Determinants Affecting Health of Slovak Population and their Quantification

Mária Vojtková[1] | *University of Economics in Bratislava, Bratislava, Slovakia*
Eva Kotlebová[2] | *University of Economics in Bratislava, Bratislava, Slovakia*
Daniela Sivašová[3] | *University of Economics in Bratislava, Bratislava, Slovakia*

Abstract

The state of health of the population is the result of various determinants, but also a barometer of the conditions that affect the formation of the individual's health. In a healthy society, a healthy individual can develop, and healthy individuals can develop from a healthy society. This article deals with the analysis of the impact of selected factors on the health status of the Slovak population. This is based on data from the latest EHIS (The European Health Interview Survey). We worked with the respondent's answer to the question whether he / she suffers from a long-term health problem (variable with variations yes-no). From the variables surveyed, we chose the ones we thought they could have effect on the selected indicator. With respect to the binary dependent variable, we used logistic regression for the analysis, where all calculations have been carried out in the SAS Enterprise Guide statistical program. The results are findings that have to some extent confirmed our assumptions about the impact of selected factors on health, although some of them have not been shown to the extent we expected.

Keywords

Health, health determinants, EHIS, logistic regression

JEL code

I10, C31

INTRODUCTION

Health is an important attribute of quality of life and well-being. Not only does it represent functional and instrumental value, but it also has importance for one's own identity as it determines who one is (Blaxter, 2010). In order to determine health, it is essential to define precisely when a person is healthy and when we can consider him / her sick. There is no reliable and accurate definition of health, not even that of transition from a healthy person to a sick person, because several exogenous and endogenous factors constantly cause gradual or sudden changes in human health. However, there are many definitions

[1] University of Economics in Bratislava, Faculty of Economic Informatics, Department of Statistics, Dolnozemská cesta 1, 852 35 Bratislava, Slovakia. E-mail: maria.vojtkova@euba.sk, phone: (+421)267295724.
[2] University of Economics in Bratislava, Faculty of Economic Informatics, Department of Statistics, Dolnozemská cesta 1, 852 35 Bratislava, Slovakia. E-mail: eva.kotlebova@euba.sk, phone: (+421)267295718.
[3] University of Economics in Bratislava, Faculty of Economic Informatics, Department of Statistics, Dolnozemská cesta 1, 852 35 Bratislava, Slovakia. E-mail: daniela.sivasova@euba.sk, phone: (+421)267295731.

of health. The most commonly used definition under the Constitution of the World Health Organization (WHO), which entered into force on 7 April 1948, is that "the health is a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity".[4] An important attribute is subjective feeling of the examined person, but still based on it we can only estimate the health condition.

Health and well-being are a very important part of human life, but they are influenced by many factors (Evans, Barer, Marmor, 2017; Marmot, 2005). Those related to poor health, disability, disease or deaths are known as risk factors. The risk factor is an individual's behavior or condition that increases the probability of disease or injury. They are often presented in isolation, but experience has shown that they interact. Most factors, such as lifestyle and, in part, social factors are largely up to the individual´s decision whether and to what extent they will incorporate them into their lives. The aim of this article is to identify the factors that affect the health of the population in Slovakia.

According to a survey of Citizens' Views on the Future of Slovakia (Bunčák et al., 2009), health and long life ranked second in the list of preferred life goals of the Slovak population. When it comes to future concerns, responses such as illness, deterioration of health, as well as a lack of funding for medicines and health care, came first.

In general, in the UN Human Rights Declaration, health, medical care and sickness are considered fundamental human rights. In November 2017, the Health Profile of the country was published, based on the collective work of the OECD and the European Observatory on Health Systems and Policies in cooperation with the European Commission (OECD, 2017). It is an overview of the state of health of the population and health care of the individual countries of the European Union (the EU). Based on these profiles of the 28 EU countries, it is evident that Slovakia's health has improved compared to previous years, but Slovakia still lags behind the EU average. This is evident, for example, by the average life expectancy at birth, which is one of the main synthetic indicators of population living conditions and mortality rates (indirect indicators relating to health). In 2017 it reached 77.3 years in Slovakia, which is shorter by 3.6 years compared to the EU28 average. There is a big difference between female and male sex – women live on average 7 years longer (80.7) than men (73.8). This gender gap is greater than the EU28 average (5.9 years). On the contrary, the interesting fact is that the healthy life years in Slovakia in 2017 are the same for men and women, 55.6 years, while in the EU28 there is a slight difference, 64 years for women and 63.5 years for men.

1 DETERMINANTS OF HEALTH

Differences in morbidity or mortality between countries are not only dependent on the quality of healthcare. It is true that in some countries (including Slovakia), as compared to more advanced countries, not so much money was invested to health care, either in more advanced technologies or medicines, but other important factors also affect the health status of the population.

Almost all diseases are largely initiated by risk factors, and their presence decides whether or not the disease will break out. Risk factors, in turn, are strongly influenced by the environment, which may encourage or even eliminate their occurrence. Therefore, we consider the environment as a significant determinant of health. Each risk factor has its own specifics – for some diagnoses it has a high initiation potential, for other diagnoses it can eliminate their occurrence.

We categorize health determinants into certain groups, which are:

a) lifestyle,
b) genetic basis,
c) socio-economic,
d) health care.

[4] *<https://www.who.int/about/who-we-are/constitution>.*

Among these factors, lifestyle has the highest impact on health – its impact is up to 50–60%, other factors contribute significantly lower: genetic basis 10–15%, socioeconomic and natural environment 20–25% and health care 10–15%. (Čeledová and Čevala, 2010).

The aim of this article is to identify the factors that influence the health of Slovaks, based on the European Health Survey (hereinafter referred to as "EHIS"), which was carried out in Slovakia in the second wave by the end of 2015. The number of respondents was 5 490. As we are interested in what determines our health, we decided to choose as the target variable the expression of the respondent, whether he has a certain disease or a long-term health problem for more than 6 months. A complementary goal is to quantify the impact of significant determinants on the target variable. Selected determinants (factors), whose influence we decided to investigate include:[5]

- Age,
- Gender,
- Legal marital status,
- Highest level of educational attainment,
- Respondents' employment status,
- Net monthly equivalent household income,
- General health condition perceived by the respondent,
- Hospitalization in hospital over the last 12 months,
- Last visit to a general practitioner or family doctor,
- The respondent could not afford prescribed drugs in the last 12 months,
- Body mass index (BMI),
- Physical effort in performing duties – including paid and unpaid work activities,
- Frequency of fruit consumption, excluding fruit juices made from concentrate,
- Frequency of consumption of vegetables or salads, excluding potatoes and vegetable juices made from concentrates,
- Frequency of alcoholic beverages of any kind in the last 12 months,
- Smoking habits.

Since the target variable is categorical, we decided to use the logistic regression method to achieve the designed goal. Its aim is to find the most suitable model for describing the relationship between a binary dependent variable and a set of selected explanatory variables, which can be both continuous and categorical. The analysis itself was performed using the SAS Enterprise Guide statistical tool (Dhand, 2010).

2 METHODOLOGY

To assess the statistical significance of the impact of the considered factors on probability that a person will suffer from a long-term health problem, we have decided to use the logistic regress model with logit link function (Hilbe, 2016; Hosmer and Lemeshow, 2013; Bagley et al., 2001):

$$\text{logit}(p_i) = \ln \frac{p_i}{1 - p_i} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik}, \tag{1}$$

where p_i is the probability that a person will suffer from the long-term health problem. $\beta_0, \beta_1, \ldots, \beta_k$ are parameters of logit model and $x_{i1}, x_{i2}, \ldots, x_{ik}$, where $i = 1, 2, \ldots, n$ are the values of the explanatory variables $X_1, X_2, \ldots, X_k$ observed for i-th statistical unit (in this case a person). To estimate the parameters of the logistic regression model, we used the standardly applied maximum likelihood method that maximizes the likelihood function L. To obtain maximum likelihood estimates is generally used the Newton-Raphson algorithm.

[5] A detailed description of the variables with each character category is given in the Annex.

The significance of the logistic model is verified by testing the null hypothesis, according to which holds $\boldsymbol{\beta}^{\mathrm{T}} = (\beta_1 \quad \beta_2 \quad \ldots \quad \beta_k) = \mathbf{0}^{\mathrm{T}}$, against an alternative hypothesis which is claiming that at least one regression coefficient is non-zero. We used Chi-square tests (Likelihood ratio, Score statistics, Wald statistics) in our analysis. It is well known (Allison, 2012) that for large samples all tests generally give comparable results. To verify the significance of the impact of individual explanatory variables on probability *p*, we applied the Wald test in SAS Enterprise Guide. For each of the listed factors above we tested the null hypothesis according to which the explanatory variable does not affect the probability of the investigated event occurrence. To verify the null hypotheses, we used Wald´s test statistics:

$$Wald = \hat{\boldsymbol{\beta}}^{\mathrm{T}} \cdot \mathbf{S}_{\mathbf{b}}^{-1} \cdot \hat{\boldsymbol{\beta}}\,, \tag{2}$$

where $\hat{\boldsymbol{\beta}}$ is vector of estimates of regression coefficients that stand at dummy variables for the respective factor - a categorical explanatory variable and $\mathbf{S}_{\mathbf{b}}$ is a variance-covariance matrix of a vector $\hat{\boldsymbol{\beta}}$. Wald's test statistic has an asymptotic χ^2 distribution with a number of degrees of freedom equal to the number of estimated vector parameters $\boldsymbol{\beta}$. A special case of above test is the Wald test, which verifies the statistical significance of one regression coefficient. In this case Wald statistics has an asymptotic χ^2 distribution with 1 degree of freedom and it is as follows:

$$z^2_{Wald}) = \left(\frac{\hat{\beta}_j}{s_{\hat{\beta}_j}}\right)^2, \tag{3}$$

where $s_{\hat{\beta}_j}$ is an estimated standard error of the j^{th} estimated coefficient.

In logistic regression, the effect of the explanatory variable X_j on explanatory variable Y is quantified by the odds ratio (OR), which is estimated as follows:

$$OR_j = e^{\hat{\beta}_j}\,, \tag{4}$$

where $\hat{\beta}_j$ is an estimate of the relevant regression coefficient. The odds ratio in binary logistic regression represents the change of the chance that Y = 1 (in our – case that a person will suffer from a long-term health problem) versus the chance that Y = 0 (in our case a person will not suffer from a long-term health problem), influenced by unit increase of the explanatory variable X_j under the condition of ceteris paribus. If the explanatory variable is an artificial variable, the odds ratio compares the odds at two different levels of the predictor.

The quality of the logistics model can be evaluated according to various measures. One group consists of penalty models of quality, namely AIC – Akaike Information Criterion and SC – Schwarz-Criterion, which are based on the logarithmic transformation of the likelihood function. The second group consists of the measures of association between predicted and original values of the dependent variable, including Somers D, Goodman-Kruskal gamma, Kendall tau-a and c-statistics[6] (Katamuri, 2017).

3 ASSESSMENT OF SELECTED DETERMINANTS INFLUENCING HEALTH

In this section, we focused on assessment contingency and creating a model of logistic regression, where the modeled variable is a "Long Term Health Problem", specifically whether or not the respondent suffers from any disease or health problem that persists for at least 6 months. At the same time, the dependent variable is the main subject of the study, with two variations 1 – yes, 2 – no.

[6] Note that the concordance index, c, also gives an estimate of the area under the receiver operating characteristic (ROC) curve when the response is binary.

We confirmed, that the fact, whether respondent suffered from a long-term health problem, was in 2015 significantly influenced by almost all selected determinants, by the analysis of association or contingency (Šoltés, 2008) using Chi-square tests shown in table 1. In case of significant determinants, the *p*-value is lower than the commonly used significance level 0.05. Surprisingly, only factors related to the lifestyle of the respondent, namely the frequency of fruit and vegetable consumption, proved to be insignificant determinants. Due to its nature, we have omitted the numeric variable age.

To measure the intensity of this dependence, we constructed different measures. To interpret the results, we decided to use Cramer V, which is based on the average square contingency and is a useful measure when comparing the degree of association for contingency tables of different dimensions. This degree of association has shown that the risk of a respondent's suffering from a long-term health problem lasting at least 6 months is mostly affected by the respondent's General health condition, the Ability to buy prescribed medication, and the Status of the job. A moderate significant relationship between the modeled variable and the factor can be observed with the Last doctor visit and Marital status factors.[7] The lowest degree of significant dependence can be observed between the dependent variable and Physical effort in the performance of duties, Smoking habits and Gender.

By analyzing the contingency, we assessed the relationship between the dependent variable and the analyzed determinants individually, but it should also be taken into consideration that there may also

Table 1 Assessment of contingency between analyzed determinants and risk of long-term health problem of Slovak population

Statistic	SUB STATUS			DRUGS			EMPLOYMENT			VISIT		
	DF	Value	Prob	DF	Value	Prob	DF	Value	Prob	DF	Value	Prob
Chi-Square	2	2 178.0161	<.0001	2	1 389.4363	<.0001	3	1 318.0307	<.0001	2	735.0518	<.0001
Likelihood Ratio Chi-Square	2	2 590.8150	<.0001	2	1 437.7054	<.0001	3	1 491.2656	<.0001	2	741.2896	<.0001
Cramer's V		0.6299			0.5031			0.4900			0.3659	
Statistic	**MARITAL STATUS**			**BMI**			**HOSPITAL**			**ALCOHOL**		
	DF	Value	Prob	DF	Value	Prob	DF	Value	Prob	DF	Value	Prob
Chi-Square	3	719.5360	<.0001	2	303.7197	<.0001	1	236.6876	<.0001	8	214.1047	<.0001
Likelihood Ratio Chi-Square	3	805.8298	<.0001	2	304.6617	<.0001	1	263.7490	<.0001	8	224.9365	<.0001
Cramer's V		0.3620			0.2383			0.2076			0.1976	
Statistic	**INCOME**			**EDUCATION**			**PHYSICAL EFFORT**			**SMOKING**		
	DF	Value	Prob	DF	Value	Prob	DF	Value	Prob	DF	Value	Prob
Chi-Square	5	178.7282	<.0001	5	145.3824	<.0001	3	92.3897	<.0001	3	77.1706	<.0001
Likelihood Ratio Chi-Square	5	179.1298	<.0001	5	147.2321	<.0001	3	94.9327	<.0001	3	78.6991	<.0001
Cramer's V		0.1804			0.1627			0.1297			0.1186	

[7] *<http://www.acastat.com/statbook/chisqassoc.htm>.*

Table 1 (continuation)

Statistic	SEX			FRUITS			VEGETABLES		
	DF	Value	Prob	DF	Value	Prob	DF	Value	Prob
Chi-Square	1	63.6577	<.0001	4	11.89	0.0239	4	5.0003	0.2873
Likelihood Ratio Chi-Square	1	63.6358	<.0001	4	11.45	0.0231	4	5.0063	0.2867
Cramer's V		–0.1077			0.0453			0.0302	

Source: EHIS 2015, created in SAS Enterprise Guide

be certain relationships between some factors. For example, a group of factors where contingency analysis has shown dependence on the analyzed variable (education, income, smoking, and alcohol consumption) can be determined by subjective perception of the health status of respondents in each category. Therefore, we will also assess the impact of individual factors through logistic regression, in which the impact of other relevant variables included in the model will be fixed.

We first considered the impact of all selected variables using the full regression model (see Table 2). We can see from the results of Table 2 that not all variables have a statistically significant effect on the dependent variable, so we decided to modify the model and gradually eliminate insignificant factors from the model by a stepwise regression method.

Table 2 Estimation of regression model expressing dependence of long-term health problem of person on selected factors (the full model)

Testing Global Null Hypothesis: BETA = 0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	3 151.8170	48	<.0001
Score	2 524.2487	48	<.0001
Wald	1 271.4345	48	<.0001
Effect	**DF**	**Wald Chi-Square**	**Pr > ChiSq**
Sub. Status	2	345.1817	<.0001
Drugs	2	251.9051	<.0001
Visit Doctor	2	55.9095	<.0001
Age	1	44.4643	<.0001
BMI	2	17.9255	0.0001
Employment	3	12.8506	0.0005
Physical Effort	3	6.8247	0.0777
Marital Status	3	8.5354	0.0362

Table 2 (continuation)

Effect	DF	Wald Chi-Square	Pr > ChiSq
Sex	1	4.1589	0.0414
Hospitalization	1	6.4816	0.0109
Education	5	4.6868	0.4553
Alcohol	8	9.8140	0.2783
Vegetables	4	7.6045	0.1072
Fruits	4	2.2136	0.6965
Income	5	2.8937	0.7164
Smoking	2	0.6401	0.7261

Source: EHIS 2015, created in SAS Enterprise Guide

The resulting adjusted model (see Table 3) contains ten statistically significant factors. The degree of influence of individual explanatory variables can be seen by the value of chi-square statistics. The existence of a long-term health problem for Slovak population in 2015 is mainly influenced by the subjective perception of the subject's difficulties, the possibility of affording prescribed drugs over the last 12 months, and age. To some extent, it is surprising for us to find out that the variables related to the lifestyle of the population (alcohol consumption and smoking) have been excluded from the model.

Table 3 Estimation of regression model expressing dependence of long-term health problem of person on selected factors (reduced model)

Step	Effect Entered	DF	Number In	Score Chi-Square	Pr > ChiSq
1	**Sub. Status**	2	1	1 794.5969	<.0001
2	**Drugs**	2	2	586.6429	<.0001
3	**Age**	1	3	250.8074	<.0001
4	**Doctor's Visit**	2	4	66.6645	<.0001
5	**Employment**	3	5	13.5674	0.0036
6	**BMI**	2	6	11.2212	0.0037
7	**Sex**	1	7	9.3152	0.0023
8	**Hospitalization**	1	8	5.9978	0.0143
9	**Physical Effort**	3	9	8.4544	0.0375
10	**Marital Status**	3	10	8.0824	0.0443

Source: EHIS 2015, created in SAS Enterprise Guide

To verify the significance of the model of dependence of long-term health problem of the Slovak population on selected factors, the plausibility test, score test and Wald test were used (see Table 4). For all three tests, the p-value was shown to be less than the commonly used significance level (0.05), so we can reject the hypothesis according to which all model parameters are zero. However, this result does not exclude the possibility of a zero value for any of the model parameters. In the second part of the output, there are three measures of model quality (Akaik's information criterion, Schwartz-Bayes criterion and logarithmic transformation the likelihood function), separately for the model with an intercept only and separately for the specially estimated logistic model (intercept and covariates). Since all of the above measures are lower in the logistics model, we consider it to be better than the model with an intercept only.

Table 4 Quality assessment of the reduced logistic regression model

Testing Global Null Hypothesis: BETA = 0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	3 123.7083	20	<.0001
Score	2 508.4425	20	<.0001
Wald	1 271.5884	20	<.0001

Model Fit Statistics		
Criterion	**Intercept Only**	**Intercept and Covariates**
AIC	7 233.722	4 150.014
SC	7 240.305	4 288.268
–2 Log L	7 231.722	4 108.014

Source: EHIS 2015, created in SAS Enterprise Guide

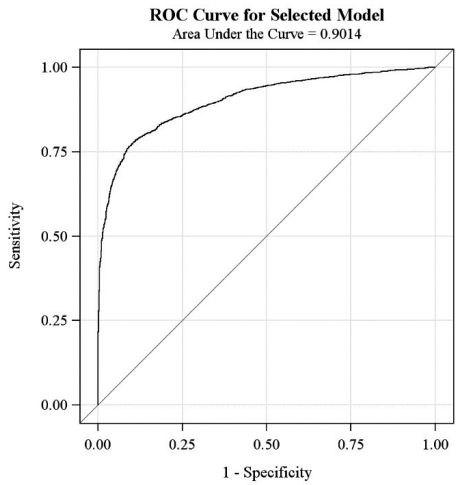
Table 5 Association between predicted probabilities obtained from the model of logistic regression of long-term health problem of Slovak population and observed values

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	90.1	**Somers' D**	0.803
Percent Discordant	9.9	**Gamma**	0.803
Percent Tied	0.0	**Tau-a**	0.388
Pairs	6 904 106	**c**	0.901

Source: EHIS 2015, created in SAS Enterprise Guide

In evaluating the model quality, we also used association measures (see Table 5) to assess the association between the predicted probabilities for the modeled variation of the dependent variable and the actual values of the dependent variable. As can be stated from column 2 of the Table 5, the proportion of matched pairs of observations is significantly higher than the proportion of opposite pairs, which indicates

Figure 1 ROC curve of logistic model of long-term health problem of Slovak population



Source: EHIS 2015, created in SAS Enterprise Guide

the quality of the model. The last column of the table contains the association measures values (Somers D, Goodman-Kruskal gamma, Kendall tau and c-statistics), which, with the exception of Kendall tau, are high, which is another argument in favor of the model accuracy.

The value of statistics c (0.901) can be represented graphically by using the ROC (see Figure 1) curve – its value is the area under the curve. As we can see, the curve is placed high above the diagonal of the square, so the quality of the model is confirmed.

In the next step, we used the unconditional maximum likelihood method to estimate model parameters. The results of the estimated parameters for each model category, point and interval estimates of the odds ratio for 2015, which we will use for the interpretation, are shown in Table 6. We will mainly focus on statistically significant variations of the variables compared to the reference

Table 6 Estimates of logistic model coefficients and odds ratios of long-term health problem of Slovak population

Analysis of Maximum Likehood Estimates		Coefficient		Odds Ratio Estimates		
Parameter	Effect	Estimate	Pr > ChiSq	Point Estimate	95% Wald Confidence Limits	
Intercept		0.1272	0.6975			
Age		0.0319	<.0001	1.032	1.023	1.042
Sex	Male	–0.1091	0.0106	0.804	0.680	0.951
	Female					
Marital Status	Single	–0.0958	0.3067	1.108	0.905	1.355
	Widower	0.3061	0.0436	1.655	1.120	2.447
	Divorced	–0.0123	0.9089	1.204	0.921	1.575
	Married					
Hospitalization	Yes	0.1784	0.0141	1.429	1.075	1.899
	No					
Visit Doctor	More than 12 months ago	0.0371	0.7466	0.542	0.453	0.648
	Less than 12 months ago	–0.6872	0.0009	0.263	0.143	0.484
	Never					

Table 6 (continuation)

Analysis of Maximum Likehood Estimates		Coefficient		Odds Ratio Estimates		
Parameter	**Effect**	**Estimate**	**Pr > ChiSq**	**Point Estimate**	**95% Wald Confidence Limits**	
Physical Effort	**Stand or sit mainly**	0.2084	0.0076	1.113	0.941	1.318
	Manual labour	0.1238	0.2681	1.023	0.779	1.344
	No work done	–0.4334	0.0086	0.586	0.382	0.900
	Moderate activity/ Walking					
BMI	**Underweight**	–0.1487	0.4149	0.953	0.557	1.633
	Overweight / obese	0.2497	0.0152	1.420	1.200	1.680
	Normal weight					
Drugs	**Yes**	0.7456	0.0001	6.177	3.426	11.135
	No	0.3296	0.0018	4.075	3.419	4.855
	Not needed					
Sub Status	**Neither good nor bad**	–0.1655	0.3628	6.876	5.524	8.558
	Bad / very bad	252 051	<.0001	77.674	28.486	211.797
	Good / very good					
Employment	**Unemployed**	–0.0407	0.6880	1.196	0.917	1.560
	Other	0.0379	0.7258	1.294	1.003	1.669
	Retired	0.2225	0.0846	1.556	1.141	2.123
	Employed					

Source: EHIS 2015, created in SAS Enterprise Guide

categories, where the *p*-value is less than the significance level of 0.1. The reference variations of each category are listed for each variation in the last empty line. All parameter interpretations are given under the ceteris paribus condition, and this will not be repeated for each individual interpretation given the scope of the article.

As we have already stated on the basis of the values in Table 1, the greatest impact on the long-term health problem suffered by the Slovak population is the variable General health state perceived by the respondent (Sub status). Overall, we can say, that this variable is statistically significant for one variation, with the category of good or very good general health being chosen as the reference category, given its most frequent occurrence. The probability of a long-term health problem of a person who perceives his general health condition as bad or very bad is up to 77.674 times higher than that in group of a persons with a good or very good general health condition. A generally perceived state of health, neither good nor bad, appears to be a statistically insignificant category.

Another statistically significant variable by both criteria is Respondent's ability to afford prescribed medication (drugs) over the last 12 months. Persons who cannot afford them are 6.177 times more likely to suffer from a long-term health problem compared to those who do not need medicines. Somewhat lower the probability of a long-term health problem was also observed for persons who, on the other hand, can afford medicines, 4.075 times higher than in those who do not need medicines. The two odds ratios presented are in line with our expectations: The absence of a prescribed medication is a strong indication of good health, and it is logical that in the case of prescribing drugs, those who can afford it, are in a better contition comparing with those, who cannot afford it.

The model results also confirmed the well-known fact that increasing age has a negative impact on health. If a person's age increases by a year, the probability of a risk of suffering from a long-term health problem is 1.032 times higher.

Our expectations were also confirmed by the variable Status of employment. In comparison with the reference category, there was a statistically significant difference (at the significance level of 0.1) only for the pensioner category – compared to the employed person, the chance of a long-term health problem is 1.556 times higher.

Analysis of variable a Doctor visit showed that those who visited a doctor less than 12 months ago had a 3.8-fold (1 / 0.263) lower statistically significant probability of risk of a long-term health problem than those who had never visited a doctor. This finding shows the importance of a doctor's visit also in terms of disease prevention.

The importance of a healthy lifestyle is confirmed by the influence of the body mass index. A person with a high weight or obesity is likely to suffer from a long-term health problem by up to 42% higher than a person with a normal weight.

For us, an interesting fact has been shown in the comparison of sexes. Although women's life expectancy is higher than that of men, this does not necessarily mean that they are generally healthier – we have found that the men´s risk of a long-term health problem is 1.244 times lower than women´s.

The group of people who have been hospitalized in the hospital for the last 12 months also proved to be a risk category. The probability of risk of suffering from a long-term health problem is 1.429 times higher than that of those who have not been hospitalized.

A person who does not perform any work tasks has a 1.7-fold lower risk of a long-term health problem than a person who usually walks or performs tasks with moderate physical exertion. This leads to the idea that physical exertion, both during and outside work, has a negative impact on health. On the other hand, it should be noted that the group of people who do not perform work tasks is formed predominantly by students or retired people living a healthy lifestyle. In assessing the impact of physical exertion in the performance of duties, there was also a statistically significant difference between those who are at work and those who are moderately physically stressed – such people are 1.1 times more likely to have a long-term health problem.

We chose a married person as the reference category of the marital status variable due to the ever-increasing importance of the harmonious family in Slovakia. The statistically significant difference was only in one category: widowed persons are likely to suffer from a long-term health problem 1.655 times higher. However, this situation needs to be seen in a broader context: the worst position of widowed persons is probably also related to the fact that they are often elderly.

DISCUSSION AND CONCLUSION

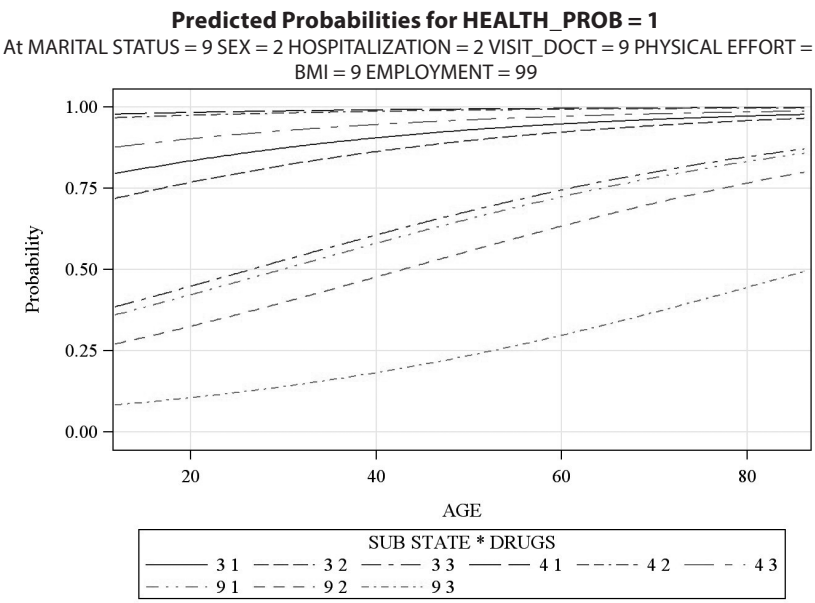
A new health policy for Europe from the WHO Regional Office for Europe Health 2020 underlines that its main objective is "to significantly improve the health and prosperity of the population, reduce the extent of health inequalities, strengthen public health and secure universal, fair, sustainable and high-quality health systems".

Concerning long-term health problems or diseases, in 2017 about one third (36.9%) of the EU28 population reported having suffered from these problems.[8] Up to 30.5% of people in the EU28 with the highest income (above fourth quintile) reported having a long-term illness or health problem, the equivalent share for people with a lower income threshold (first quintile) was up to 44.0%. While some researches suggest that health problems are more common for people with lower incomes, according to our results in Slovakia, income does not play such an important role. The results of the analysis showed that the most significant factor is the subjective perception of the subject's difficulties. In 2015, up to 60.7% of respondents perceived their health as very good or good, while only 14.2% of respondents perceived their health as bad or very bad.

In Figure 2 shows the simultaneous action of the three most important factors (general health status perceived by the respondent, the possibility to afford prescribed drugs and the person's age), while the other factors have been fixed at the reference levels. Under these conditions, it has been shown that with increasing age the probability of a person suffering from a long-term health problem increases. The results also confirmed the general fact that as the population is aging older, the risk of disease increases. The riskiest category consists of people who perceive their health as very bad or bad (SUB_STATE 4) and at the same time had (DRUGS 1) or had not (DRUGS 2) the ability to afford prescribed drugs over the last 12 months, combinations of variations appear to be not very significant. On the other hand, the least risk of a person suffering from a long-term health problem is among young people who perceive their health as very good or good (SUB_STATE 9) without needing health care (DRUGS 3).

Figure 3 highlights the importance of prevention in healthcare. The probability of a person suffering from a long-term health problem is the lowest if the person has been visited a doctor less than 12 months

Figure 2 Estimates of the risk of a person suffering from a long-term health problem depending on age, general perception and the possibility of affording prescribed drugs



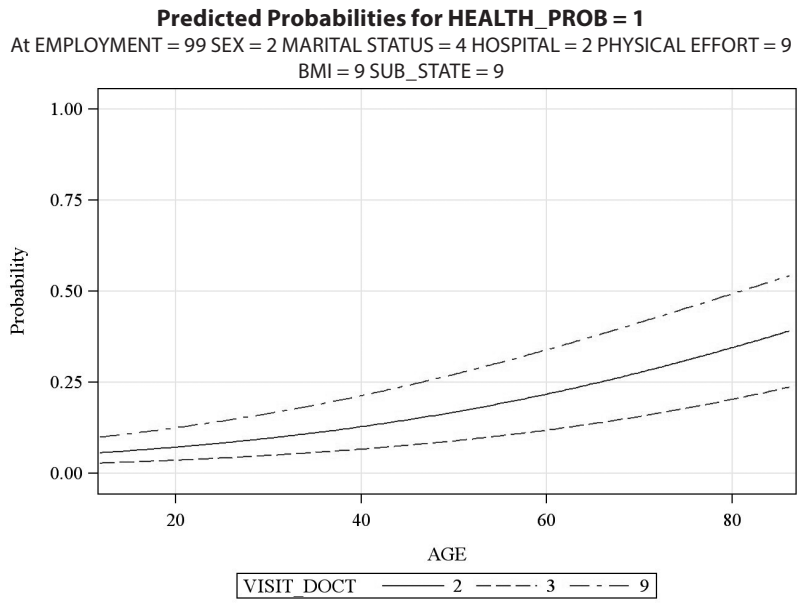
Note: For coloured figure see the online version of Statistika journal No. 4/2019.
Source: EHIS 2015, created in SAS Enterprise Guide

[8] *<https://ec.europa.eu/eurostat/statistics-explained/index.php/Quality_of_life_indicators_-_health>.*

ago. On the contrary, it is highest if the person has not visited the doctor at all, and if the person is over 80, the probability value is 50% or more.

The riskiest category in terms of all the factors analyzed can be considered an elderly man, a widower or pensioner who has perceived his condition as bad or very bad and has been hospitalized for the last 12 months, never visited a doctor, usually sits or stands, while he couldn't afford prescribed drugs for the last 12 months.

Figure 3 Estimates of the risk that a person will suffer from a long-term health problem depending on the doctor's visit



Note: For coloured figure see the online version of Statistika journal No. 4/2019.
Source: EHIS 2015, created in SAS Enterprise Guide

Analysis of association confirmed by using of Chi-square tests showed that the long-term health problem from which the Slovak population in 2015 suffered was significantly influenced by almost all selected categorical variables. The p-value of the tests is in all cases lower than the commonly used significance level.

Lifestyle (Beblavá, 2003) is a frequently discussed determinant that affects the health of the population, but according to our findings, variables such as fruit and vegetable consumption in the contingency analysis and also in the logistic regression model were proved to be insignificant. Moreover, in fixing the impact of other relevant variables included in our model, they have shown to be insignificant to alcohol consumption or smoking. Only the influence of the BMI factor and the Physical Exertion in the fulfillment of duties were significant. These findings are surprising to us; to some extent, they can be explained by the fact that the issues of consumption of vegetables and fruits about smoking were present, while the respondent's health problem lasts for at least half a year and it is not clear what the problem is and what causes it.

The strategic role of Slovak health care is to strengthen citizens' interest and responsibility for their own health, which can be achieved by informing them about the determinants affecting them. This paper provides, through the results of the present analysis, a list of potential factors that may affect health, while

quantifying their impact on the expression of the Slovak population, whether it suffers from a long-term health problem.

Especially nowadays it is important to realize that other factors affecting the health of the population (which go beyond the scope of the present analysis) are environmental. Understanding and assessing the impact of environmental factors on human health (both physical and mental) is a multidisciplinary approach. It depends mainly on the knowledge of the quality of the environment, from the internal environment (working and non-working), through the outdoor environment in urbanized units to the natural environment. Good environmental quality of man, which significantly effects his health, is a sum of good quality of air, water and food.

The World Health Organization is actively monitoring the impact of environmental factors on the occurrence of various types of diseases[9] and is actively seeking effective measures to improve the situation. However, it is necessary, especially now that global warming is objectively proven, to carry out relevant research on its impact on the health of the population in individual (and developed) countries and to take appropriate measures based on the findings.

ACKNOWLEDGMENT

The paper was supported by a grant agency of the Ministry of Education, Science, Research and Sport of the Slovak Republic VEGA. Project: VEGA No. 1/0770/17 Availability and affordability of housing in Slovakia.

References

ALLISON, P. D. *Logistic Regression using SAS. Theory and Application.* 2nd Ed. North Carolina, USA: SAS Institute, 2012.

BAGLEY, SC., WHITE, H., GOLOMB, BA. Logistic Regression in the Medical Literature: Standards for Use and Reporting, with Particular Attention to One Medical Domain. *Journal of Clinical Epidemiology*, 2001, Vol 54, pp. 979–985.

BEBLAVÁ, E. et al. *Národná správa o ľudskom rozvoji SR 2001–2002* (in Slovak). Bratislava: PR 1, Centrum pre hospodársky rozvoj, 2003.

BLAXTER, M. *Health*. Cambridge: Polity Press, 2010.

BUNČÁK, M. et al. *Názory občanov na budúcnosť Slovenska* [online]. Bratislava: Ekonomický ústav Slovenskej akadémie vied, 2009, 96 p. [cit. 25.3.2019] <https://archiv.vlada.gov.sk/old.uv/data/files/5615.pdf>.

ČELEDOVÁ, L. AND ČEVELA, R. *Výchova ke zdraví: Vybrané kapitoly* (in Czech). Prague: Grada Publishing, a. s., 2010.

DHAND, N. K. UniLogistic: A SAS Macro for Descriptive and Univariable Logistic Regression Analyses [Code Snippet 1] [online]. *Journal of Statistical Softwar*e, 2010, Vol 35. <https://www.jstatsoft.org/article/view/v035c01>.

EVANS, R. G., BARER, M. L., MARMOR, T. R. eds. *Why are some people healthy and others not? The determinants of the health of populations.* New York: Routledge. 2017.

HILBE, J. M. *Practical Guide to Logistic Regression.* Boca Raton: Chapman & Hall/CRC, 2016.

HOSMER, D. W. AND LEMESHOW, S. *Applied Logistic Regression.* New York: John Wiley & Sons, 2013.

KATAMURI, S. S. *Predictive modeling with SAS Enterprise Miner. Practical Solutions for Business Applications.* 3rd Ed. North Carolina, USA: SAS Institute Inc., 2017.

MARMOT, M. Social determinants of health inequalities. *The lancet*, 2005, Vol 365, Iss. 9464, pp. 1099–1104.

ŠOLTÉS, E. *Regresná a korelačná analýza s aplikáciami* (in Slovak). Bratislava: Iura Edition, 2008.

OECD/EUROPEAN OBSERVATORY ON HEALTH SYSTEMS AND POLICIES. *Slovensko: Zdravotný Profil Krajiny 2017* [online]. State of Health in the EU, OECD Publishing, Paris/European Observatory on Health Systems and Policies, Brussels, 2017. [cit. 20.8.2019] <https://www.oecd-ilibrary.org/social-issues-migration-health/slovensko-zdravotny-profil-krajiny-2017_9789264285408-sk;jsessionid=j9rRgLzQUXFVuXb72icZzaqX.ip-10-240-5-163>.

WHO. *Zdravie 2020: a European policy framework and strategy for the 21st century.* Copenhagen, WHO Regional Office for Europe, 2013.

[9] *<https://apps.who.int/iris/bitstream/handle/10665/204585/9789241565196_eng.pdf?sequence=1&isAllowed=y>.*

ANNEX

Table A1 Description of input variables from EHIS database

Name of the artificial variable	Original variables in EHIS	Variations		Position in EHIS
HEALTH PROB	Long-term health problem: Suffers any illness or heath problem	1	Yes	ST02
		2	No	
DRUGS	The respondent could not afford any prescription medication within past 12 month	1	Yes	CR24
		2	No	
		3	No healthcare need	
AGE	Respondent's age (number of completed years)	15–80	Persons age of 80 and over are listed as 80	HH04
MARITAL STATUS	Legal marital status	1	Single	RE03
		9*	Married	
		3	Widowed	
		4	Divorced	
EDUCATION	Highest level of education	1	Primary education	RE05
		2	Secondary education	
		3	Secondary diploma	
		4	Post-secondary education	
		5	Undergraduate education	
		6	Graduate and post graduate education	
SUB STATUS	Respondent's general heath status: How person perceives his/her own health	9*	Very good or good	ST01
		3	Neither good nor bad	
		4	Very bad or bad	
HOSPITALIZATION	Hospital stay within past 12 months	1	Yes	CR01
		2	No	
EMPLOYMENT	Respondent's employment status	99*	Employed or self-employed	RE06
		20	Unemployed	
		31	Others	
		32	Retired	

Table A1 (continuation)

Name of the artificial variable	Original variables in EHIS	Variations		Position in EHIS
INCOME	Net monthly equivalent household income	1	Under 1st quintile	HH06
		2	Between 1st quintile and 2nd quintile	
		3	Between 2nd quintile and 3rd quintile	
		4	Between 3rd quintile and 4th quintile	
		5	Above 4th	
VISIT DOCTOR	Last visit to general practice or family doctor	9*	Never	CR06
		2	Over 12 months	
		3	Less than 12 moths	
SMOKING	Habits, in terms of smoking (Are you a smoker?)	1	Yes, daily	DT15
		2	Yes, occasionally	
		3	No	
PHYSICAL EFFORT	Physical activity while accomplishing tasks	1	Mostly sitting or standing	DT03
		9*	Mostly walking or doing tasks with moderate physical activity	
		3	Mostly hard manual labour	
		4	No work done	
SEX	Sex of respondent	1	Male	HH03
		2	Female	
BMI	Body Mass Index BMI = weight (kg) / height $(m)^2$; calculated only in adults (18 years and over)	1	Underweight BMI < 18,5	
		9*	Normal weight 18,5 ≤ BMI < 25	
		3	Overweight 25 ≤ BMI < 30 obesity BMI ≥ 30	
FRUITS	Frequency of fruit consumption	1	Once or twice per day	DT11
		2	4 to 6 times per week	
		3	1 to 3 times per week	
		4	Less than once a week	
		5	Never	

Table A1 (continuation)

Name of the artificial variable	Original variables in EHIS	Variations		Position in EHIS
VEGETABLES	Frequency of vegetables or salads consumptions	1	Once or twice per day	DT13
		2	4 to 6 times per week	
		3	1 to 3 times per week	
		4	Less than once a week	
		5	Never	
ALCOHOL	Frequency of alcohol consumption	1	Every day or almost every day	DT19
		2	5 to 6 days per week	
		3	3 to 4 days per week	
		4	1 to 2 days per week	
		5	2 to 3 days per month	
		6	Once a month	
		7	Less than once a month	
		8	No alcohol within past 12 months, because I quit drinking alcohol	
		9	Never, or only few drinks throughout the life	

Note: * the highest variation number is always selected as the reference category.
Source: EHIS 2015, created in SAS Enterprise Guide