USING R FOR THE ANALYSIS OF SEVERE COGNITIVE IMPAIRMENT

Kornélia Cséfalvaiová

Abstract

This article includes the analysis of risk factors of severe cognitive impairment, based on socio-demographic and health variables from the database SHARE (Survey of Health, Ageing and Retirement in Europe). The R – Statistical Software and also the method of logistic regression were used for the analysis of risk factors of severe cognitive impairment. The aim is to highlight the main risk factors in order to pay higher attention to the prevention and control of the disease. Ongoing dramatic increase in the number of people with cognitive impairment will put high demands on health and social care in the Czech Republic. Population aging and the increase of elderly persons aged 80+ evoked a need to address this issue, since age is the major risk factor for Alzheimer's disease, dementia and severe cognitive impairment. Currently there has been little attention paid to statistical methods and approaches that may contribute to early identification of dementia and cognitive disorders. In this research logistic regression helps to explain important variables associated with severe cognitive impairment in the Czech Republic.

Key words: Ageing, Czech Republic, Severe Cognitive Impairment, Risk Factors, Logistic Regression.

JEL Code: I1, J14

Introduction

Population ageing is an ongoing demographic process that is closely associated with the health status of the elderly. The more older people are in the population, the higher is the share of old people suffering from diseases that are affected by age. To these diseases definitelly belong dementia and different types of cognitive impairment. It is generally known that some disease, e.g., diabetes, cardiovascular disease or poor physical and mental condition, increase the risk of occurrence dementia and severe cognitive impairment. The situation is complicated by the fact that the individual may suffer at the same time at more than one simultaneously disease:

diabetes, hypertension or heart disease. Equally important is appreciated that not all AIDS patients with a given disease visit the practitioner and are introduced into the statistics. Therefore, a number of diseases which are characterized by, but not limited too course of the patient, e.g., Elevated blood pressure, it can be seen only very roughly. One approach to solving this problem is to try to model development morbidity from chronic disease on the basis of knowledge of the risk factors.

The source of data used is the SHARE database (The Survey of Health, Aging and Retirement in Europe), which by its multidisciplinary nature provides a comprehensive picture of the aging process in Europe. The results in this article are of significance with respect to the issue of dementia useful material for future analysis of different areas of research (medicine, statistics, social care, demography). According to projections for the Czech Republic in 2050, the number of seniors aged 65-74 years will increase by 50%, the number of seniors aged 75-84 years will grow by almost 94% (Stejskal, 2014). It is highly urgent to take into account the high current and future treatment costs of dementia and other mental diseases (Cséfalvaiová, Stejskal, 2016).

1 SHARE − The Survey of Health, Ageing and Retirement in Europe

For the analysis of risk factors that are associated with dementia and severe cognitive impairment in the Czech Republic the database called SHARE (The Survey of Health, Ageing and Retirement in Europe) was used. Among the main topics of multidisciplinary research include demography, family, education, physical and mental health, cognitive function, medical care and risks, quality of life, employment and income, housing, income and consumption of households, social support, etc. Data set SHARE provides full advice socio-demographic variables, variables relating to lifestyle and physical and mental health, which help to elucidate acting factors. The investigation so far to the 5 waves in different European countries, including Czech Republic. It was on a panel database of microdata from the area of the economic situation, health, social and family bonds. It provides real-tracking data on a sample of 123 thousand individuals in 27 European countries and Israel older than 50 years. Czech Republic was involved in the project in the 2ⁿᵈ wave of investigations in 2006. The variables characterizing the state of physical and mental health and variables from which it was possible to calculate a variable cognitive function, found only in the 2ⁿᵈ , 4th and 5th wave and therefore we used in this article only data from these waves.

2 Determinants of occurrence of severe cognitive impairment

Generally in the literature,there were introduced several risk factors of dementia. From these results it can be assumed that higher education and active lifestyle reduce the chance of developing dementia. Furthermore, some diseases such as diabetes and cardiovascular diseases or poor physical and mental health should generally increase the chances of developing dementia. The aim of this part is the analysis and identification of factors that affect the risk of severe cognitive impairment in the Czech Republic. Researchers question is whether there are any assumptions or risk factors, which when exposed to a certain person more frequently, thereby increasing their chances of developing a cognitive disorder? Admission variables related to socio-demographic characteristics, physical and mental health and lifestyle were drawn from the SHARE (Börsch-Supan, 2013), which was described in the section 1 of this article.

2.1 Model of logistic regression using R – health variables

Multi-dimensional analysis can exclude relationships that exist between the explanatory variables. To determine associations between basic demographic characteristics and other variables, and severe cognitive impairment model was constructed logistic regression:

The model included the following variables: sex (male / female), age, comprising four age groups (<60 years, 60-69 years, 70-79 years, 80+), education are two groups of education (elementary+secondary+others and tertiary) partner (living with a partner / without a partner), and other variables associated with the medical condition: weight, smoking, alcohol and selected diseases. The estimated regression model was created in the R statistical program (www.r-project.org). Results of logistic regression model in this work will be further interpreted according to odds ratio (OR). According to the results it is evident that the chance of the development of severe cognitive impairment is strongly influenced by high age. With high ages more people are at risk of the disease. In subsequent models, the effect of age also confirmed - independently of the other variables included in the model, the chances of occurrence of severe cognitive impairment increases rapidly with age. In the first model, women have a lower chance

of developing severe cognitive impairment (OR = 0.24). Similar results were confirmed e.g. for Austria, Sweden, the Netherlands (Ziegler, 2010). It is important to note that the results associated with character input data for that country. Results for Germany, Spain, Italy demonstrate a higher chance occurrence of disorders for women (Ziegler, 2010). After the inclusion of education variable in the model was shown a protective effect of levels of educational attainment. With higher education chances of occurrence of severe cognitive impairment decreases (for tertiary education OR is 0.59). Further, from the results it is predicted that the chances of developing disorder affects also the fact whether the person is living or living with partner (Living without a partner: OR = 5.10). Moving away from the balance of normal body weight, whether it is a weight increase or decrease may be due to a general disease that can affect the cognitive functions as well.

In the research there were used alltogether 4 models of logistic regression, but in this article we show only one model of logistic regression (see below the model 1 of logistic regression that analyses health variables, including the main demographic variables: gender, age, education and partnership).

Model 1: Analyses of risk factors of dementia and severe cognitive impairment – health variables

$$\ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \gamma * gender + \sum_{i=1}^{4} \delta_i * age_i + \sum_{j=1}^{3} \varepsilon_j * education_j + \theta * partner + \sum_{k=1}^{4} \mu_k$$
$$* weight_k + \sum_{l=1}^{2} \varepsilon_l * smoking_l + \sum_{m=1}^{7} \sigma_m * drinking_m + \sum_{n=1}^{7} \tau_n * illnesses_n$$

The results of the analyses and the model of the logistic regression for health variables are seen on the Figure 1:

Fig. 1: ROC Curve
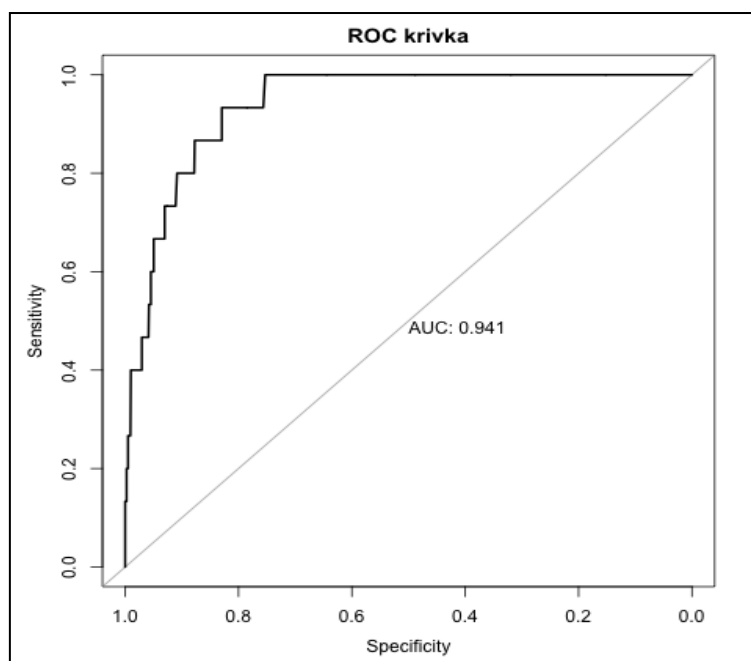
Source: own calculations

ROC Curve and the value of AUC (Area Under Curve) is equal to 0.941 that means that the selected variables well fit the model of logistic regression (using R software).

Conclusion

Age was shown (like within the literature) as the major risk factor of severe cognitive impairment. The risk of severe cognitive impairment increases with age, some studies have suggested that the highest age groups is slower increase. Factors, such as lifestyle can prevent and slow down the development of cognitive disorders. Lifestyle is also affected by the presence of other diseases such as hypertension, diabetes, heart attack, vascular diseases that are associated with severe cognitive impairment. It can be concluded that healthy diet reduces the risk of developing severe cognitive impairment, both directly and indirectly. There is no direct correlation between the different pathologies and pursued the development of severe cognitive impairment, but generally it can be said that a combination of factors, the monitored increases the likelihood of its development. A higher level of education and healthy lifestyle appear to be the factors which delay disease incidence in the higher age categories. Living with a spouse/living without a partner appeears to be more important that it was considered before. Using logistic regression is a widely and frequently used statistical method when we want to see the associations among variables. In the medicine and analysis of health data logistic regression (and statistical software R) is the most frequently used statistical method. Many

authors used the R software in the research (see e.g. Šimpach, 2014) and also in other analysis (see e.g. Malá, 2016 ; Bašta, 2016).

Acknowledgment

References

BAŠTA, M. (2016). A wavelet-based approach to breakpoint detection. In: *10th International Days of Statistics and Economics* (pp. 123-133). ISBN 978-80-87990-02-5. https://msed.vse.cz/msed_2016/article/213-Basta-Milan-paper.pdf

BÖRSCH-SUPAN, A., A. BRUGIAVINI, H. JÜRGES, A. KAPTEYN, J. MACKENBACH, J. SIEGRIST AND G. WEBER. (2008). First results from the Survey of Health, Ageing and Retirement in Europe (2004-2007).Starting the longitudinal dimension. Mannheim: Mannheim Research Institute for the Economics of Aging (MEA).

BÖRSCH-SUPAN, A., BRANDT, M., HUNKLER, C., KNEIP, T., KORBMACHER, J., MALTER, F., SCHAAN, B., STUCK, S., ZUBER, S. (2013). Data Resource Profile: The Survey of Health, Ageing and Retirement in Europe (SHARE). International Journal of Epidemiology DOI: 10.1093/ije/dyt088.

CSÉFALVAIOVÁ, K., STEJSKAL, J. (2016). *Statistical methods used for identifying dementia and mild cognitive impairment*. In: *10th International Days of Statistics and Economics* (pp.286-293). ISBN 978-80-87990-02-5. https://msed.vse.cz/msed_2016/article/193-Csefalvaiova-Kornelia-paper.pdf

HALLAUER, J. F. (2002). Epidemiologie für Deutschland mit Prognose. In J. Hallauer and A. Kurz (Eds.), Weissbuch Demenz. Versorgungssituation relevanter Demenzerkrankungen in Deutschland, Stuttgart, pp. 15-18. Georg Thieme Verlag.

HEBÁK, P. A kol. (2015). Statistické myšlení a nástroje analýzy dat. Informatorium: 2015. Vyd. 1. 880 s. ISBN: 978-80-7333-105-4

JAGGER, C., A. M. D., M. M. B. BRETELER, J. R. M. COPELAND, C. HELMER, M. BALDERESCHI, L. FRATIGLIONI, A. LOBO, H. SOININEN, A. HOFMAN, AND L. J. LAUNER (2000). Prognosis with dementia in Europe: a collaborative study of population-based cohorts. Neurology 54, 16.

JORM, A. F. (1994). A short form of the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE): development and cross-validation. Psychological Medicine 1994; 24 (1), s. 145−153.

MALÁ, I. (2016). Properties of moment method estimates based on L moments from right censored data. In: *10th International Days of Statistics and Economics* (pp. 1159-1169). ISBN 978-80-87990-02-5. https://msed.vse.cz/msed_2016/article/245-Mala-Ivana-paper.pdf

NIKOLAI, T., VYHNÁLEK, M., ŠTĚPÁNKOVÁ, H., HORÁKOVÁ, K. (2013). Neuropsychologická diagnostika kognitivního deficitu u Alzheimerovy choroby. Praha: Psychiatrické centrum Praha 2013, 64 s. ISBN: 978-80-87142-25-7.

STEJSKAL, J., BARTOŠOVÁ, J. (2014). *Innovative approaches focused on population ageing in the Czech Republic*. In: *9th International Days of Statistics and Economics* (pp. 1450-1457). ISBN 978-80-87990-02-5. http://msed.vse.cz/msed_2014/article/395-Stejskal-Jakub-paper.pdf

ŠIMPACH, O. (2014). *Detection of Outlier Age-specific Mortality Rates by Principal Component Method in R Software: The Case of Visegrad Four Cluster. In:* 8th International Days of Statistics and Economics. Slaný: Melandrium, p. 1505–1515. ISBN 978-80-87990-02-5.

ZWEIG, M. H., CAMPBELL, G. (1993). Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clinical Chemistry. 1993. Apr. 39 (4): 561−77.

Contact

Kornélia Cséfalvaiová

University of Economics, Prague,

Department of Demography

W. Churchill sq. 4

130 67 Prague 3, Czech Republic

kornelia.csefalvaiova@vse.cz