberal welfare state regime (e.g. United Kingdom) that is characterized by a higher degree of globalization and flexibility in its labor market. Occupational status is the main determinant of individual's state of health in Social-Democratic welfare state regimes (Denmark, Finland, Norway and Sweden) where health inequalities are related to the labor market, not to education or income. The reason for this relationship is probably linked to the universal provision system and therefore to the more egalitarian welfare policies in Nordic countries that could possibly reduce the effect of educational and economic determinants. Meanwhile, in Southern countries (Greece, Italy, Portugal, Spain, etc.), educational inequalities seem to be an evident determinant of health outcomes. And this should be seen as a proof of the validity of the present analysis. Obviously, although there are some regularities that are not surprising for a researcher in this field of study, this work offers a more complex and wider perspective on how specific socioeconomic determinants (income, education and occupational status) are related to subjective general health (SRH), labor market dualization (insider), socio-demographic factors (i.e. gender and age), and thus to self-perceived discrimination (SPD).

According to previous literature, this study shows, on the one hand, that health may be interrelated with diverse and multiple socioeconomic determinants while, on the other, it demonstrates that this relationship may be inverted depending on concrete circumstances associated with specific contexts (e.g. different public policies among welfare state regimes). For instance, health seems to be a determinant of income in Social-Democratic welfare state regimes, and a determinant of occupational status in conservative ones (i.e. general subjective health is more likely to affect income and occupational status). In addition, the impact of health on SES determinants seems to be stronger in Eastern countries (Bulgaria, Estonia, Hungary, Romania, etc.). These specific interrelations indicate that under country-specific conditions, a factor such as SRH, typically defined as a dependent variable in statistical models (e.g. in multivariate logistic regression analysis), may even be defined as an independent or explanatory dimension in our analysis. Now the question is: What conditions define the change in these relationships? Or, in other words, what factors define causes and effects in this relationship?

Clearly, it is difficult to give a single answer to this question. Of course, it would be naive to think that it is possible to find a simple response to explain the emergence of such a complex phenomenon. In fact, from these results we can infer there are as

many answers as singular contexts, and that is the basic reason why the accuracy of our response will depend on the complexity and appropriateness of our model. This analysis has focused on eight specific variables that are relevant for understanding the complex interrelationship between socioeconomic factors and individuals' states of health. Although it would be possible to introduce new variables into the analysis, this basic model provides a great deal of information from a relatively simple network, especially compared to traditional approaches that are mainly focused on (logistic) regression models. The five resulting networks make it possible to confirm the validity of previous findings (e.g. to describe the relevance of income or education, respectively in Liberal and Southern regimes), while, at same time, they identify the most probable structure of relationships, a result that is relevant to understand why certain associations may appear.

For example, countries characterized by a dualized labor market such as Southern European ones, with a dual workforce of insiders and outsiders (Blossfeld et al., 2012), are clearly depicted in the network. This set of connections identifies, on the one hand, the main socioeconomic determinant of health (i.e. education) and, on the other, describes the most probable correlational path between education and health. In our analysis, gender is found to be the initial determinant of the variables related to labor market structure (i.e. insider and occupational status), which becomes the reason behind individuals' education and, ultimately, their income and health. Thus, this interdependency structure is especially plausible since we know from literature on the welfare state that women—as well as young people, older and/or less skilled workers—are identified as outsiders in the divided structure of the labor market in Southern European countries (Esping-Andersen, 1999; Bambra, 2007). In other words, we could say that the more unfavorable position of women in the labor market in these countries may also be related to their worse subjective state of health.

In the case of the Social-Democratic welfare state regime, the absence of correlation between insiderness and occupational status is due to the inexistence of a dualized labor market structure in Nordic European countries. In these countries, generalized access to benefits and services based on citizenship and the promotion of social equity through a generous redistributive social security system impede the link between insiderness and occupational status and, as a result, labor market dualization (Blossfeld et al., 2012). Meanwhile, this relationship is visibly depicted in other welfare systems such as Liberal or Conservative ones. On the other hand, it is interesting to note how individual education might even be determined by occupational status. Although initially this structural relationship could be seen as contradictory, this association might be due to welfare systems characterized by promoting active labor market policies aimed at upgrading education and skills, and thus fostering the employability of the unemployed (Blossfeld et al., 2012). In addition, this relationship could also be an effect of current trends in over-education related to the under-utilization of human capital in globalized labor markets. In other words, it could be thought that workers that cannot find their desired occupation (i.e. a job according to their level of qualification and skills) will continue their education since they are not comfortable with their current position in the labor market (Blossfeld et al., 2012). Although these might be two possible hypotheses for this unexpected relationship, this question will require additional research in future studies.

Finally, the generalized effect of the subjective state of health on self-perceived discrimination should be highlighted. Despite all the attention that the literature has dedicated to the effect that discrimination (especially based on race or ethnicity) produces on health outcomes, our analysis points out that this relationship may be inverted in practice. In fact, it is possible that those individuals with a poor state of health (both subjective and/or objective), and probably lower self-esteem (Corning, 2002), were more susceptible to perceiving a certain type of discrimination related to personal stigmas (either based on gender, sexuality, physical or intellectual disabilities, etc.). Unfortunately, empirical studies that describe the influence of psychological distress on self-perceived discrimination have appeared only recently (Corning, 2002; Fischer and Shaw, 1999; Waldo, 1999), so future studies should pay particular attention to this relationship.

Thus, it is interesting to observe that the SPD variable does not determine any other variables in the model, although in other previous studies this indicator has been found to be a fundamental determinant of health (Kim and Williams, 2012; Alvarez-Galvez and Salvador-Carulla, 2013). In the present model, SPD and SRH are also related, but the path of influence is inverted. Although this question should be studied in greater depth in future work, this alternative result might be related to the lower prevalence of respondents reporting self-perceived discrimination in the final samples [$<0.10\%$]. In addition, it is difficult to find a concrete explanation for the effect of education on SPD in the liberal welfare state regime. This association might be related to the idiosyncrasies of an educative system based on market dominance, which could generate differences in skill among graduates who can or cannot afford the cost of studies, which could ultimately increase social stratification. Hypothetically, a division between skilled and unskilled students could affect the discrimination perceived in these groups. However, this is a question that cannot be answered by this exploratory study, and in the future will require further research.

At methodological level, although we could have used other techniques for the study of dependency, such as multiple regressions, structural equation models (SEM) or standard Bayesian networks, these methods were rejected because they are more appropriate for a theory-driven approach. For example, in SEM the researcher defines the structure of the relationships in the model, so the final result might be clearly biased by previously-known theoretical assumptions (Sorooshian et al., 2011). However, if we want to find alternative explanations directly based on empirical data, the additive Bayesian Networks (ABN) approach could be more appropriate as a prior step to the modeling of possible directional associations between socioeconomic determinants and health. Thus it may be more useful to understand the latent structure of correlations before describing the relationship and intensity of these multiple interdependencies. This approach permits the automatic identification of the dependency structure in a previously specified set of attributes and a minimum set of rules (e.g. in our model, sex and age cannot be affected by other determinants). The number of predefined rules (i.e. arcs imposing a specific direction

in the dependency structure) will depend on our previous knowledge of the data (Lewis and McCormick, 2012). Ultimately, results can be validated by expert information and evidence-based knowledge.

Finally, it is necessary to mention a clear limitation in this study, which is particularly related to the selection of welfare state regimes. This study has identified five welfare state models, although this classification may be problematic. In fact, some regimes could be questioned due to their specific character (e.g. Eastern/Post-Communist models); others with similar characteristics could be added depending on the literature (e.g. Central and Southern European regimes as the Conservative model); others could be included (e.g. Catholic countries) (Esping-Andersen, 1999; Bamba, 2007). Consequently, taking into account this weakness, future studies should aim to improve this classification, trying out other possibilities to increase the variability of the final model and, at the same time, to identify countries' idiosyncrasies. In line with this consideration, in future studies it could be interesting to use primary indicators (e.g. alternative measures of insidership or occupational status which do not capture the effect of several variables), but also to compare the isolated effect of these variables by acting out different scenarios of multiple interdependency.

Acknowledgments

The working group of research on health and social policy at Loyola University Andalusia.

Appendix A. Multi-collinearity test

Variable	VIF	SQRT VIF	Tolerance
Age	1.21	1.10	0.82
Gndr	1.48	1.22	0.68
Insider	1.57	1.25	0.64
SRH	1.12	1.06	0.89
ISEI	1.43	1.19	0.70
Education	1.41	1.19	0.71
Income	1.22	1.10	0.82
SPD	1.02	1.01	0.98

Appendix B

Table 1

Bayesian network model for Liberal regime.

Dependent	Intercept	Age	Gender	Insider	Health	Isei	Eduyrs	Income	Discrim.
Age	-1.90E-17								
Gender	0.122242								
Insider	2.881699	0.4278728	-3.2224542						
Health	-5.18E-14	-2.22E-01						1.47E-01	
Isei	-0.2453007			0.3697984			0.4634802		
Eduyrs	1.05E-15	-2.85E-01							
Income	-3.92E-15	-2.01E-01				3.24E-01			
Discrim.	-2.28E+00				-0.2788579		0.3014557		

Table 2

Bayesian network model for Conservative regime.

Dependent	Intercept	Age	Gender	Insider	Health	Isei	Eduyrs	Income	Discrim.
Age	-2.39E-15								
Gender	0.0253064								
Insider	13.585396			-1.4513213		0.6384464			
Health	-4.08E-13	-3.02E-01						1.41E-01	
Isei	5.21E-15				9.27E-02			2.86E-01	
Eduyrs	1.69E-14	-2.70E-01				5.09E-01			
Income	7.25E-02	-1.26E-01	-0.1431155						
Discrim.	-2.77E+00	-0.4140525			-0.3962382				

Table 3

Bayesian network model for Liberal regime.

Dependent	Intercept	Age	Gender	Insider	Health	Isei	Eduyrs	Income	Discrim.
Age	2.41E-15								
Gender	0.05429513								
Insider	2.721338		-2.994487			1.232341			
Health	9.02E-02	-5.08E-01	-0.1755917						
Isei	-8.77E-16	9.82E+00					6.20E-01		
Eduyrs	-5.43E-15	-3.05E-01			1.76E-01				
Income	1.02E-14				1.14E-01		2.73E-01		
Discrim.	-2.77E+00	-0.1866903			-0.467299				

Table 4

Bayesian network model for Social-Democratic regime.

Dependent	Intercept	Age	Gender	Insider	Health	Isei	Eduyrs	Income	Discrim.
Age	-1.02E-14								
Gender	0.0021564								
Insider	13.0525266	-0.4309713	-1.3475471						
Health	1.20E-13	-2.99E-01				1.53E-01			
Isei	-8.24E-01		0.3736488						
Eduyrs	-6.83E-15	-2.67E-01					4.66E-01		
Income	-2.81E-14				1.46E-01		2.77E-01		
Discrim.	-2.91E+00	-0.3062643			-0.4146724				

Table 5

Bayesian network model for Southern regime.

Dependent	Intercept	Age	Gender	Insider	Health	Isei	Eduyrs	Income	Discrim.
Age	1.39E+15								
Gender	-0.1680115								
Insider	2.0060198	0.1966246	-2.1388335						
Health	-2.78E-15	-3.65E-01					2.15E-01		
Isei	-4.64E-01	-0.2231357		0.6730209					
Eduyrs	-1.01E-14	-3.64E-01				5.14E-01			
Income	4.70E-15					1.70E-01	2.62E-01		
Discrim.	-2.79E+00	-0.5272143							-0.1456604

References

- Adams, P., Hurd, M.D., McFadden, D., Merrill, A., Ribeiro, T., 2003. Health, wealthy and wise. Tests for direct causal paths between health and socioeconomic status. *J. Econ.* 112, 3–56.
- Adams, P., Hurd, M.D., McFadden, D.L., Merrill, A., Ribeiro, T., 2003. Healthy, wealthy, and Wise? Tests Direct Causal Paths between Health Socioecon. Status. *J. Econ.* 112, 3–56.
- Alvarez-Galvez, J., Salvador-Carulla, L., 2013. Perceived discrimination and self-rated health in Europe: evidence from the European social survey (2010). *PLoS ONE* 8 (9), e74252. <http://dx.doi.org/10.1371/journal.pone.0074252>.
- Alvarez-Galvez, J., Rodero-Cosano, M.L., Motrico, E., Salinas-Perez, J.A., Garcia-Alonso, C., Salvador-Carulla, L., 2013. The impact of socioeconomic status on self-rated health: study of 29 countries using European social surveys (2002–2008). *Int. J. Environ. Res. Public Health* 10 (3), 747–761.
- Bambra, C., 2007. Going beyond the three worlds of welfare capitalism: regime theory and public health research. *J. Epidemiol. Commun. Health* 61, 1098–1102.
- Bartley, M., Sacker, A., Clarke, P., 2004. Employment status, employment conditions, and limiting illness: Prospective evidence from the British household panel survey 1991–2001. *J. Epidemiol. Community Health* 58, 501–506.
- Bleakley, H., 2003. Disease and Development: evidence from the American South. *J. Eur. Econ. Assoc.* 1, 376–386.
- Bloom, D., Canning, D., 2008. Population Health and Economic Growth. Working Paper no.24, for the Commission on Growth and Development (Washington, DC).
- Blossfeld, H.-P., Buchholz, S., Hofäcker, D., Kolb, K. (Eds.), 2012. *Globalized Labour Markets and Social Inequality in Europe*. Palgrave-MacMillan, pp. 79–107.
- Caillet, P., Klemm, S., Ducher, M., Aussem, A., Schott, A.-M., 2015. Hip Fracture in the Elderly: a Re-analysis of the EPIDOS study with causal bayesian networks. *PLoS ONE* 10 (3), e0120125.
- Corning, A., 2002. Self-esteem as a Moderator between perceived discrimination and psychological distress among women. *J. Couns. Psychol.* 49 (1), 117–126.
- Costa-Font, J., Hernández-Quevedo, C., 2012. Measuring inequalities in health: what do we know? what do we need to know? *Health Policy* 106, 195–206.
- Eikemo, T.A., Bambra, C., Joyce, K., Dahl, E., 2008. Welfare state regimes and income-related health inequalities: a comparison of 23 European countries. *Eur. J. Publ. Health* 18, 593–599.
- Eikemo, T.A., Bambra, C., Judge, K., Ringdal, K., 2008. Welfare state regimes and differences in self-perceived health in Europe: a multilevel analysis. *Soc. Sci. Med.* 66, 2281–2295.
- Esping-Andersen, G., 1999. *Social Foundations of Postindustrial Economies*. Oxford University Press.
- ESS Round 1-4: European Social Survey Round 1-4 Data (2002-2008). Norwegian Social Science Data Services, Norway—Data Archive and distributor of ESS data.
- Fischer, A.R., Shaw, C.M., 1999. African Americans' mental health and perceptions of racist discrimination: the moderating effects of racial socialization experiences and self-esteem. *J. Couns. Psychol.* 46, 395–407.

- Frijters, P., Haisken-DeNew, J.P., Shields, M.A., 2005. The causal effect of income on health: evidence from German reunification. *J. Health Econ.* 24, 997–1017.
- Ganzeboom, H.B.G., Donald, J.T., 1996. Internationally comparable measures of occupational status for the 1988 International Standard Classification of Occupations. *Soc. Sci. Res.* 25, 201–239.
- Hand, D.J., McConway, K.J., Stanghellini, E., 1997. Graphical models of applicants for credit. *IMA J. Maths Appl. Bus. Ind.* 8, 143–155.
- Hanibuchi, T., Nakaya, T., Murata, C., 2012. Socioeconomic status and self-rated health in East Asia: a comparison of China, Japan, South Korea and Taiwan. *Eur. J. Public Health* 22 (1), 47–52.
- Häusermann, S., Schwander, H., 2012. Varieties of Dualization? Labor market segmentation and insider-outsider divides across regime. In: Patrick, E., Häusermann, S., Palier, B., Seeleib-Kaiser, M. (Eds.), *The Age of Dualization. The Changing Face of Inequality in Deindustrializing Societies*. Oxford University Press, Oxford and New York.
- Heckerman, D., Geiger, D., Chickering, D.M., 1995. Learning Bayesian networks—the combination of knowledge and statistical data. *Mach. Learn.* 20 (3), 197–243.
- ILO (International Labour Office), 1990. *International Standard Classification of Occupations : ISCO-88*. International Labour Office, Geneva.
- Kim, S.-S., Williams, D.R., 2012. Perceived discrimination and self-rated health in South Korea: a Nationally Representative survey. *PLoS ONE* 7 (1), e30501.
- Koivisto, M., Sood, K., 2004. Exact Bayesian structure discovery in Bayesian networks. *J. Mach. Learn. Res.* 5, 549–573.
- Kunst, A.E., Bos, V., Lahelma, E., Bartley, M., Lissau, I., Regidor, E., Mielck, A., Cardano, M., Dalstra, J.A., Geurts, J.J., Helmer, U., Lennartsson, C., Ramm, J., Spadea, T., Stronegger, W.J., Mackenbach, J.P., 2005. Trends in socioeconomic inequalities in self-assessed health in 10 European countries. *Int. J. Epidemiol.* 34, 295–305.
- Lewis, F.I., McCormick, B.J.J., 2012. Revealing the complexity of health determinants in resource-poor Settings. *Am. J. Epidemiol.* 176 (11), 1051–1059.
- Lewis, F.I., Ward, M.P., 2013. Improving epidemiologic data analyses through multivariate regression modelling. *Emerg. Themes Epidemiol.* 10 (1), 4. <http://dx.doi.org/10.1186/1742-7622-10-4>.
- Needham, C.J., Bradford, J.R., Bulpitt, A.J., et al., 2007. A primer on learning in Bayesian networks for computational biology. *PLoS Comput. Biol.* 3 (8), pe129.
- Nummela, O.P., Sulander, T.T., Heinonen, H.S., Uutela, A.K., 2007. Self-rated health and indicators of SES among the ageing in three types of communities. *Scand. J. Publ. Health* 35, 39–47.
- Smith, J.P., 1999. Healthy Bodies and Thick Wallets: The dual Relation between health and economic status. *J. Econ. Perspect.* 13 (2), 145–166.
- Sorooshian, S., Norzima, Z., Rosnah, M.Y., Yusof, I., 2011. Bias of structural equation modeling. *J. Comput. Model.* 1 (2), 157–160.
- Waldo, C.R., 1999. Working in a majority context: a structural model of heterosexism as minority stress in the workplace. *J. Couns. Psychol.* 46, 218–232.
- Wilkinson, R.G., 1996. *Unhealthy Societies: the Afflictions of Inequality*. Routledge, London, UK.
- Wilkinson, R., 1997. Health inequalities: Relative or absolute material standards. *Br. Med. J.* 314, 591–595.
- Wilkinson, R., Marmot, M., 2003. *Social Determinants of Health: the Solid Facts*. World Health Organization, Copenhagen, Denmark.
- Yannis, G., Nicholas, T., Ping, Y.P., 2009. Personal values as mitigating factors in the link between income and life satisfaction: evidence from the European Social Survey. *Soc. Indic. Res.* 91 (3), 329–344.