Non-communicable diseases, socio-economic status, lifestyle and well-being in Italy: An additive Bayesian network model

Malattie croniche, status socio-economico, stile di vita e benessere: un approccio basato sui Modelli di Rete Bayesiani additivi

Laura Maniscalco and Domenica Matranga

Abstract The aim of the paper is to investigate the statistical association, on a sample of Italian subjects, extracted by Survey of Health, Ageing and Retirement in Europe (SHARE) dataset, between chronic diseases (occurrence or number of chronic diseases) and socio-economic and behavioural determinants (lifestyle indicators, QoL indicators, cognitive functioning variables). To this aim, additive Bayesian network (ABN) analysis was used. The resulting ABN model shows that better-educated individuals have better health outcomes, age is direct and gender is an indirect determinant of the number of chronic diseases. Furthermore, self-perceived health is associated with lower number of chronic diseases, lower physical limitations and higher quality of life and these indicators can be considered within a unitary vision to represent well-being of elderly people, as they share a similar distribution by gender and age.

Abstract *Lo scopo dello studio è stato quello di verificare, con dati su un campine italiano, l'associazione statistica tra la malattia cronica ed i determinanti che rigurdano lo status socio-economico, gli indicatori dello stile di vita e della qualità di vita e la funzione cognitiva. Per raggiungere lo scopo posto, sono stati usati i modelli di rete Bayesiani additivi. Il modello ABN scelto mostra che gli individui più istruiti hanno una migliore salute, tenendo costante le altre esplicative, l'età è un fattore determinante diretto e il sesso è un fattore determinante indiretto del numero di malattie croniche. Inoltre, la salute percepita è associata a un minor numero di malattie croniche, a minori limitazioni fisiche e ad una migliore qualità di vita, questi indicatori possono essere considerati nel loro insieme perché rappresentano il benessere degli anziani e condividono una distribuzione simile per genere ed età.*

Laura Maniscalco
Università degli Studi di Palermo , Dipartimento di Chirurgia Neurosensoriale e Motoria, Medicina Orale con Odontoiatria per pazienti a rischio, e-mail: maniscalco.laura92@gmail.com

Domenica Matranga
Università degli Studi di Palermo, Dipartimento di Scienze per la Promozione della Salute e Materno-Infantile "G. D'Alessandro" e-mail: domenica.matranga@unipa.it

Key words: GLM, Additive Bayesian Network, lifestyle, well-being

1 Introduction

In Italy, the main non-communicable diseases (NCDs) all together account for heavy disease burden, due to the increasing population aging. Data from the Italian National Institute of Statistics in 2015 show that the mortality rate for cardiovascular diseases is 512 per 100000 people, while mortality for other diseases is lower but still worrying (294 per 100000 for cancer, 137 per 100000 for chronic respiratory diseases and 82 per 100000 for mental disorders). The WHO NCD global surveillance strategy is based on a multidimensional view of NCD determinants, including physiological and lifestyle influences and environmental and social factors. People with low socioeconomic status (SES) have less access to NCD care and treatment, are less aware of correct lifestyles to hamper the onset of NCDs and to prevent advanced-stage disease and complications. Some studies show that poor living conditions and primary education are associated with physical inactivity [7] and that there is an inverse social gradient for the feminine obesity [4]. Other studies show the positive impact of education on the consumption of healthy nutrients and on the reduction of individual body-mass index [3]. The majority of chronic diseases affects the overall health of patients by limiting their well-being, the functional status, productivity and health-related quality of life. The main limitation of psychosocial well-being regards the minor involvement in social activities that implicates a reduction of positive reinforcement [5]. A classical approach to investigate the statistical association between chronic diseases (presence or not of chronic diseases) and covariates of interest (SES variables, lifestyle indicators, QoL indicators, cognitive functioning variables) is based on generalized linear models, which make a net classification of variables into covariates and the response. In a multifactorial complex disease system, such that one of NCDs, it should be desirable to analyze the associations between all covariates with all variables being potentially dependent, using Additive Bayesian Networks (ABN) [10]. In order to explain this interrelationship, a data-driven approach using Additive Bayesian Network was applied to find the most probable structure.

2 Material and Methods

2.1 The sample

For the purpose of our study, it was analysed the Survey of Health, Ageing and Retirement in Europe (SHARE) dataset. This dataset is a multidisciplinary and cross-national panel database of micro data on health, socio-economic status and social and family networks on 27 European countries and Israel. In particular, for this analysis we focused on Italian data. The dataset contains, 5288 observations and 17 variables, after having removed missing values. Categorical variables with multiple categories have been transformed into binary, in order to analyze the data with the ABN methodology.

2.1.1 Dataset structure

The variables under study are: gender with two categories “Male ” and “Female ”, age of the respondent, years of education (yedu), marital status (mstat) with categories “Married” and “Not Married”, household total net income (thinc), household net worth (hnetw), current job status (cjs), with categories “Labor force”and “Not labor force”, Body Mass Index (BMI) of the respondent, smoking (esmoked), with categories “Yes” and “No”, physical inactivity (phinact), with categories “Yes” and “No”, number of chronic disease (chronic), US version of self-perceived health (sphus), with categories “Less than good” and “More or equal than good”, global activity limitation (gali), a binary variables with categories “Limited” and “Not Limited”, score of verbal fluency test (fluency), score regarding “first trial ten words list learning” test (cf008tot), score regarding “delayed trial ten words list learning” test (cf016tot), quality of life in older ages (casp), ranging between 0 to 12. Fluency, cf008tot and cf016tot measure subjective and objective aspects of the respondent’s cognitive function, like memory and verbal fluency.

2.2 Statistical methods

Descriptive statistics were carried out. Continuous variables were summarized with mean, median and standard deviation. Categorical variables were analysed with frequencies distributions. Additive Bayesian network (ABN) analysis was used to identify factors associated with the non-communicable disease. Bayesian networks (BNs) belong to the family of probabilistic graphical models (GMs). These models represent a set of variables and their conditional dependencies through a directed acyclic graph (DAG). They are enabled to represent the joint probability distribution (JDP) over a set of random variables. The structure of a DAG is represented by a set of nodes that represent random variables and a set of edges visualized by arrows between nodes that are direct dependencies among variables. Additive Bayesian networks (ABN) are a special type of BN models, where the parameters instead to be based on contingency table, are based on generalized linear model (GLM). Each node in the DAG is the equivalent to the response variable in a GLM. To all DAG was imposed a uniform prior to allow a full data-driven approach. The identification of the best DAG was doing with an exact search method and the identification of the best DAG was based on the marginal log-likelihood. The marginal log-likelihood represents the goodness of fit metric in Bayesian modeling and it includes an implicit penalty for model complexity. The first step to identify the best DAG consists in increasing the maximum number of parents allowed per node (that are the number of allowed covariates in each model) until the goodness of fit remained constant. The model selection procedure started from one to twelve possible parents per node. The best DAG was identified, with eleven maximum number of possible parents per node. The second step regards the model adjustment for over-fitting, that usually is doing when the sample is small and there are many variables. In this step before to generate dataset with MCMC simulation it useful check if the area under the curve of the posterior density integrates to one. In the third step, the marginal posterior log odds ratio was estimated for each parameter from the integration of the posterior distribution with respect to the set of parameters. The maximum likelihood estimates were obtained by the joint posterior distribution. Since ABN methodology allows to evaluate the association between all variables, an arc between two variables in the final ABN model indicates a "direct" relationship, whereas an "indirect" relationship is defined as a relationship between two variables through an intermediate variable. Data were analyzed with R software (version 3.3.2), and the ABN methodology was implemented with "abn" package [11].

3 Results

The exploratory data analysis revealed that there is a strong correlation between cf008tot and cf016tot (p=0.72) and a negative correlation between age and cf008tot (r=-0.39), a positive correlation between chronic and age (r=0.35) and between years of education and cf008tot. All other quantitative variables show a correlation < 0.33. By exploring the scatter-plot of the marginal log-likelihood, the optimal number of parents allowed was found to be six. The optimal DAG (see Fig. 1), showed that the number of chronic disease showed a direct association with BMI (coeff= 0.13), years of education (coeff= -0.073) and age (coeff= 0.12) and an indirect association with gender (coeff= 0.26). Very interestingly, gender is a direct determinant of smoking status. Relating to self-perceived health, the DAG shows direct associations with chronic diseases (coeff= -0.52), cognitive function indicator (cf008 coeff=0.07, cf016tot coeff= 0.29 and fluency coeff=0.04), global activity limitations (coeff= 1.99), quality of life (coeff= 0.12) and with socio-behavioural determinants (marital status coeff=-0.36, working status coeff= -0.65 and physical inactivity coeff= -0.63).

4 Discussion

Additive Bayesian networks analysis of the Share dataset confirms the multidimensional approach of NCD determinants which is suggested by WHO [8]. According to this framework, all determinants are distributed along four successive causation levels, going from physiological influences to social structure, passing through lifestyle and environmental influences. Among physiological and lifestyle influences, our study found out BMI as direct and smoking as indirect determinant of the number of chronic diseases in line with other literature [2, 6]. Within the social structure, another important finding was the role of education and age as direct determinants and of gender as indirect determinant of the number of chronic diseases. Better-educated individuals obtain better health outcomes from a fixed set of inputs because they have the abilities and information to make better choices for their lifestyles. Education indirectly facilitates individual development and interpersonal relationships, enabling people to pursue personal and professional success, which has a positive impact on health [1]. Another finding of our study is that better self-perceived health is associated with lower number of chronic diseases, lower physical limitations and higher quality of life. All these indicators can be considered within a unitary vision to represent well-being of elderly people, as they share a similar distribution by gender and age. In fact, a higher proportion of males yield positive

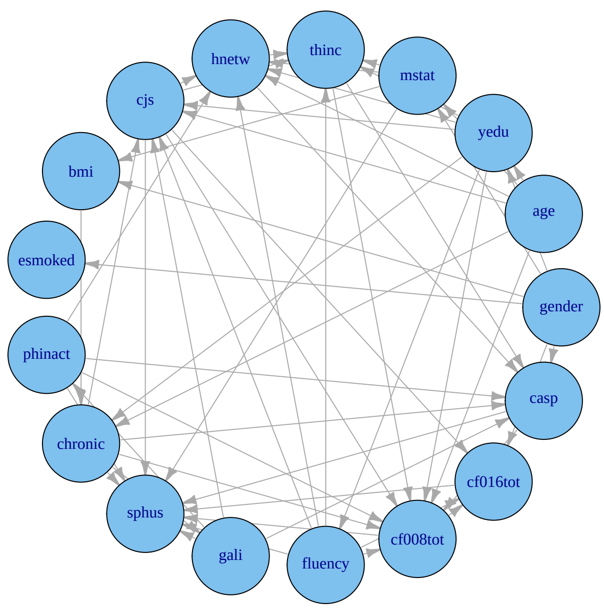

Fig. 1 Optimal ABN model, in terms of marginal likelihood, with a maximum number of six parents per node

outcomes on psychological distress, self-rated health, life satisfaction and chronic diseases and a slightly higher percentage of females display better outcomes regarding their quality of life and BMI. Furthermore, a higher level of well-being appeared to be prevalent among elderly people embedded more frequently in social activity participation [9]. One strength of our study relates to the statistical method used. Additive Bayesian networks supply a general framework to understand the multiple association process that can emerge from the complex interrelationship among health indicators, socio-economic and behavioural determinants. Moreover, through prior elicitation, it should be possible to embed all available information about the statistical association among them. Finally, the data-driven approach could help us to formulate new research hypotheses and to design and construct subsequent theoretical models.

References

1. Di Cesare, M. et al.: Inequalities in non-communicable diseases and effective responses. The Lancet, Elsevier **381**, 585–597 (2013)
2. Giampaoli, S. and Palmieri, L. et al.: Cardiovascular health in Italy. Ten-year surveillance of cardiovascular diseases and risk factors: Osservatorio Epidemiologico Cardiovascolare/Health Examination Survey 1998–2012. Eur J Prev Cardiol, SAGE Publications Sage UK: London, England **22**, 9–37 (2015)
3. Matranga, D. and Tabacchi, G. and Cangialosi, D.: Sedentariness and weight status related to SES and family characteristics in Italian adults: exploring geographic variability through multilevel models. Scandinavian journal of public health, SAGE Publications Sage UK: London, England, (2017)
4. McLaren, L.: Socioeconomic status and obesity. The Oxford handbook of the social science of obesity. http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199736362.001.0001/oxfordhb-9780199736362-e-016. Accessed 23 September 2016
5. Megari, K.: Quality of life in chronic disease patients. Health Psychology Research, PAGEPress **1**, (2013)
6. Panico, S. and Palmieri, L. et al.: Preventive potential of body mass reduction to lower cardiovascular risk: the Italian Progetto CUORE study. Prev Med, Elsevier **47**, 53–60 (2008)
7. van Der Berg, J.D. and Bosma, H. et al.: Midlife determinants associated with sedentary behavior in old age.. Med Sci Sports Exerc, NIH Public Access **46**, 1359 (2014)
8. World Health Organization: STEPS: A framework for surveillance. (2003) Available at $http://www.who.int/ncd_surveillance$
9. Vozikaki, M. and Linardakis, M. et al.: Activity participation and well-being among European adults aged 65 years and older. Social Indicators Research, Springer **131**, 769–795 (2017)
10. Lewis, F.I. and McCormick, B.:Revealing the complexity of health determinants in resource-poor settings. American journal of epidemiology **176**, 1051–1059 (2012)
11. Pittavino, M. and Lewis, F. and Furrer, R.: an R package for modelling multivariate data using additive Bayesian networks. The Comprehensive R Archive Network (CRAN), 1–37, (2016)