Assessment of Healthy Life Years Factors Across European countries based on Neural Networks and Multiple Linear Regression Analysis

Igor Kirshin [1,2,*]

[5] Doctor of Economics, Full professor, Higher School of Business, Institute of Management, Economics and Finance, Kazan (Volga Region) Federal University, 18 Kremlyovskaya str., Kazan, Russia, 420008; kia112524@gmail.com
[6] Doctor of Economics, Full professor, Department of General Hygiene, Kazan State Medical University, 49 Butlerova Street, Kazan, Russia, 420012; kia112524@gmail.com

Abstract
Background. The objective of this paper is to identify, test and evaluate the influence of health and disability factors on the healthy life years. Research design and methods. Panel data from the Eurostat European Health Survey and Health Statistics covering 31 European countries from 2011 to 2022 were used to examine how healthy life years are associated with health and disability factors. A cross-country multiple regression analysis with dummy variables for the COVID-19 period was performed using the multiple linear regression model and the Multilayer perceptron neural network in two versions: regression and time series (regression). Results. The results obtained convincingly confirm the proposed hypothesis: Healthy life years were significantly associated with self-assessed Disability level and self-assessed long-term limitations in usual activities due to health problems, and to a lesser extent with Proportion of people with good or very good perceived health and People with long-term diseases or health problems. Global sensitivity analysis showed that all networks consider the Level of disability factor to be the most important. Conclusion. The identified factors can be used as a significant predictors of healthy life years assessment for European countries population.

Keywords: healthy life years, time series analysis and forecasting, multiple linear regression analysis, neural networks, global sensitivity analysis, health inequalities; Self-reported health

1. Introduction

Public health is considered one of the most important systems that determine the socio-economic development of any country. It affects life expectancy, labor productivity, labor supply, human capital formation and public expenditure. Investments in health are one of the priorities of national socio-economic development strategies aimed at improving the quality of life. Various indicators are used to assess improvements in public health and the effects of public health programs.

The strategic goal of a national health system and public health financing policy is to ensure financing of health care providers and ensure equal access of the population to public health. The World Health Organization emphasizes the importance of having a health system focused on universal health coverage to increase life expectancy in the context of sustainable economic development (Dye et al., 2013; Ranabhat et al., 2018).

Historically, health has been measured primarily in terms of mortality - infant mortality, life expectancy, age-specific and disease-specific mortality rates - and morbidity - days of disability, and prevalence of chronic diseases (Erickson et al., 1995a). It was necessary to develop a universal indicator that would include both information on the health of the population and life expectancy, which would provide a more complete picture of the health of the population.

As the proportion of the ageing population, particularly in European countries, is steadily increasing, one of the most important questions is whether people live additional years of life in good or poor health as a result of increasing life expectancy. Although life expectancy in the European Union (EU) is increasing, whether most of these extra years are spent in good health is unclear (Jagger et al., 2008). Since traditional health measures do not provide an answer to this question, researchers began to develop a short set of tools to simultaneously monitor mortality and various aspects of health. The Euro-REVES 2 project, 'Setting up of a coherent set of health expectancies for the European Union', was begun in 1998 under the European Health Monitoring Programme (Robine and Jagger, 2003).

A special indicator for the integrated assessment of the performance of national healthcare systems has been developed: Healthy Life Years at birth (HLY), also called disability-free life expectancy (DFLE) or Sullivan's Index (Sullivan, 1971). The HLY concept has emerged as one of the more commonly used health status measures that include both mortality and morbidity (Erickson et al., 1995b).

The indicator “Healthy life years at birth” has been specifically developed by the European Commission as a Structural Indicator within the Lisbon Strategy (2000-2010) to monitor whether increase in life expectancy is accompanied or not by a corresponding increase in healthy active life, and whether health inequalities between Member States are reducing or not (Health expectancy: Healthy Life Years (HLY), 2000). HLY is an important indicator of the relative level of health of people in the European Union (EU). The HLY at birth is a measure of the number of years a person can expect to live in good health from birth.

In the context of population ageing in European countries, HLY can influence many other socio-economic dimensions, such as the number of economically active people, consumption, investment, pension expenditure, public finances and the sustainability of economic development. Indeed, increasing HLY can lead to an increase in labor and capital factors, since higher healthy life expectancy expands investment in all types of capital, including human capital. Many studies examining sustainable economic development have found a positive correlation between HLY and the concepts of sustainable development and integrated development.

HLY is calculated as follows (Methodology for the calculation of health expectancies, 2008):

$$HLY = \frac{\sum_{i=0}^{\omega} \left[L_i (1 - prev_i) \right]}{l_i} \tag{1}$$

where Li – number of person years lived in the age group i, $prev_i$ – the fraction of disabled persons in age i, li – number of survivors of age i.

HLY combines information on mortality and morbidity, determined by age-specific prevalence (proportions) of the population in healthy and unhealthy states and mortality in different age groups. A healthy state is defined as the absence of limitations in life activities (absence of disability).

The highest average HLY value in the EU - 64.6 years - was recorded in 2019, the year preceding the COVID-19 Pandemic. Therefore, it is logical to compare HLY before and at the end of the COVID-19 Pandemic, in 2022 (Fig. 1) (Eurostat: Healthy life years at birth by sex, 2023). Paradoxically, the largest drop in HLY is observed in the countries that were leaders in this indicator before the pandemic. Thus, the maximum decrease was observed in Spain - 8.7 years, followed by Sweden - 6.8 years; and Germany rounds out the top three anti-leaders with 5.2 years. The French health barometer, using self-reported health (SRH) and the global activity limitation indicator (GALI), showed that SRH and GALI of French adults also worsened in 2017, 2019, 2020 and 2021. The prevalence of good or very good SRH and GALI and their evolution between 2017 and 2021 were studied according to sex, age, main socioeconomic positions (SEP), and regions (Lahbib et al., 2024).

On the contrary, Slovenia led in the growth of this indicator in 2022 compared to 2019, the growth of which was 5.8 years. In the EU, the number of healthy life years at birth in 2022 was 62.6 years overall, with 62.8 years for women and 62.4 years for men, a gender gap of 0.4 years. Life expectancy in 2022 was 83.3 years for women and 77.9 years for men, a gap of 5.4 years. The calculation showed that 24.6% of women's life expectancy will be spent with activity limitations. This share in the shorter life of men was smaller and amounted to 19.9%.

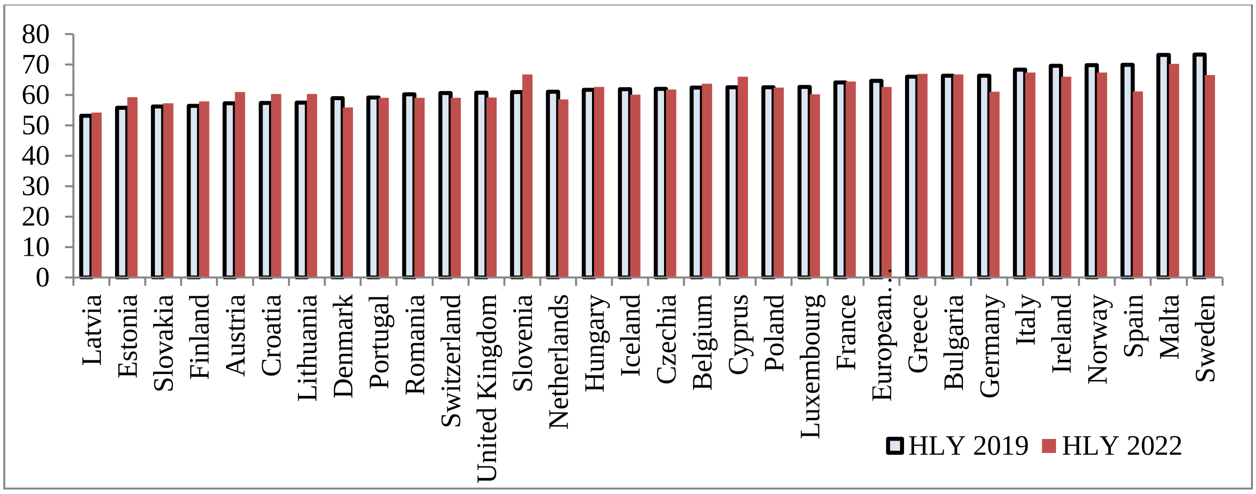


Fig. 1 - Healthy life years in absolute value at birth (Total) (Eurostat: Healthy life years at birth by sex, 2024)

In the EU countries, life expectancy at birth for women in 2022 ranged from 77.9 years in Bulgaria to 85.9 years in Spain; a difference of 8 years. A similar comparison for men showed that the lowest life expectancy in 2022 was recorded in Latvia - 69.4 years, and the highest - in Sweden - 81.4 years; a difference of 12 years.

For men, the difference between years of healthy life in Latvia (the country with the lowest HLY value for men - 51.7 years) and Malta (the country with the highest HLY value for men - 71.6 years) reached almost 20 years. For women, the difference between Lithuania (54.2 years) and Malta (72.7 years) was 18.5 years. Such a large gap shows that European countries are highly differentiated in terms of healthy life expectancy and its quality. Although periods of HLY decreases and sudden accelerations have occurred in almost all the EU countries, HLY seems to have a long-run positive trend.

The interests of all macroeconomic actors: individuals, businesses and governments coincide in the issue of increasing the duration of healthy life years. But how much can this duration be increased? How can individuals, businesses and governments manage the duration of this period? In other words: what public policy is capable of achieving this target improvement? Many researchers believe that the duration depends on two main determinants: internal - a person's lifestyle and existing external support systems in relation to the individual to which he has access (Aisa et al., 2014; Jaba et al., 2014; Shaw et al., 2005; Nixon and Ulmann, 2006). The efficiency of the country's health care system and health care expenditures have a significant impact on both determinants. While researchers' opinions differ on the inverse impact. Thus, G. Trzpiot and A. Orwat-Acedańska (Trzpiot and Orwat-Acedańska2016), studying the influence of health determinants and socioeconomic factors on HLY of men and women in 30 European countries, found no significant relationship between HLY and GDP.

The strategy of the development of national health systems includes various aspects of health care organization and public health status assessment. This strategy is implemented through the evaluation and monitoring of indicators of the functioning and effectiveness of national health systems, including the quality of and access to health services.

These indicators include:

• health expenditure;
• health resources: employment and education in the health sector, physical and technical resources;
• hospital activities: hospital care and average length of stay, diagnostic examinations and surgical procedures
• outpatient care: consultations with doctors, dentists and other health professionals;
• preventive services: in particular cancer screenings and influenza vaccination;
• drug use: prescription and non-prescription drugs;
• home care and assistance: home care services;
• unmet health needs.

2. Data and model specification

The Eurostat health dataset monitoring are used to facilitate the implementation of main directions of the health-related policies, both at European and national level (Fig. 2).

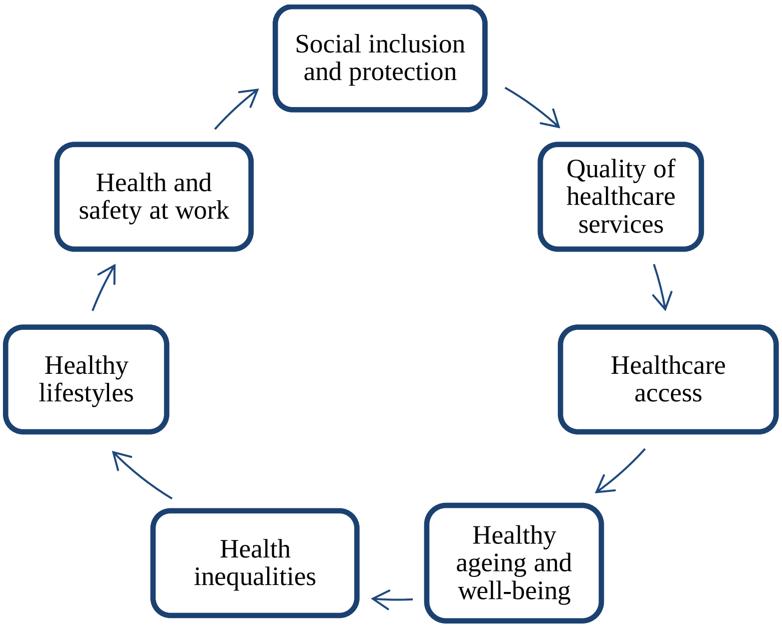


Fig. 2 - The main directions of the European health-related policies

The most of the data on health status and health determinants are determined by the EU statistics on income and living conditions (EU-SILC) and European health interview survey (EHIS) (Fig. 3) (Health methodology, 2020).

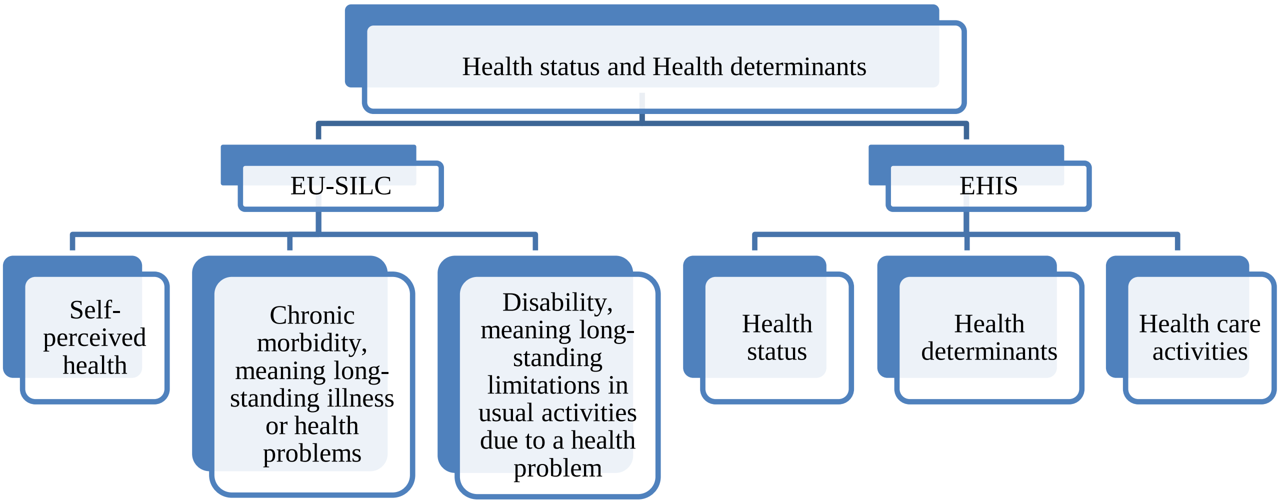


Fig. 3 - Structure of data on health status and health determinants

The first point of European nation health assessment is Health status, which focuses on various aspects of health self-perception, which enables the analysis of public health issues, demographic patterns, socio-economic trends, and disparities in health status (Table 1). A methodology for self-assessment of health was proposed within the framework of the Euro-REVES 2 project. Four metrics were proposed, reflecting self-assessment of health, physical and sensory functional limitations, activity limitations and mental health (Robine and Jagger, 2003).

Self-assessment metrics can be indicators of potential demand for health services and long-term care needs. The concept of self-assessed general health is inherently very subjective. It is limited to assessment by the individual himself and not by someone else, be it an interviewer, a health worker or a relative. Self-assessed health indicators based on this concept can be used to express general health, health inequalities and health care needs. The concept of long-term illnesses and long-term health problems is also subjective. The concept is limited to assessment by the individual himself (European Health Interview Survey wave, 2024a).

Table 1 - The Eurostat Health status datasets and indicators (Eurostat. HEALTH. Database, 2024)

№	Datasets	Selected indicators	Factors (model predictors, used in this research (2011-2022)
1.	Healthy life years at birth (HLY)	• HLY at birth by sex (total)	HLY at birth by sex (total)
		• Healthy life expectancy based on self-perceived health	
2.	Self-perceived health and well-being	• Current depressive symptoms by level of disability (activity limitation)	
		• Severity of current depressive symptoms by level of disability (activity limitation)	
		• Severity of bodily pain	
		• Self-perceived health	Share of people with very good or good perceived health by sex (total) (SGOOD)
3.	Functional and activity limitations	• Severe difficulties in personal care activities or household activities	
		• Difficulties in personal care activities or household activities	
		• Difficulties in household activities	
		• Functional limitations	
		• Self-perceived long-standing limitations in usual activities	Self-perceived long-standing limitations in usual activities due to health problem (total) (SPLSLim)
		• Level of disability (activity limitation) (Share of persons with a disability (activity limitation), %) (Population with disability, 2024)	Level of disability (activity limitation) (total) (LoD)
4.	Self-reported	• People having a long-standing illness or health	People having a long-standing

	chronic morbidity	problem	illness or health problem (total) (PLSILL)
		• Persons reporting a chronic disease	
5.	Injuries from accidents	• Persons reporting an accident resulting in injury	There is no complete data set for 2011-2022
		• Medical intervention for an accident resulting in injury	
6.	Absence from work due to health problems	Absence from work due to personal health problems by sex, age and educational attainment level	There is no complete data set for 2011-2022

Most of the data on health status come from the EU Statistics on Income and Living Conditions (EU-SILC), which presents in the following modules:

- self-perceived health and well-being
- chronic morbidity, i.e. long-term illness or health problem (people with a long-term illness or health problem);
- disability, i.e. long-term limitation in daily activities due to a health problem (Health status and health determinants, 2020).

Minimum European Health Module of Health Status includes (European Health Interview Survey wave, 2024b):

- HS1: Self-perceived general health,
- HS2: Long-standing health problem,
- HS3: Limitation in activities because of health problems.

To build a multiple linear regression model that identifies the causal relationships between HLY and determinant variables, Eurostat panel data from 2011 to 2022 for 31 countries were analyzed. The factors (model predictors) used in this study were selected from the selected tabular indicators according to the criterion of having a complete data set for 2011–2022. Thus, the model includes one dependent variable (HLY) and 4 independent variables as potential predictors of the regression model: SGOOD, SPLSLim, LoD, PLSILL.

To improve the reliability of the model, statistical procedures were used that take into account significant heterogeneity in the pre- and during-COVID-19 subsamples, since the impact of the studied HLY determinants changes more significantly during the COVID-19 years. To account for the specificity of the COVID-19 years, dummy variables were introduced into the multivariate linear regression model.

Based on the above considerations, we formulate the following hypothesis: SGOOD is positively correlated with HLY; SPLSLim, LoD, PLSILL are negatively correlated with HLY. To test the research hypothesis, the dependence of HLY on the above-described indicators was identified and tested.

To better reflect the characteristics of the multiple linear regression model, the mean absolute percentage error (MAPE) was used as an indicator of quality and accuracy, calculated as:

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|xi - \overline{xi}\right|}{xi}\cdot 100\% \tag{1}$$

where xi denotes the observed value of HLY at year i, $\overline{xi}$ denotes the predicted value (model outputs) of xi. If the MAPE is less than 5%, the model performance can be regarded as being excellent (Box, 1994).

3. Results

The results of the preliminary visual data analysis on scatter diagrams in 3D are presented in Fig. 4, 5, 6. This visual representation helped to clarify the nature of the relationship between the independent variable and dependent variables.

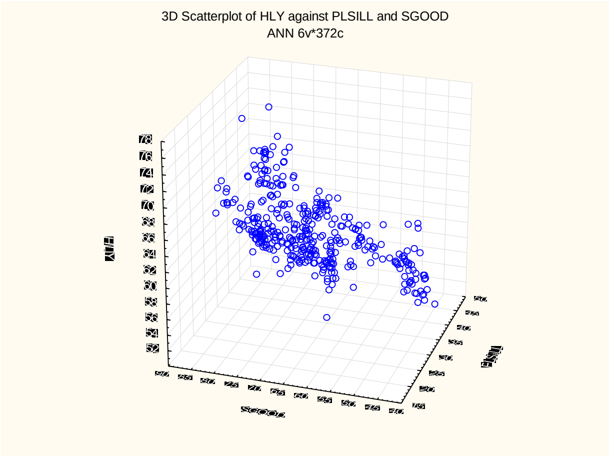


Fig. 4 - 3D Scatterplot of HLY against PLSILL and SGOOD

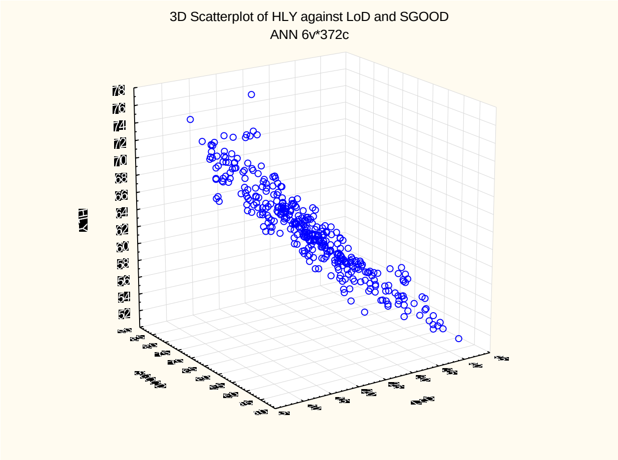


Fig. 5 - 3D Scatterplot of HLY against LoD and SGOOD

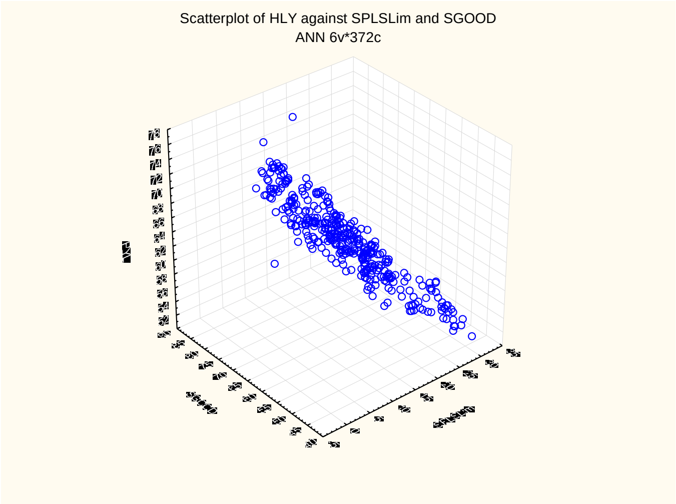


Fig. 6 - Scatterplot of HLY against SPLSLim and SGOOD

The obtained 3D images demonstrate quite strong dependencies between the pairs of predictors SGOOD, SPLSLim, LoD and the target variable. For strongly related variables, certain shapes of the set of points in 3D space are obtained. While the spatial distribution of the values of the PLSILL variable is characterized by numerous outliers, the points form a “shapeless scattering cloud”, indicating the weakness of the relationships between HLY and PLSILL. Preliminary analysis showed that the time series of HLY data is nonlinear and nonstationary. Therefore, the suitability of the ANN approach in this study was noted, since this approach can properly handle noisy nonlinear and nonstationary processes. In this paper, the ANN approach based on the multilayer perceptron neural network (MLPNN) was applied, which belongs to a general class of ANN structures called Feedforward Neural Networks (FNN). In order to obtain greater reliability of the identified cause-and-effect relationships between HLY and determinants, in our opinion, it is advisable to use two different approaches to constructing regression models provided by the Gretl and Statistica analytical software packages:

• the least squares method;

• the Multilayer Perceptron (MLP) neural network (the Neural Networks module of the STATISTICA 12 package) for two options: regression and time series (regression).

STATISTICA neural networks are a powerful environment for analyzing neural network models, which implements not only modern neural network architectures and learning algorithms, but also new approaches to selecting input data and constructing a network. The Neural Networks module of the STATISTICA system includes a procedure that organizes the search for the desired network configuration. This procedure consists of constructing and testing a large number of networks with different architectures and then selecting the network that is best suited to solving the problem. On the selected data set, 15 models of multiple linear regression (MLR) based on ANN were built (Table 2). The sizes of the training, control and test subsamples, selected randomly, are 70%, 15% and 15% of the total sample, respectively. Initially, five networks were trained on five different subsamples. Then, a visual graphical analysis was carried out to determine the networks that most adequately reflect the regression dependence.

An assessment of the histograms of the distribution of residuals for each network indicates a distribution close to normal. The scatter plot of the target (initial) and output values confirmed the fairly good quality of the selected neural network models (Fig. 7).

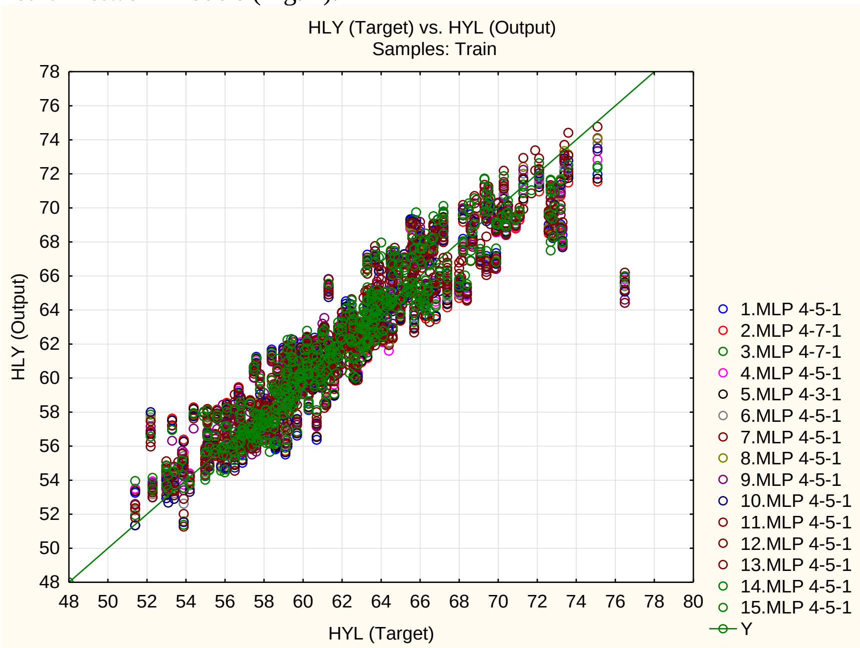


Fig. 7 - Scatterplot of HLY (Target) against HLY (Output) of 15 MLP, sample Train

The model with the highest train and test performance on the test and control subsamples was chosen as the best-fit model. It was seventh MLP 4-5-1 network. The network approximates the target data quite well. The scatter plot illustrating the dependence of the HLY values predicted by the neural network on the target values is shown in figure 8. In the ideal case (if all forecasts coincide with the real data), all points of the graph should lie on the shown line. Analysis of the histograms of the distribution of the network residuals indicates a distribution close to normal (Fig. 9).

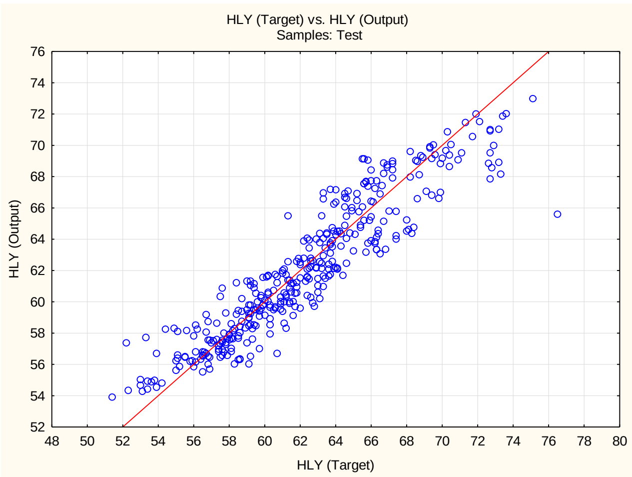


Fig. 8 - Scatterplot of HLY (Target) against HLY (Output) of MLP 4-5-1, sample Test

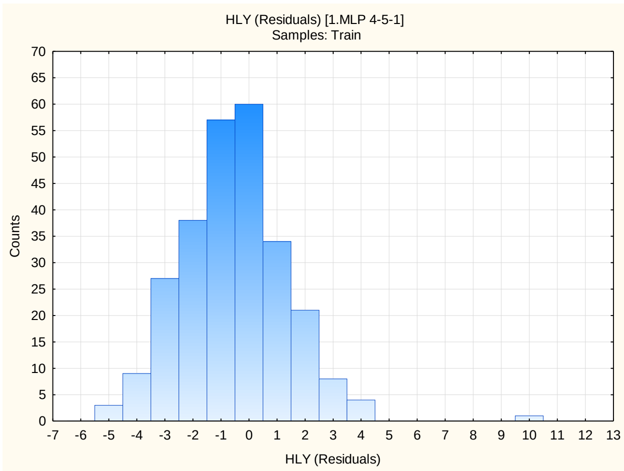


Fig. 9 - Distribution of MLP 4-5-1 residuals, sample Test

Then, response surfaces ware constructed using the input and output variables. To construct the surface, we used a point fit of a 3D scatter plot. The surface allows us to reveal the hidden structure of the data and the relationships between the three variables. Using experiments with the rotation of the response surfaces, complex nonlinear relationships between the variables are identified. The dependences of HLY on SGOOD and LoD (Fig. 10a and 10b), as well as on SGOOD (Input) and SPLSLim, are described quite well by the response surfaces (Fig. 11a and 11b). While the response surface for PLSILL has deviations (Fig. 13a and 13b).

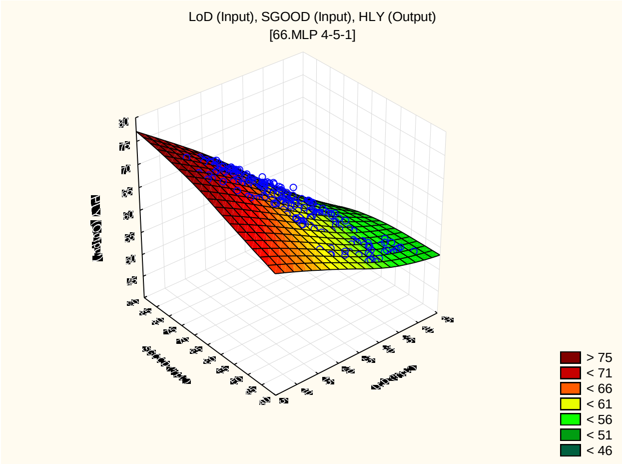


Fig. 10a - 3D Scatterplot of HLY (Output) against SGOOD (Input) and LoD (Input), MLP 4-5-1

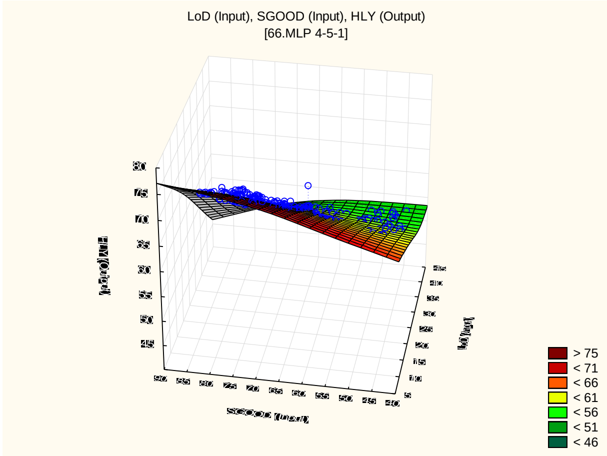


Fig. 10b - 3D Scatterplot of HLY (Output) against SGOOD (Input) and LoD (Input), MLP 4-5-1

The response surface shows how well neural networks recognize the relationship between HLY (Output) and inputs: SGOOD, SPLSLim, LoD, PLSILL . However, they cannot answer the question of whether it is possible to model a neural network more accurately than the obtained ones. In our opinion, If the results of the obtained neural network are acceptable to the researcher, then there is no need for further improvement.

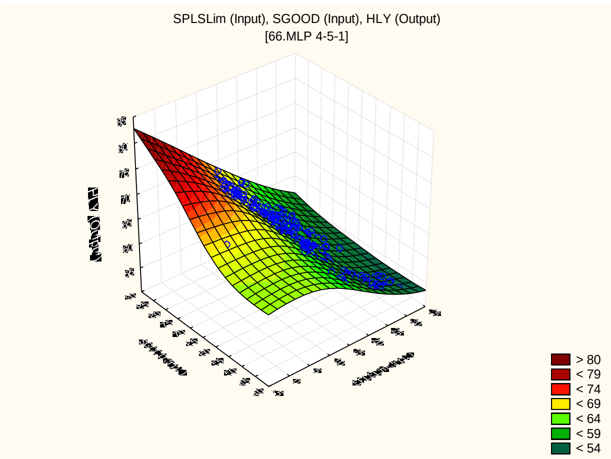


Fig. 11a - 3D Scatterplot of HLY (Output) against SGOOD (Input) and SPLSLim (Input), MLP 4-5-1

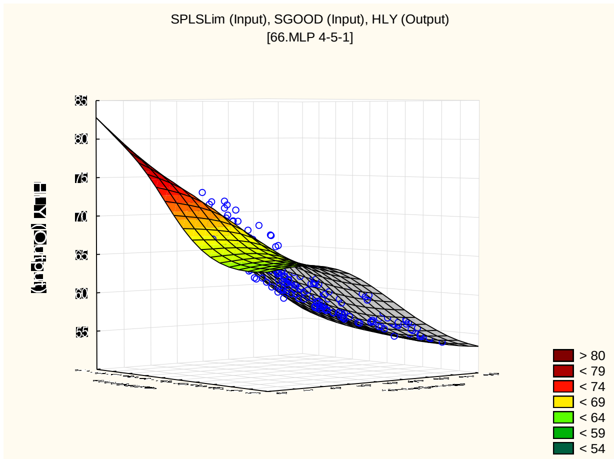
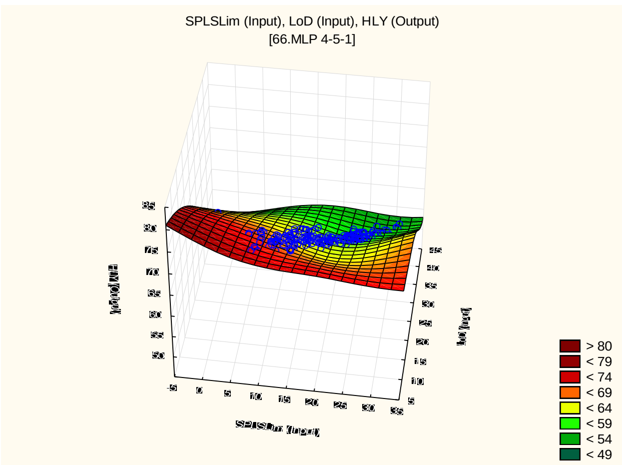


Fig. 11b - 3D Scatterplot of HLY (Output) against SGOOD (Input) and SPLSLim (Input), MLP 4-5-1

The dependence of HLY on SPLSLim and LoD is described quite well by the response surfaces (fig. 12a and 12b).



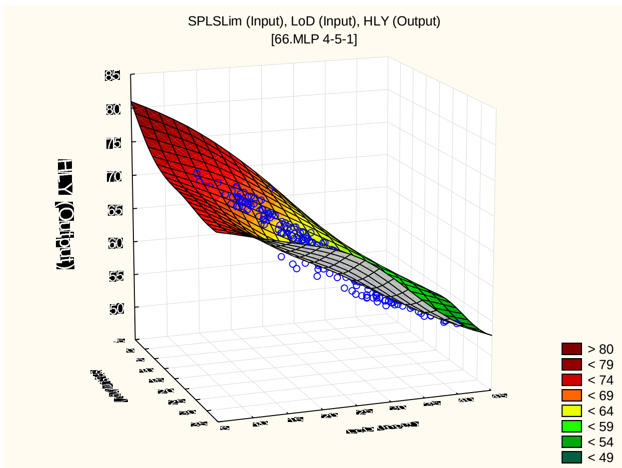


Fig. 12a - 3D Scatterplot of HLY (Output) against LoD (Input) and SPLSLim (Input), MLP 4-5-1

Fig. 12b - 3D Scatterplot of HLY (Output) against LoD (Input) and SPLSLim (Input), MLP 4-5-1

Figure 10-12 shows that the response surface of the successful model almost completely matches the scatterplot, indicating a sufficient level of dependence recognition. In contrast, the response surface of the factors: Proportion of people with good or very good perceived health and People with long-term illnesses or health problems - is characterized by a lower degree of dependence recognition (Fig. 13a, 13b).

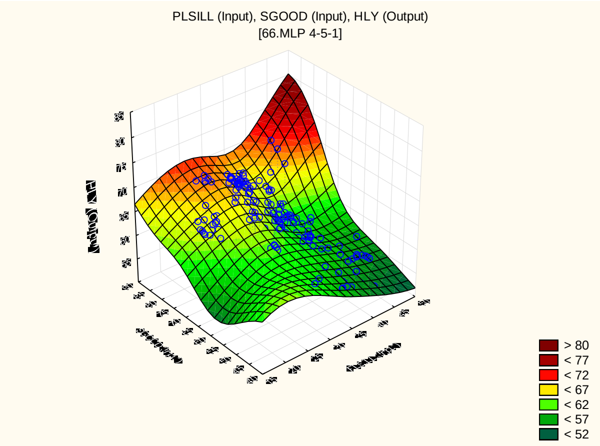


Fig. 13a - 3D Scatterplot of HLY (Output) against SGOOD (Input) and PLSILL (Input), MLP 4-5-1

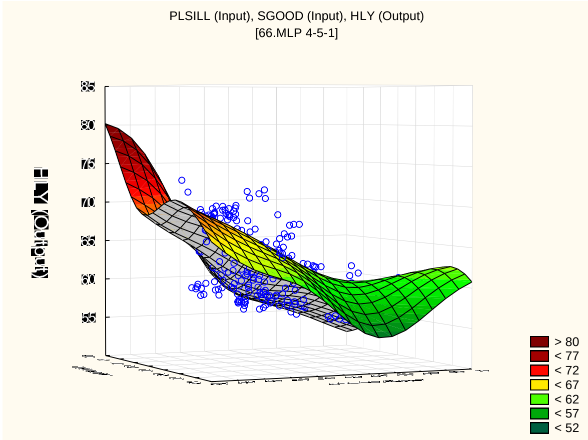
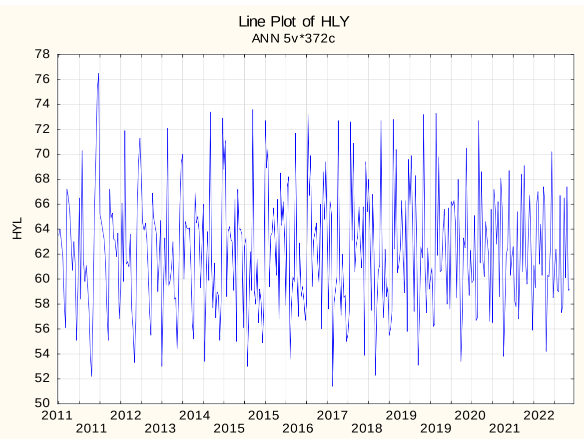


Fig. 13b. - 3D Scatterplot of HLY (Output) against SGOOD (Input) and PLSILL (Input), MLP 4-5-1

Next, we used global sensitivity analysis to determine the relative importance of the input variables for a given MLP neural network. Global sensitivity analysis is a statistical method for analyzing the effects of a relative change in the model output values due to a change in the input parameters. The result of global sensitivity analysis will be more accurate if we can build a series of models (as implemented in the STATISTICA Neural Networks package). When considering several models, global sensitivity analysis allows us to identify key variables that are always important for the response and have a high sensitivity index, as well as variables with low sensitivity and points to "doubtful" variables that change their rating and may contain redundant information.

The results of the analysis of the importance of the input variables in predicting HLY are presented in Table 3. Global sensitivity analysis showed that all networks consider the LoD variable to be the most important. This result is generally correct and logically explainable: a level of disability means any condition of the body or mind (impairment) that makes it difficult for a person with this condition to perform certain activities (activity limitation) and interact with the outside world (participation limitation). Comparative analysis of the target and output results of the best networks on the test subsample showed a fairly good quality of the constructed neural network. MAPE was only 1.73% (Table 4, 5).

Next, the analysis of the results of neural network modeling was carried out based on the time series method (regression). In Fig. 14, the original data set is presented as a time series. The best of the constructed neural networks MLP 124-20-1 was selected (Table 6). The training results of this network can be considered quite relevant to the target sample.



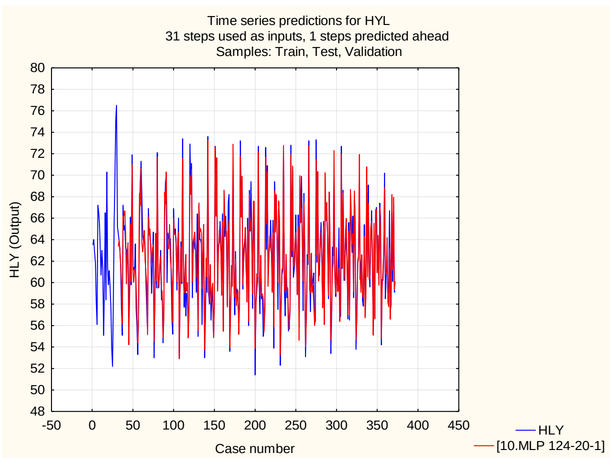
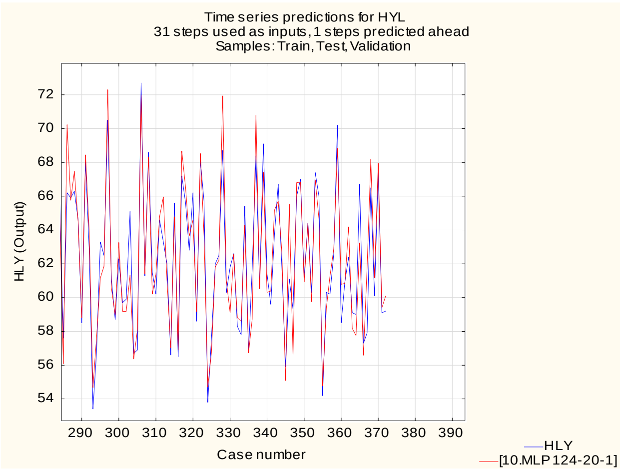


Fig. 14 - A graph of the original values of the HLY

Fig. 15 - A graph of the original (blue) and predicted (red) values of the HLY by the 124-20-1 MLP network

It is clear that for most countries the neural network managed to predict the value of the HLY indicator quite well. This graph also shows how much HLY differs in different countries. For example, HLY in 2011 in Slovakia was 52.2 years, and in Switzerland - 76.5 years. Let's compare the target and simulated values on the train and test (Fig. 15 and 16). The train and test subsamples shows a good approximation. The test subsample also shows a good approximation (Fig. 17).



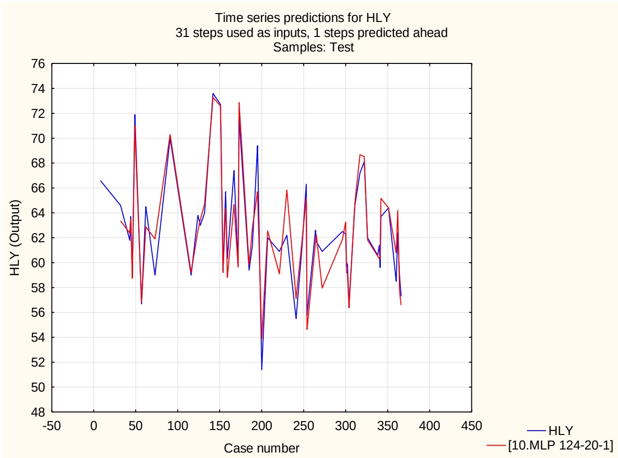
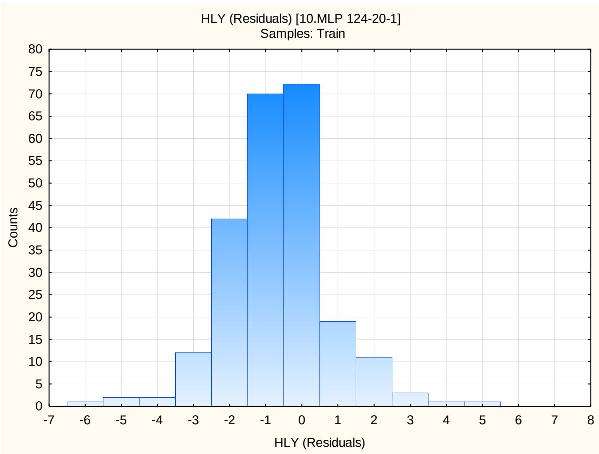


Fig. 16 - A graph of the original (blue) and predicted (red) values of the HLY by the MLP 124-20-1 network

Fig. 17 - A graph of the original (blue) and predicted (red) values of the HLY by the MLP 124-20-1 network

An analysis of the histograms of the distribution of network residuals on train and test indicates a distribution close to normal (Fig. 18, 19).



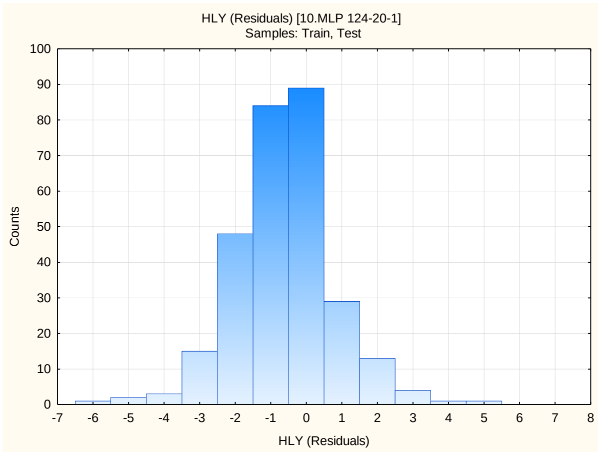


Fig. 18 - Distribution of MLP 124-20-1 residuals, sample Train

Fig. 19 - Distribution of MLP 124-20-1 residuals, sample Train, Test

A scatterplot analysis showed that the network approximates the target data quite well (Fig. 20, 21).

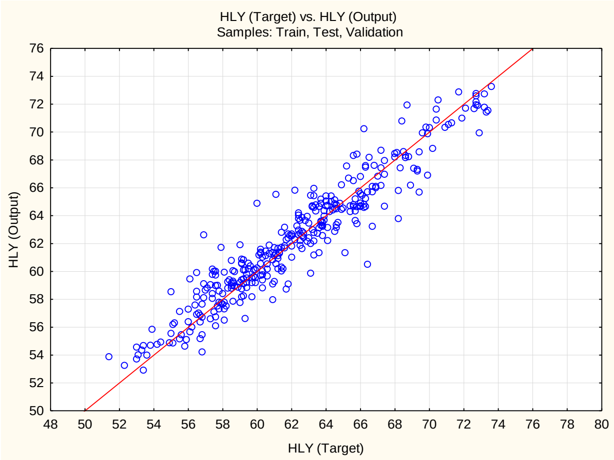


Fig. 20 - Scatterplot of HLY (Target) against HLY (Output) of MLP 124-20-1 network, sample Train, Test

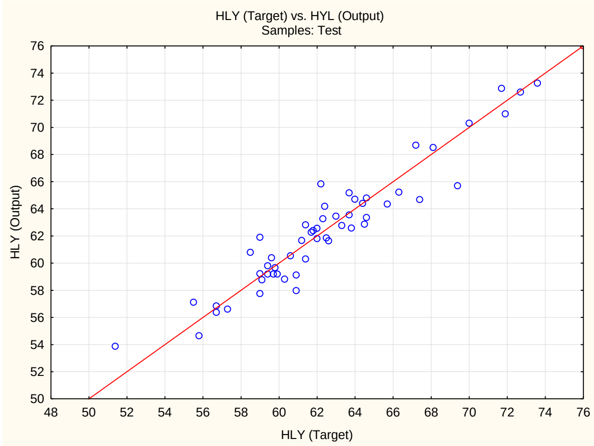
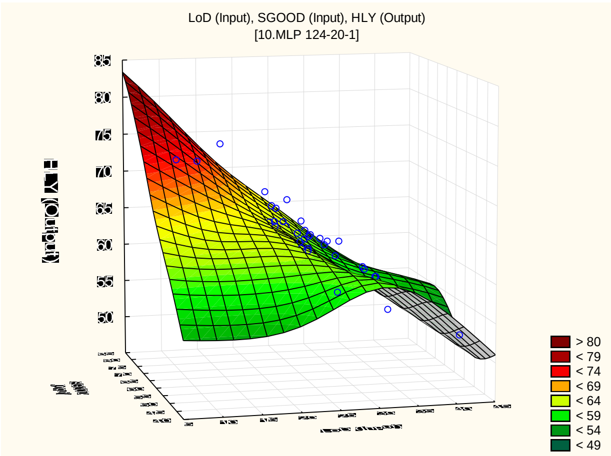


Fig. 21 - Scatterplot of HLY (Target) against HLY (Output) of MLP 124-20-1 network, sample Test

The MAPE estimate on the test subsample is presented in table 7 and 8. The MAPE on the test was 1.69%, which indicates a fairly good quality of the constructed neural network model. Next, we analyzed the response surfaces of the MLP 124-20-11 neural network.

The times series method constructed 3-D surfaces (Fig. 22 - 25) confirmed the results obtained using the MLP neural network for regression method.



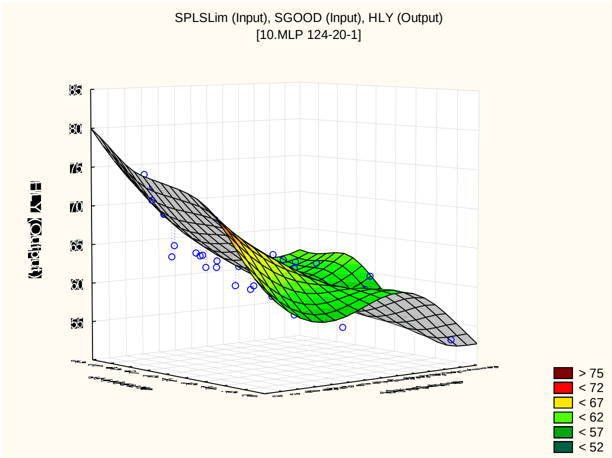
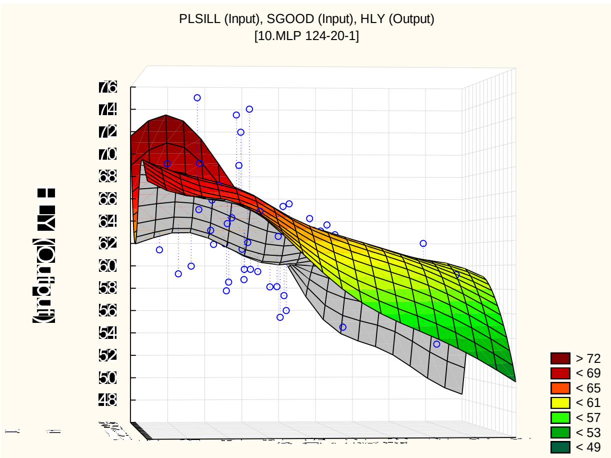


Fig. 22 - 3D Scatterplot of HLY (Output) against SGOOD (Input) and LoD (Input), MLP 124-20-11

Fig. 23 - 3D Scatterplot of HLY (Output) against SGOOD (Input) and SPLSLim (Input), MLP 124-20-11

A 3D visualization showed that the response surface of the successful model almost completely matches the scatterplot, indicating a sufficient level of dependence recognition.



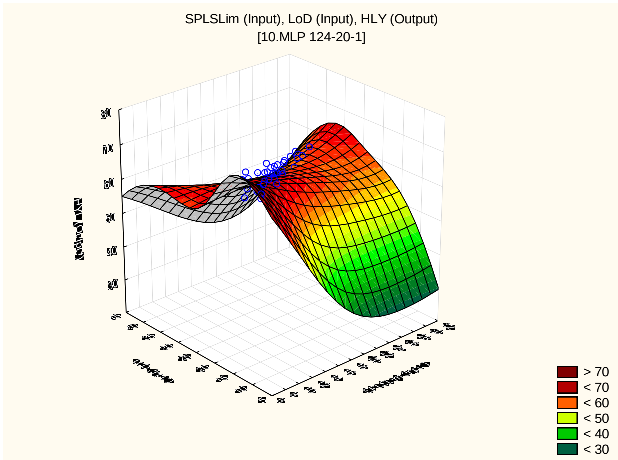


Fig.24 - 3D Scatterplot of HLY (Output) against SGOOD (Input) and PLSILL (Input), MLP 124-20-11

Fig. 25 - 3D Scatterplot of HLY (Output) against LoD (Input) and SPLSLim (Input), MLP 124-20-11

Next, the software package Gretl was also applied for revealing a regression equation, including dummy variables (Table 9). The built model of multivariate linear regression in Gretl indicates, as expected, a positive impact of SGOOD on the HLY over the analyzed period and negative impact of PLSILL, LoD and SPLSLim on the HLY over the analyzed period.

$$D_i = 1,\ for\ 2011-2019,$$

$$D_i = 0,\ for\ 2020-2022\ (COVID-19\ Pandenic\ period).$$

Table 9 – Results of multiple regression model estimation.
Dependent variable – HLY

Variables	(1)	(2)
SGOOD	0,0674 *** (0,0114)	0,06366 *** (0,0114)
PLSILL	−0,0502 ** (0,0207)	−0,0545 *** (0,0205)
LoD	−0,4649 *** (0,0373)	−0,4687 *** (0,0369)
SPLSLim	−0,1827 *** (0,0412)	−0,1826 *** (0,0408)
Const	74,7165 *** (1,2574)	75,675 *** (1,29)
D (Dummy variable)		−0,6102 *** (0,2145)
Standard error of regression	1,79	1,77
R-squared	0,86	0,86
Countries	31	31
Observatios	372	372

The fit of the model was generally high (R-squared: 0,86), which suggests that this specification is able to capture a large part of the variability in the data. Since P-value<0.01, we concluded that at the 1% significance level, all variable in model 2 is significant.

The obtained results convincingly confirmed our hypothesis that the SGOOD factor is positively correlated with HLY; while the SPLSLim, LoD, PLSILL factors are negatively correlated with HLY.

4. Conclusion, limitations and future research

Health for each person is a natural vital value, occupying the highest level in the hierarchical system of values. Therefore, the phenomenon of healthy life years is currently being actualized as a valuable resource for investing in the future. But, first of all, the person himself must realize that only he himself can and should live a high-quality, full life, without falling into the category of people with long-standing limitations in their usual life activities and disabilities. Self-assessment of a person's health in combination with behavioral, social and other factors can affect the number of years of healthy life. Given the interrelations and interdependence of these factors, a "holistic" comprehensive approach should be used that considers these variables together in one model and analyzes their impact on a sample of 31 countries for the period from 2011 to 2022. The correct choice, identification of HLY factors and assessment of the impact of the identified factors on HLY are necessary for the implementation of successful health and social policy in European countries. This study showed that the variability of HLY can be explained by the factors: SGOOD, SPLSLim, LoD, PLSILL. Based on these variables, a predictive model can be developed to predict HLY. The predictive model will allow government officials, individuals, health professionals, researchers and the general public to better estimate the expected HLY. Since the multiple regression equation requires the measurement of simple parameters, it can be time-effective, inexpensive and realistic in public health activities.

The presented study identified the determinants affecting HLY in European countries using an approach combining a regression model with autoregressive MLP. The results show that the determinants of self-rated health are statistically significant.

As a result of this study, fairly reliable neural network models were built that allow, based on Eurostat Health statistics, to determine the degree of influence of individual factors on HLY. HLY was significantly associated with the level of disability (self-assessed) and self-perceived long-term limitations in usual activities due to health problems, and to a lesser extent with the proportion of people with good or very good health perception and people with a long-term illness or health problem ($P < 0.001$ in all cases). It can be assumed that the last two indicators reported by respondents are largely dependent on the subjective perception of their level of health and long-term illness or health problem.

The social and cultural status of the respondents should also be taken into account, since people of different social, racial and ethnic backgrounds may evaluate their health differently.

Other limitations of this study include the fact that self-assessment of the same health status may vary greatly among different people. Much depends on the methods of measuring health status. People from lower social strata may not even take into account some existing limitations of their life activities, individual aspects of their health in self-assessment and assess their health as good based on their usual life activities, daily lifestyle, limited in comparison with more advanced ideas about a healthy life and developed needs of people from higher social strata of society. Researchers note that health is a dynamic balance of physical, mental, social and existential well-being in adaptation to living conditions and the environment (Krahn et al., 2021).

In our analysis, we included only the variables SGOOD, SPLSLim, LoD, PLSILL. Therefore, the neural network model MLP and the multiple regression equation developed in this study do not analyze the influence of other factors, such as lifestyle features and health care costs, on HLY. In the future, it will be necessary to develop models that take these and other factors into account.

The developed models allow us to assess the influence of the analyzed determinants on HLY, promptly diagnose emerging problems and develop solutions to eliminate them, which helps to increase the effectiveness of health-saving policies.

Appendix

Table 2 - ANN-based multiple linear regression (MLP) models

Index	Net. name	Training perf.	Test perf.	Validation perf.	Training error	Test error	Validation error	Output activation
1	MLP 4-5-1	0,925853	0,947168	0,920753	1,784823	1,004226	1,397927	Logistic
2	MLP 4-7-1	0,926716	0,947866	0,919956	1,778297	0,939558	1,399322	Logistic
3	MLP 4-7-1	0,933751	0,947227	0,917451	1,598773	0,940684	1,420862	Tanh
4	MLP 4-5-1	0,935344	0,948669	0,925602	1,570763	0,916281	1,301011	Logistic
5	MLP 4-3-1	0,927445	0,947431	0,918827	1,748050	0,965192	1,429505	Logistic
6	MLP 4-5-1	0,935063	0,948153	0,919319	1,569584	0,920180	1,385300	Identity
7	MLP 4-5-1	0,947118	0,955378	0,922507	1,284495	0,797768	1,358605	Identity
8	MLP 4-5-1	0,935132	0,946121	0,917678	1,565942	0,961259	1,426703	Identity
9	MLP 4-5-1	0,942693	0,945354	0,915305	1,389130	0,981955	1,465663	Identity
10	MLP 4-5-1	0,935263	0,948540	0,914905	1,563381	0,913434	1,445224	Identity
11	MLP 4-5-1	0,953233	0,898539	0,929216	1,035479	2,138522	1,616246	Identity
12	MLP 4-5-1	0,936825	0,947964	0,914430	1,526441	0,933175	1,496239	Identity
13	MLP 4-5-1	0,936680	0,944321	0,954057	1,327557	1,118430	1,362186	Identity
14	MLP 4-5-1	0,938136	0,906799	0,912790	1,395926	2,279501	1,697260	Identity

15	MLP 4-5-1	0,945153	0,950960	0,934240	1,180829	0,970448	1,888290	Identity

Table 3 - Results of global sensitivity analysis

Networks	Sensitivity analysis (ANN) Samples: Train			
	LoD	SPLSLim	SGOOD	PLSILL
1.MLP 4-5-1	2,749729	1,436970	1,119466	1,002683
2.MLP 4-7-1	3,065003	1,514784	1,173668	1,004418
3.MLP 4-7-1	4,400724	1,367479	1,112627	1,128507
4.MLP 4-5-1	4,729756	1,388129	1,198083	1,053940
5.MLP 4-3-1	3,125240	1,427670	1,143524	1,009356
6.MLP 4-5-1	4,520067	1,438752	1,122601	1,107629
7.MLP 4-5-1	8,011707	4,918591	1,234175	1,441946
8.MLP 4-5-1	4,810756	1,372319	1,106004	1,134237
9.MLP 4-5-1	6,645698	3,876052	1,087940	1,259330
10.MLP 4-5-1	5,056953	1,364065	1,115763	1,207627
11.MLP 4-5-1	5,618194	3,400124	1,441819	1,459231
12.MLP 4-5-1	4,779838	1,337446	1,104310	1,136728
13.MLP 4-5-1	4,611246	2,245784	1,148694	1,249025
14.MLP 4-5-1	4,007391	1,174490	1,131414	1,007686
15.MLP 4-5-1	5,427111	3,733998	1,620899	1,290157
Average	4,770628	2,133110	1,190732	1,166167

Table 4 - MLP 4-5-1 MAPE. Samples: Test

Case name	Predictions spreadsheet for HLY (ANN) Samples: Test			
	HLY Target	HLY - Output 4. MLP 4-5-1	HLY - Abs. Res. 4. MLP 4-5-1	MAPE = v3/Abs(v1)
2011	64,00000	67,02634	3,026343	0,047286611
2011	67,20000	69,02108	1,821085	0,0270994787
2011	55,10000	57,08914	1,989137	0,0361004897
2011	65,20000	65,98059	0,780591	0,0119722598
2012	65,30000	66,33187	1,031874	0,015802045
2012	56,80000	58,38978	1,589779	0,027989068
2012	56,70000	56,36264	0,337358	0,00594987112
2013	63,90000	64,89926	0,999261	0,0156378915
2013	63,30000	62,43841	0,861588	0,0136111799
2013	61,00000	62,22275	1,222749	0,020045062
2013	69,30000	70,29431	0,994313	0,0143479445
2014	60,90000	60,85650	0,043495	0,00071420524
2014	56,50000	55,58667	0,913330	0,0161651255
2014	59,30000	58,74651	0,553485	0,00933364847
2014	59,70000	60,24421	0,544212	0,00911577615
2014	58,20000	57,35352	0,846483	0,0145443886
2015	64,20000	64,88146	0,681464	0,0106146996
2015	63,20000	65,70748	2,507484	0,0396753859
2015	66,40000	64,40051	1,999488	0,0301127656
2015	59,20000	61,82604	2,626043	0,0443588375
2015	54,90000	58,04805	3,148050	0,0573415285

2015	57,80000	57,05016	0,749835	0,0129729269
2015	68,90000	69,21541	0,315412	0,00457782601
2015	70,40000	69,20627	1,193726	0,016956338
2015	63,50000	63,93360	0,433601	0,00682835607
2016	68,50000	69,25472	0,754717	0,0110177602
2016	66,20000	64,53430	1,665704	0,0251616854
2016	59,80000	60,17967	0,379671	0,0063490158
2016	71,70000	70,47170	1,228299	0,0171310863
2016	60,30000	59,97388	0,326115	0,0054082108
2016	63,10000	63,34614	0,246138	0,00390075646
2017	59,70000	59,37980	0,320201	0,00536350373
2017	63,70000	61,72203	1,977972	0,0310513645
2017	72,70000	71,19070	1,509297	0,0207606255
2017	63,10000	63,89878	0,798779	0,0126589419
2018	65,80000	67,13771	1,337705	0,0203298652
2018	65,80000	63,93591	1,864092	0,0283296669
2018	59,20000	58,77110	0,428900	0,00724493016
2019	58,90000	58,79038	0,109622	0,00186115197
2019	66,00000	65,63953	0,360472	0,00546169901
2019	57,40000	56,38005	1,019951	0,0177691816
2019	61,70000	61,16574	0,534257	0,00865895029
2019	69,80000	66,49431	3,305692	0,0473594779
2020	65,90000	65,40251	0,497486	0,00754909757
2020	62,90000	64,06141	1,161408	0,0184643507
2020	61,30000	61,81149	0,511485	0,00834397058
2020	61,50000	61,10158	0,398424	0,00647844084
2021	66,20000	63,70129	2,498711	0,0377448716
2021	58,60000	57,97827	0,621729	0,010609711
2021	57,60000	57,40016	0,199835	0,00346936066
2021	60,30000	59,65978	0,640216	0,0106171764
2021	61,70000	59,89488	1,805124	0,0292564673
2022	58,50000	58,07112	0,428876	0,00733120714
2022	60,90000	59,98773	0,912267	0,014979758
2022	59,10000	58,37641	0,723591	0,0122434947

Table 5 - MLP 4-5-1 MAPE case 1-55. Samples: Test

Case name	Predictions spreadsheet for HLY (ANN) Samples: Test			
	HLY Target	HLY - Output 4. MLP 4-5-1	HLY - Abs. Res. 4. MLP 4-5-1	MAPE
MEAN case 1-55	62,5236364	62,4877397	1,08685308	0,0173101725

Table 6 - ANN-based multiple linear regression (Time-series MLP) models

Index	Net. name	Training perf.	Test perf.	Validation perf.	Training error	Output activation
1	MLP 124-30-1	0,951114	0,928580	0,923929	0,996698	Identity
2	MLP 124-30-1	0,963943	0,924555	0,920197	0,777236	Identity

3	MLP 124-30-1	0,958363	0,898227	0,946296	0,794021	Identity
4	MLP 124-30-1	0,960534	0,907098	0,897892	0,820476	Identity
5	MLP 124-30-1	0,961110	0,948836	0,925089	0,761412	Identity
6	MLP 124-20-1	0,957452	0,935592	0,932859	0,864926	Identity
7	MLP 124-20-1	0,967780	0,926209	0,902248	0,695394	Identity
8	MLP 124-20-1	0,956178	0,897646	0,949098	0,845268	Identity
9	MLP 124-20-1	0,962755	0,908206	0,897308	0,772099	Identity
10	MLP 124-20-1	0,953597	0,951353	0,940297	0,909211	Identity
11	MLP 124-10-1	0,956459	0,935859	0,929473	0,887524	Identity
12	MLP 124-10-1	0,966298	0,927381	0,916376	0,725297	Identity
13	MLP 124-10-1	0,957822	0,900482	0,947674	0,809068	Identity
14	MLP 124-10-1	0,972400	0,907903	0,897443	0,575687	Identity
15	MLP 124-10-1	0,952259	0,947649	0,941380	0,936510	Identity

Table 7 - MLP 124-20-1 MAPE. Samples: Test

Case name	Predictions spreadsheet for HLY (ANN) Network: 10.MLP 124-20-1 Samples: Test			
	HLY Target	HLY - Output 10.MLP 124-20-1	HLY - Abs. Res. 10.MLP 124-20-1	MAPE =v3/Abs(v1)
2011	66,60000			
2012	64,60000	63,35090	1,249097	0,0193358643
2012	61,80000	62,38192	0,581919	0,00941615903
2012	63,70000	63,53762	0,162382	0,00254916187
2012	59,10000	58,75186	0,348143	0,00589074589
2012	71,90000	70,99221	0,907791	0,0126257473
2012	56,70000	56,84121	0,141209	0,00249046513
2012	64,50000	62,86150	1,638500	0,025403096
2013	59,00000	61,90245	2,902453	0,049194113
2013	70,00000	70,30135	0,301352	0,0043050356
2014	59,00000	59,21120	0,211198	0,00357962719
2014	63,80000	62,57534	1,224664	0,019195364
2015	63,00000	63,45593	0,455925	0,00723690974
2015	64,00000	64,69837	0,698374	0,0109120909
2015	73,60000	73,26598	0,334017	0,00453826815
2015	72,70000	72,59685	0,103147	0,00141880031
2015	59,40000	59,19879	0,201213	0,00338743213
2016	65,70000	64,33721	1,362794	0,0207426829
2016	63,30000	62,77005	0,529949	0,00837202789
2016	60,30000	58,81486	1,485139	0,0246291682
2016	67,40000	64,67528	2,724718	0,0404260874
2016	59,80000	59,64800	0,151996	0,00254174585
2016	71,70000	72,87784	1,177840	0,0164273401
2016	59,40000	59,80627	0,406267	0,00683951949
2017	61,40000	62,80832	1,408320	0,0229368049
2017	69,40000	65,69453	3,705475	0,0533930068
2017	51,40000	53,86928	2,469276	0,0480403837
2017	62,00000	62,55027	0,550269	0,00887529933

2018	60,90000	59,10413	1,795873	0,0294888755
2018	62,20000	65,82992	3,629922	0,0583588794
2018	55,50000	57,11187	1,611866	0,029042629
2018	61,20000	61,65512	0,455120	0,00743660164
2019	66,30000	65,22027	1,079729	0,0162855115
2019	55,80000	54,63560	1,164398	0,0208673551
2019	62,60000	61,63842	0,961576	0,0153606311
2019	61,70000	62,26695	0,566948	0,00918878449
2019	60,90000	57,96199	2,938015	0,0482432664
2020	62,50000	61,85359	0,646406	0,0103424993
2020	62,30000	63,25825	0,958253	0,0153812754
2020	59,70000	59,18563	0,514374	0,00861597921
2020	59,90000	59,18468	0,715322	0,0119419419
2020	56,70000	56,36369	0,336311	0,00593140259
2021	64,60000	64,79584	0,195841	0,00303158913
2021	67,20000	68,67572	1,475721	0,0219601337
2021	68,10000	68,52127	0,421267	0,00618600889
2021	62,00000	61,81337	0,186628	0,00301013319
2021	60,60000	60,53496	0,065039	0,00107324855
2021	61,40000	60,30799	1,092015	0,0177852532
2021	59,60000	60,38841	0,788407	0,0132283045
2022	63,70000	65,16670	1,466698	0,0230250909
2022	64,40000	64,39219	0,007809	0,000121255843
2022	58,50000	60,78765	2,287652	0,039105157
2022	62,40000	64,18028	1,780276	0,0285300664
2022	59,00000	57,75315	1,246852	0,0211330764
2022	57,30000	56,59445	0,705546	0,012313202

Table 8 - MLP 124-20-1 MAPE case 1-55. Samples: Test

Case name	Predictions spreadsheet for HLY (ANN) Network: 10.MLP 124-20-1 Samples: Test
	MAPE
MEAN case 1-55	0,0168831685

Conflicts of interest
"All authors declare that they have no conflicts of interest".

References

Aisa, R., Pueyo, F. and Clemente, J. (2014) 'The influence of (public) expenditure on longevity', *International Journal of Public Health,* Vol. 59 No 5, pp. 867–875.

Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. (1994) Time Series Analysis; Forecasting and Control. 3rd Edition, Prentice Hall, Englewood Cliff, New Jersey.

Jagger, C., Gillies, C., Mascone, F., Cambois, E., Van Oyen, H., Nusselder, W.J., Robine, J.-M., EHLEIS team. (2008) 'Inequalities in healthy life years in the 25 countries of the European Union in 2005: a cross-national meta-regression analysis', *The Lancet,* Vol. 372 Issue 9656, pp. 2124-2131.

Dye, C., Reeder, C., and Terry, R. R. (2013) 'Research for universal health coverage', *Science Translational Medicine,* Vol. 5 No 199, DOI:10.1126/scitranslmed.3006971.

Erickson, P., Wilson, R., and Shannon, I. (1995) 'Years of Healthy Life', *Healthy People 2000. Statistical notes*. No 7. [online] https:// DOI: 10.1037/e583992012-001

European Health Interview Survey wave 4 – Methodological manual – 2024 edition. (2024) Available online at: https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/w/ks-gq-24-018 (accessed 26 July 2024)

Eurostat. HEALTH. Database. (2024) Available online at: https://ec.europa.eu/eurostat/web/health/database (accessed 22 October 2024)

Eurostat: Healthy life years at birth by sex. (2024) Available online at: https://ec.europa.eu/eurostat/databrowser/view/tps00150/default/table?lang=en&category=t_hlth.t_hlth_state (accessed 12 June 2024)

Health expectancy: Healthy Life Years (HLY). (2024) Available online at: https://ec.europa.eu/health/indicators/docs/echi_40.pdf (accessed 20 December 2024)

Health methodology. (2024) Available online at: https://ec.europa.eu/eurostat/web/health/methodology#Health%20status%20and%20health%20determinants (accessed 21 November 2024)

Health status and health determinants. (2024) Available online at: https://ec.europa.eu/eurostat/web/health/methodology#Health%20status%20and%20health%20determinants (accessed 21 November 2024)

Robine, J.-M., Jagger C., Euro-REVES group. (2003) 'Creating a coherent set of indicators to monitor health across Europe: the Euro-REVES 2 project'. *European Journal of Public Health*, Vol. 13 No 3, pp. 6-14.

Jaba, E., Balan, C., and Robu, I. (2014) 'The relationship between life expectancy at birth and health expenditures estimated by a cross-country and time-series analysis', *Procedia Economics and Finance,* No 15, pp. 108–114.

Krahn, G.L., Robinson, A, Murray, A.J., Havercamp, S.M., Nisonger RRTC on Health and Function. (2021) 'It's time to reconsider how we define health: Perspective from disability and chronic condition', *Disability and Health Journal,* Vol. 14 No 4, p.101129 [online] doi: 10.1016/j.dhjo.2021.101129.

Lahbib, H., Carcaillon-Bentata, L., Beltzer, N., Delpierre, C. and Coste, J. (2024) 'Trends and social inequalities in self-reported health and activity limitations in France between 2017 and 2021: results from four nationwide representative surveys', *BMC Public Health*. Vol 24 No 1, p. 1916. [online] https://doi: 10.1186/s12889-024-19437-2.

Methodology for the calculation of health expectancies. (2008). Available online at: https://ine.es/en/daco/daco42/discapa/meto_evld_en.pdf (accessed 22 November 2024)

Nixon, J. and Ulmann, P. (2006) 'The relationship between health care expenditure and health outcomes', *The European Journal of Health Economics, Vol. 7 No* 1, pp. 7–18.

Population with disability. (2024) Available online at: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Population_with_disability#Self-reported_disability_. 28limitation_in_usual_activities_due_to_health_problems.29 (accessed 31 July 2024)

Ranabhat, C. L., Atkinson, J., Park, M. B., Kim, C. B. and Jakovljevic, M. (2018) 'The influence of universal health coverage on life expectancy at birth (LEAB) and healthy life expectancy (HALE): A multi-country cross-sectional study', *Frontiers in Pharmacology, Vol. 9,* [online] https://doi.org/10.3389/fphar.2018.00960.

Shaw, J., Horrace, W. and Vogel, R. (2005) 'The determinants of life expectancy: An analysis of the OECD health data', *Southern Economic Journal, Vol. 71 No* 4, pp. 768–783.

Sullivan, D.F. (1971) 'A single index of morbidity and mortality', *HSMHA Health Reports,* Vol. 86 No 4, p. 347–55.

Trzpiot, G. and Orwat-Acedańska, A. (2016) 'Spatial quantile regression in analysis of healthy life years in the European Union countries', *Comparative Economic Research. Central and Eastern Europe*, Vol. 19, Iss. 5, pp. 179-199, https://doi.org/10.1515/cer-2016-0044