Evidence-based analysis of the situation with tuberculosis in Kyrgyzstan and Pakistan

Svetlana Ts. Manzhikova (0000-0001-6164-203X)[1(1)], **Adil Iqbal** (0000-0002-2838-3483)[1], **Humna Mustafa**[2]

[1] ADAM University, Bishkek, Kyrgyz Republic
[2] Bahauddin Zakaria University, Multan, Punjub, Pakistan

Abstract. The target of this work is to implement an evidence-based approach in a comparative analyzing of the current tuberculosis situation in the Kyrgyz Republic and the Islamic Republic of Pakistan. Such analysis is of particular interest for assessing the so-called "country" contribution to the implementation of strategies to stop tuberculosis, proclaimed by the World Health Organization. Two-factor linear regression models are presented for both countries. Models take into account how state budget expenditures on health care and the level of poverty impact onto the tuberculosis incidence. The latter circumstance is due to the significant difference between the National Statistical Services in considered countries, as well as the lack of initial data and their heterogeneity. The estimated quality characteristics of the models show them to use in prediction values of the tuberculosis incidence until 2024-2025. The modeling results show that if the established trend of strengthening of the state role continues, then we can talk about the successful implementation of the WHO strategy to decrease and stop tuberculosis in the Kyrgyz Republic and the desired annual reduction in incidence to 10% by 2025. For Pakistan, the same parameter will be no more than 1.5%. The novelty of the presented work lies in the very formulation of the problem of monitoring the current outcomes of the implementation of WHO recommendations in different countries and the methodological development of appropriate situational tasks for learning and teaching medical students the methods of evidence-based medicine/health care.

Keywords: Evidence-based approach in medicine/health care · Specification of initial statistical data · Regression modeling · Situational problems for teaching and learning.

1. Introduction

Evidence-based medicine is one of the modern trends of today's medical science. It requires the training of a physician in the collection and processing of statistical data accompanying him/her in the process of professional activity. The students obtain initial information about the types of medical statistical data and corresponding methods of their processing as early as the first year in the computer science class when they master MS Excel spreadsheet [1]. Further, they are introduced to medical statistics as part of the educational process. However, as a rule, students are not in a hurry to implement the knowledge and skills acquired during clinical practice and special disciplines. One

[1] Corresponding author: msvetlana_88@mail.ru

reason for this is insufficient basic mathematical training of medical students. It does not allow them to identify and extract meaningful information from the original data, which is the basis for the adoption or evaluation of already adopted and/or current management decisions in the organization of treatment and/or healthcare.

Taking into account that students of the Islamic Republic of Pakistan are also studying in the Graduate School of Medicine at the Adam University, the authors decided to assess the probability of the WHO goals [2,3,4] of eliminating the tuberculosis epidemic in the Kyrgyz Republic and in the Republic of Pakistan can be realized. In other words, the authors had to conduct a comparative analysis of the current tuberculosis (TB) situation using open relevant statistical data from both countries. This work is a logical continuation and development of methodological approaches to the implementation of systems analysis methods in medicine/healthcare in the educational process by setting and solving situational tasks of real content presented earlier in the article [5].

The WHO "Stop TB" strategy calls for action by national governments, societies and individual organizations to "first, accelerate the annual decline in global tuberculosis rates from 2% in 2015 to 10% per year by 2025." [4].

This problem statement is of interest to students, both Kyrgyz and foreign, and may motivate them to get an answer by performing a very rigorous, logical, but multiway and "branching" algorithm for statistical research in the context of situational tasks.

2. Materials and Method

2.1. Selection of baseline data for assessing the current TB situation in the KR and Pakistan and their specification. Case Study 1.

As the variable *y* under study, the authors chose the incidence of active TB (number of cases), and as explanatory (exogenous) variables, they defined the indicators of the role of the state in overcoming the TB epidemic in the Kyrgyz Republic, presented in Table 1. These indicators include state budget expenditures on housing and communal services, healthcare, social protection, and environmental protection. The authors also included the poverty indicator as an intermediate indicator for the overall economic condition of the country. The data (Table 1) are official data from the National Statistical Committee of the Kyrgyz Republic (NSC KR) and the Ministry of Health [6,7,8,9,17].

Table 1. Active TB morbidity, the structure of state budget expenditures, and the level of poverty in KR.

Years	Incidence of active TB, number of cases	Structure of state budget expenditures, million soms				Total below income poverty line, %
		Housing and utilities	Healthcare	Social security	Environment protection	
	y	x_1	x_2	x_3	x_4	x_5
2012	5851	3002.7	9714	15951	438.1	36.02
2013	5859	3121.4	10171.1	18429.9	411.2	36.98
2014	5898	3563.2	10443	21060.9	516.8	30.6
2015	5853	3858.6	10538.6	22703	580.8	32.1
2016	5680	4444.4	11188	23313.1	680.1	25.4
2017	5616	4761.4	13168.1	24384.3	574.5	25.6
2018	5249	4381.3	11955.6	27376.6	670.2	22.4
2019	5096	4428.3	12139.8	29830.3	595.2	20.1
2020	3518	4623.8	15164.4	31149.5	514.3	25.29
2021	3891	5109.2	18925.7	33088.4	541.5	23.27

Paired correlation coefficients were determined using MS Excel tool Data Analysis/Correlation add-on and presented in Table 2 as a correlation matrix.

Table 2. Correlation matrix.

	x_1	x_2	x_3	x_4	x_5
y	-0.669	-0.875	-0.847	-0.035	0.497
x_1	1	0.806	0.890	0.626	-0.891
x_2		1	0.843	0.117	-0.659
x_3			1	0.448	-0.881
x_4				1	-0.781
x_5					1

As seen from Table 2, the pairwise correlation between the variable under study y – incidence of TB – and the first influencing factor x_1 is moderate and inverse, since the corresponding correlation coefficient $r_{yx1} \approx -0.67$ and is in the range of $0.3< \left| r_{yx_1} \right| < 0.7$. The value of the correlation coefficient between y and $\boldsymbol{x_5}$ also indicates a moderate, but positive/direct relationship. The presence of moderate correlation is grounds for including the factors x_1 and $\boldsymbol{x_5}$ in the model. The correlation between the resultant variable – incidence of TB y – and the factors x_2 and x_3 is strong and inverse, i.e. $r_{yx_2} \approx -0.87$, a $r_{yx_3} \approx -0.85$. The influence of the factor x_4 on y is insignificant, so it cannot be included in the model.

Sufficiently strong inter-factor correlation can be explained by the fact that the factors are dynamic series, but this does not prevent application of regression analysis to them as in the case of cross-sectional data [10].

Thus, the following can approximate the dependence of TB incidence in KR multiple linear regression model:

$$y_{KR} = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_5x_5 + \varepsilon. \quad (1)$$

Unfortunately, it was not possible to obtain a similar set of raw data for the Islamic Republic of Pakistan. The explanation could be the significant difference in the state systems of maintaining statistical records of socio-economic indicators of the two countries. However, as a result of processing the data provided in the sources [11,12,13] the indicator values presented in Table 3 were obtained.

Table 3. Tuberculosis prevalence, current health expenditures and poverty in Pakistan.

Year	TB incidence, N of cases per 100.000 people	TB incidence, (N of cases per 100.000 people)/10^{-2}	Current health expenditure, % of GDP	Poverty, % of population
		y_{Pak}	z_1	z_2
2009	276	2.76	2.61	37.9
2010	276	2.76	2.6	36.8
2011	276	2.76	2.34	36.3
2012	276	2.76	2.36	32.5
2013	275	2.75	2.6	29.6
2014	270	2.7	2.72	25
2015	270	2.7	2.69	24.3
2016	268	2.68	2.89	23.4
2017	267	2.67	2.9	22.9

2018	265	2.65	3.2	21.9
2019	263	2.63	3.38	20.4
2020	259	2.59	3.65	19.7

In Table 3 y_{Pak} – is the incidence of TB, presented as the number of cases per 100,000 population. For further processing, this number was converted to the $N/10^2$ (or $N/10^2$) form. The influencing factor z_1 was taken to be the current expenditure of the state budget of the Republic of Pakistan as a percentage of GDP. The influencing factor z_2 is the poverty rate as a percentage of the total population.

It should be noted that the percentage of poor population is almost the same in both countries under consideration. How effective is the role of the government in the Kyrgyz Republic and Pakistan in striving to stop tuberculosis?

The correlation matrix for Pakistan's baseline data is shown in Table 4.

Table 4. Correlation matrix for Pakistan's baseline data.

	z_1	z_2
y_{Pak}	-0.931	0.933
z_1	1	-0.795
z_2		1

The correlation coefficients between the dependent variable y_{Pak} and the z_1 and z_2 factors are close to unity in absolute value and are $r_{yz1} = -0.931$ and $r_{yz2} = 0.933$. A negative r_{yz1} indicates that an increase in state budget expenditures on healthcare has a significant inverse effect on the incidence of TBI, causing it to decrease. Conversely, a positive r_{yz2} indicates that the incidence of TB in Pakistan will increase with increasing levels of poverty. The correlation between the factors is also quite high, which is due to the dynamic nature of the raw data. Consequently, the available data allow the authors to construct and estimate qualitative characteristics only for a two-factor linear regression model of the form

$$y_{Pak} = b_0 + b_1 z_1 + b_2 z_2 + \varepsilon. \quad (2)$$

Since the authors defined above the task of comparative analysis of the current TB situation in the two countries, the composition of the variables to be included in the models should be similar. This means that for the KR it would be reasonable to run also a regression analysis with a two-factor model of the form

$$y_{KR} = a_0 + a_2 x_2 + a_5 x_5 + \varepsilon, \quad (3)$$

where x_2 and x_5 are the inputs in Table 1.

3. Results

3.1. Construction and quality analysis of a 2-factor linear regression model of TB incidence for Kyrgyzstan

Using the Data Analysis/Regression add-on of MS Excel spreadsheets allowed obtaining the following results of regression analysis for Kyrgyzstan (Fig. 1):

	A	B	C	D	E	F
1	SUMMARY OUTPUT					
2						
3	Regression statistics					
4	Multiple R	0.8993602				
5	R-square	0.808848769				
6	Normalized R	0.754234131				
7	Standard error	427.8922454				
8	Observations	10				
9						
10	Analysis of variance					
11		*df*	*SS*	*MS*	*F*	*Significance of F*
12	Regression	2	5423218.484	2711609.242	14.81010963	0.003053654
13	Residual	7	1281642.416	183091.7737		
14	Total	9	6704860.9			
15						
16		*Coefficients*	*Standard error*	*t-statistics*	*P-value*	*Lower 95%*
17	Y-intersection	11073.59895	2096.35943	5.282299778	0.001145184	6116.496602
18	Variable X 1	-0.094210161	0.112481477	-3.504667097	0.009930419	-0.660186559
19	Variable X 2	-37.3733499727	29.49345084	-1.265874159	0.246058717	-107.0759264

Fig. 1. Regression analysis outputs for the 2-factor model of TB incidence in Kyrgyzstan.

According to the results shown in Fig. 1, the multiple correlation coefficient has a value of R = 0,899 ≈ 0,9, which indicates a very strong relationship between the variable under study y and the factors included in the regression model. Excel also produces a normalized coefficient of determination. Its value is slightly lower and equals 0.754. As the sample size increases, the difference between these parameters decreases and the square of the correlation coefficient can be used. In current case, the value of R^2 = 0.81 means that 81% of the variation of the outcome variable – morbidity y – can be explained by the variation of the chosen factors, which are the state budget expenses on health care and the poverty level of the population, and the remaining 19% by other influencing factors not considered in the model. The R^2 = 0.81 is sufficiently high and acceptable to assess the current TB situation in the Kyrgyz Republic. Since the R^2 coefficient of determination is a characteristic of model quality, the built regression model can be considered as a model with high quality level [14].

Fisher's criterion F characterizes statistical significance of the regression equation as a whole. Actual (observed) value of the criterion is F = 14.81 and the corresponding significance level α = 0.003. The value of α is close to zero and since it equals the probability of error, it means that the regression equation can be recognized as statistically significant with a very high probability for the built model.

The equation of the 2-factor regression model (3) in accordance with Fig. 1 can be written in the following way

$$y_{KR} = 11073.6 - 0.394x_2 - 37.335x_5 + \varepsilon. \tag{4}$$

Statistical significance of model coefficients (4) is estimated by the Student's t-test. The table (Fig. 1) shows that significance of the coefficients a_0 = 11073.6 and a_2 = 0.394 are high, and coefficient a_5 = 37.335 can be considered "more or less significant" with probability from 0.7 to 0.95 [15].

When the regression model (3) and, respectively, (4) were built, it was suggested "that the real interrelation of variables is linear and deviations from the regression model are random, independent of each other and have zero mean and constant variance" [15]. Therefore, the next stage of checking the quality of model (4) is to check the listed conditions. In current case, all variables are time series, so it is necessary to calculate the Durbin-Watson statistics [15]:

$$DW_{KR} = \frac{\sum_i (e_i - e_{i-1})^2}{\sum_i e_i^2},$$

where e_i – are the deviations of the initial values of y_i (table 1) from the values of y_{KR}, calculated by the regression formula (4); e_{i-1} – are the same deviations, but shifted by one level of the time series. Calculations give the value of $DW_{KR} = 1.309$, which proves the absence of residuals autocorrelation.

Therefore, taking into account the entire set of quality indicators of the model (4), it can be concluded that formula (4) reflects the real relationship and that another non-linear model will not be better than this one by its characteristics.

3.2. Construction and quality analysis of a 2-factor linear regression model of TB incidence for Pakistan

The results of the regression analysis for Pakistan are presented in Figure 2. According to these results, the 2-factor linear regression model of TB incidence for Pakistan is

$$y_{Pak} = b_0 + b_1 z_1 + b_2 z_2 + \varepsilon = 2.814 - 0.078\, z_1 + 0.004 z_2 + \varepsilon \qquad (5)$$

	A	B	C	D	E	F
1	SUMMARY OUTPUT					
2						
3	*Regression statistics*					
4	Multiple R	0.9767728				
5	R-square	0.954085103				
6	Normalized R	0.940966561				
7	Standard error	0.010587632				
8	Observations	10				
9						
10	Analysis of variance					
11		*df*	*SS*	*MS*	*F*	*Significance of F*
12	Regression	2	0.016305314	0.008152657	72.7279831	2.07414E-05
13	Residual	7	0.000784686	0.000112098		
14	Total	9	0.01709			
15						
16		*Coefficients*	*Standard error*	*t-statistics*	*P-value*	*Lower 95%*
17	Y-intersection	2.813794911	0.078537724	35.82730383	3.42711E-09	2.628082703
18	Variable X 1	-0.077952333	0.021345526	-3.651928446	0.008157057	-0/128426482
19	Variable X 2	0.003956463	0.00086718	4.562448428	0/002597339	0.001905909

Fig. 2. Regression analysis outputs for the 2-factor model of TB incidence in Pakistan.

- Determination coefficient $R^2 = 0.954$;
- Fourier criterion value F = 72.728 has a high significance of 2.074·10-5;

- All three coefficients of the model (5) have Student's *t*-statistics of very high significance because *P*-values are very small;
- Durbin-Watson statistic DW_{Pak}= 2.046. To reliably determine whether there is "autocorrelation of residuals in the model (5), the significance level $\alpha = 0.05$ is set. Using Durbin-Watson criterion tables for the number of observations $n = 10$ and the number of independent variables of the model $m = 2$, the critical values of the statistics $dL = 0.697$ and $dU = 1.641$ are determined. The intervals within the interval [0; 4] shown in Fig. 3 are obtained. The actual value falls within the interval from dU to (4 - dU). So, there is no reason to reject the H_0 about the absence of autocorrelation in the residuals" [16].

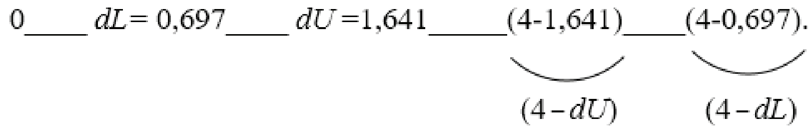

Fig. 3. Intervals within the interval [0; 4].

4. Discussion

4.1. Evaluating the Effectiveness of the Role of the State in the Kyrgyz Republic and Pakistan in Implementing the WHO Global Stop TB Program. Case Study 2.

This assessment can be made by applying the constructed regression models (4) and (5) to predict the incidence of tuberculosis in these countries. Unfortunately, the baseline data for both countries cover a period of approximately ten years. Consequently, it is only possible to obtain predictions of TB incidence with reliability corresponding to the values of the coefficient of determination R^2, for the next 3 to 3.3 years, i.e. no further than one third of the period for which the raw data for y and x_m factors (m are the x factors in models 4 and 5).

According to the results of the regression analysis of models (4) and (5), the standard errors are S_{KR} = 427.89 (Fig. 1) and S_{Pak} = 0.0106, and the mean values for the original samples are $\bar{y}_{KR}$ = 5251.1; $\bar{y}_{Pak}$ = 2.7. The ratio of these values is denoted by V [15] and can be considered as a relative forecast error of the regression model. Thus,

$$V_{KR} = \frac{S_{KR}}{\bar{y}_{KR}} = \frac{427.89}{5251.1} = 0.0815 = 8.15\%$$

$$V_{Pak} = \frac{S_{Pak}}{\bar{y}_{Pak}} = \frac{0.0106}{2.7} = 0.0039 = 0.39\%$$

Calculations using model (4) show that the decrease of TB incidence in the Kyrgyz Republic by the end of 2024 could be about 7.45%. Taking into account the forecast error of 8.15%, a deviation in the range of (7.45 ± 0.6%) can be expected. From the authors' point of view, the role of the Kyrgyz Republic state in implementing the WHO Stop TB program can be considered satisfactory, especially if the extraordinary situation with COVID-19 during the period of 2020 – 2021 is taken into account.

For Pakistan, the annual decrease in yPak TB incidence by the end of 2023 is expected, according to model (5), from 259 to 258 cases per 100,000 population, and the annual decrease in TB incidence to 257 cases per 100,000 population is expected by 2024, which is ≈ 0.4% as a percentage. Taking into account the value of relative forecast errorV_{Pak}, the decrease in y_{Pak} could be in the range of (1 ± 0.4) %. But such a decrease, unfortunately, is far from the desired target of 10%.

5. Conclusion

Two case studies have been developed to demonstrate the practical application of evidence-based medicine/healthcare as part of the medical school curriculum to assess the current implementation of WHO TB control strategies in two different countries.

The 2-factor linear regression models (4) and (5) were built basing on open statistical data on TB incidence rate, state budget expenditures on healthcare and poverty rate in the Kyrgyz Republic and Islamic Republic of Pakistan, taking into account the influence of such factors as state allocations for healthcare and percentage of poor population on TB incidence in these countries.

The quality characteristics of the models obtained are given below and allow considering them suitable for estimating the current TB situation and its short-term forecast for the next 3 to 3.3 years (Table 5):

Table 5. Quality characteristics of the models obtained.

No.	Parameter	Kyrgyz Republic	Islamic Republic of Pakistan
1.	Determination coefficient R^2	0.81	0.954
2.	Fourier criterion value F	14.81 has a high significance of 0.003	72.728 has a high significance of 2.074·10-5
3.	Model (5)	Two coefficients of model (5) have Student's t-statistics of very high significance, since P-values are very small; the third is more or less significant	All three model coefficients (5) have Student's t-statistics of very high significance because the P-values are very small
4.	Durbin-Watson statistics DW	1.309. There is no autocorrelation of residuals	2.046. There is no autocorrelation of residuals
5.	Predictability of the model V	8.15 %	0.4 %

Calculations using model (4) show that the reduction of TB incidence in the Kyrgyz Republic by the end of 2024 could be about 7.45%. Given a forecast error of 8.15%, a deviation in the range of (7.45 ± 0.6%) can be expected. From the authors' point of view, the role of the KR state in implementing the WHO Stop TB program can be considered satisfactory, especially if the extraordinary situation with COVID-19 during the period of 2020 – 2021 is taken into account.

For Pakistan, the annual decrease in TB yPak incidence is expected by the end of 2023, according to model (5), from 259 to 258 cases per 100,000 population, and by 2024 the annual decrease in TB incidence to 257 cases per 100,000 population, which is ≈0.4% in percentage terms. Taking into account the value of relative forecast error V_{Pak}, the decrease of y_{Pak} could be in the range of (1 ± 0.4) %. But such a decrease, unfortunately, is far from the WHO desired target of 10% [7].

The structure and predictive qualities of the models, unfortunately, are objectively limited by the time frame of the initial samples of statistical data, because the data are heterogeneous or absent at all outside this frame.

References

1. S. Manzikova, Bases of Statistical Analysis Medical and Biological Data Using Excel. Educational allowance for foreign students (Neo Print, Bishkek, 2020)
2. WHO. Implementing the end TB strategy: the essentials (2016). Accessed on: November 08, 2022. [Online]. Available: https://www.afro.who.int/publications/implementing-end-tb-strategy-essentials
3. WHO. The global plan to stop TB 2011-2015: transforming the fight towards elimination of tuberculosis (2010). Accessed on: November 08, 2022. [Online]. Available: https://apps.who.int/iris/handle/10665/44437

4. WHO. Global tuberculosis report (2016). Accessed on: November 08, 2022. [Online]. Available: https://www.aidsdatahub.org/resource/global-tuberculosis-report
5. S.Ts. Manzhikova, Ch.S. Ismailova, KRSU Bul. **21(4)**, 96-104 (2021)
6. Ministerstvo zdravookhraneniya Kyrgyzskoi Respubliki. Zabolevaemost tuberkulezom [Ministry of Health of the Kyrgyz Republic. Tuberculosis incidence] (2021). Retrieved from https://data.gov.kg/dataset/zabolevaemost-tuberkulezom
7. Strategiya VOZ po likvidatsii tuberkuleza k 2035 godu [WHO strategy to eliminate tuberculosis by 2035] (2022). Retrieved from https://www.trbzdrav.ru/article/infection/who-strategy-for-elimination-of-tuberculosis-by-2035.php
8. Natsionalnyi statisticheskii komitet KR. Sotsialno-ekonomicheskaya situatsiya v Kyrgyzskoi Respublike [National Statistical Committee of the KR. Socio-economic situation in the Kyrgyz Republic] (2014). Accessed on: November 08, 2022. [Online]. Available: http://www.stat.kg/media/publicationarchive/3ccac53d-06d3-4728-b380-3434473214e7.pdf
9. Natsionalnyi statisticheskii komitet KR. Sotsialno-ekonomicheskaya situatsiya v Kyrgyzskoi Respublike [National Statistical Committee of the KR. Socio-economic situation in the Kyrgyz Republic] (2022). Accessed on: November 08, 2022. [Online]. Available: http://www.stat.kg/ru/publications/doklad-socialno-ekonomicheskoe-polozhenie-kyrgyzskoj-respubliki/
10. E. Förster, B. Rönz, Metody korrelyatsionnogo i regressionnogo analiza. Rukovodstvo dlya ekonomistov [Methods of correlation and regression analysis. A guide for economists] (Finance and Statistics, Moscow, 1983)
11. Pakistan: Statistika infektsionnykh zabolevanii [Pakistan: Infectious disease statistics] (2020). Accessed on: November 08, 2022. [Online]. Available: https://dibit.ru/p/stats/pak/infect#stats-source-data-tb-e-inc-num
12. Pakistan Economic Indicators (2022). Accessed on: November 08, 2022. [Online]. Available: https://www.theglobaleconomy.com/Pakistan/
13. Current health expenditure (% of GDP) – Pakistan (2022). Accessed on: November 08, 2022. [Online]. Available:https://data.worldbank.org/indicator/SH.XPD.CHEX.GD.ZS?locations=PK
14. Model lineinoi mnozhestvennoi regressii. Fiktivnye peremennye. Spetsifikatsiya modeli [Linear multiple regression model. Dummy variables. Model specification] (2018). Accessed on: November 08, 2022. [Online]. Available: http://www.agpu.net/fakult/ipimif/metodmater/ddv005_linmnogregr2.pdf
15. O.O. Zamkov, Yu.N. Cheremnykh, A.V. Tolstopyatenko, Matematicheskie metody v ekonomike [Mathematical Methods in Economics] (Business and service, Moscow, 2009)
16. V.A. Akimov, A.A. Bykov, E.U. Schetinin, Vvedenie v statistiku ekstremalnykh znachenii i ee prilozheniya [An introduction to extreme value statistics and its applications] (All-Russian Research Institute for Civil Defense and Emergencies of the Ministry of Emergency Situations of Russia, Moscow, 2009)
17. Natsionalnyi statisticheskii komitet KR. Dinamicheskie tablitsy [National Statistical Committee of the KR. Dynamic tables] (2022). Accessed on: November 08, 2022. [Online]. Available: http://www.stat.kg/ru/statistics/zdravoohranenie/