pHealth 2022
B. Blobel et al. (Eds.)

doi:10.3233/SHTI220990

Forecasting Trends in the Tuberculosis Epidemic Situation in the Region of the Russian Federation by Dynamic Simulation Model

Igor A. CHERNIAEV [a,1], P. F. CHERNAVIN [b], A. I. TSVETKOV [a], Y. P. CHUGAEV [a], U. I. CHERNIAEVA [a], N. P. CHERNAVIN [b]
[a] *The Ural State Medical University, Yekaterinburg, Russian Federation*
[b] *The Ural Federal University, Yekaterinburg, Russian Federation*

Abstract. The spread of a new coronavirus infection in the last two years together with HIV infection preserves and even increases the potential for the spread of tuberculosis in the world. Sverdlovsk oblast (SO) of Russian Federation is the region with high levels of HIV and tuberculosis (TB). The search for new methods of forecasting of the future epidemic situation for tuberculosis has become particularly relevant. The aim was to develop an effective method for predicting the epidemic situation of tuberculosis using an artificial intelligence (AI) method in the format of a dynamic simulation model based on AI technologies. Statistical data was loaded from the state statistical reporting on TB patients for the period 2007-2017. The parameters were controlled through a system of inequalities. The proposed SDM made it possible to identify and reliably calculate trends of TB epidemiological indicators. Comparison of the predicted values made in 2017 with the actual values of 2018-2021 revealed a reliable coincidence of the trend of movement of TB epidemiological indicators in the region, the maximum deviation was no more than 14.82%. The forecast results obtained with SDM are quite suitable for practical use. Especially, in operational resource planning of measures to counteract the spread of tuberculosis at the regional level.

Keywords. Machine learning, epidemiology of tuberculosis, dynamic simulation model

Introduction

Over the past few years, there has been a tendency in the world to reduce the incidence of tuberculosis by about 2% per year. The fastest decline at the regional level was observed in the WHO European Region (5% per year), as well as in the Russian Federation (5% per year) [1]. For a long time, the Russian Federation was on the list of 22 two countries with the greatest burden of tuberculosis, but thanks to systematic efforts to counteract the spread of the infection [1,2,3,4,5,7], it was possible to achieve a significant reduction in the morbidity and mortality of the country's population from tuberculosis and to be excluded from this list. However, Russia remains among the

[1] Corresponding Author. Igor A. Cherniaev, The Ural State Medical University, Yekaterinburg, Russian Federation: Email: icherniaev@gmail.com

countries with the highest prevalence of MDR and XDR Mycobacterium tuberculosis, as well as tuberculosis in combination with HIV [8].

According to WHO, the spread of a new coronavirus infection and HIV infection preserves and even increases the potential for the spread of tuberculosis in the world [4]. The Sverdlovsk Region is a region with high levels of HIV and tuberculosis (TB), one of the leaders in the number of patients with HIV+TB co-infection, and the general epidemiological trends indicated above are also evident in this region [9,10]. The need to combat the new coronavirus infection has caused the redistribution of already limited health resources. Therefore, in the new epidemiological conditions, there are required detailed planning and calculation of resource requirements for countering the spread of tuberculosis [11,12]. In this connection, a search for new methods of analyzing the current and predicting the characteristics of the future epidemic situation for tuberculosis becomes particularly relevant.

The aim of this study is to develop an effective method to predict the tuberculosis epidemic situation by using of the AI method on the example of the Sverdlovsk region.

1. Material and Methods

The mathematical models of the spread of tuberculosis have more than half a century of history. In the early 60s of the twentieth century, Hans Waaler [13] and his collaborators proposed using of a mathematical model based on difference equations to analyze the epidemiological situation. In the subsequent years, mathematical modeling was actively used in the works of S. Brogger, K. Stiblo, E. Vinniki and P. Fine and a number of other authors [14,15,16]. A detailed review of these models and a proposal of their own are contained in the work of K. K. Avilov and A. A. Romaniukha [17].

In 2017, the authors of this article began to work on modeling of the tuberculosis epidemiological situation in the Sverdlovsk region. A comprehensive assessment of the epidemiological situation for the period 2012-2017 was carried out in the real socio-economic conditions of a large industrial region in appliance with the recommended methods [18, 19]. According to the results of the assessment, it was found that there are multidirectional trends in the main epidemiological indicators of tuberculosis in the subject. As the in-depth analysis of variance has shown, despite the high level of variation of indicators in the context of the municipalities, a unified epidemic process formed in the Sverdlovsk region, which makes it possible to identify typical situations and their corresponding organizational, managerial and planning measures.

The task was to give a forecast of the development of the tuberculosis epidemic situation in the near future and until 2030 in the whole region and for each of the districts, provided that the prevailing trends continue. It was decided to use a single mathematical model to assess the situation in the context of each municipality and the region as a whole, in total. This approach is due to the fact that the socio-economic and climatic conditions of the individual areas differ significantly, so the development of the epidemic situation can be described by the different features.

At the initial stages of the study, K. Stiblo's epidemiological model modified by the authors, including a system of the difference equations by G. Vaaler, was used, taking into account the limitations and recommendations of Avilov K.K., Romaniukha A.A. [15, 17]. The study was conducted on the material of the Sverdlovsk region, a region with a population of more than 4.5 ml with a high prevalence of tuberculosis and HIV. Information from more than 45,000 medical records of patients was used.

The model is based on statistical data obtained from the official forms of state statistical reporting: Form # 8 "Information on active tuberculosis diseases" and Form # 33 "Information on tuberculosis patients", forms 089-u/tub, data from the Federal Register of Tuberculosis Patients, registers of tuberculosis patients of health care facilities of the Sverdlovsk region, received for the period 2007-2017. [20,21,22,23]. Based on the information obtained using Excel, complex analytical tables of absolute values and epidemiological coefficients for variables were compiled in the context of 63 municipalities of the Sverdlovsk region and the region as a whole. Additional controllable parameters are provided to improve the accuracy of the model. When working with the model, the parameters were controlled by setting the absolute rates of their growth, from the previous year to the next. This approach is legitimate, since the formation of the rate of change of parameters is of a long-term nature and is formed by a set of measures carried out over many years. The use of the Greek-Latin alphabet was minimized, mnemonically understandable abbreviations based on Cyrillic alphabet were used to designate variables. The main and additional controlled parameters of the model are listed in Table 1. They are conditionally constant values for each municipal district and have a weak influence on trends in general. In the mathematical record, they are combined as a single variable other changes (at).

Table 1. Model parameters

Main and additional controlled model parameters	Additional managed model parameters
The main controlled parameters of the model	Relapses
The percentage of cure (transferred to clinical cured)	got sick in contacted with MBT+ TB cases
The percentage of MBT + in active TB	got sick in contacted with MBT- TB cases
The coefficient of contact with MBT +	Mortality from other causes
Percentage of newly identified (new cases) MBT+ in contact	
Percentage of in/in MBT- in contact	
Percentage of mortality	

Description of designations of quantitative and qualitative variables of the model are presented in Table 2.

Table 2. Designations of quantitative and qualitative variables of the model

Quantitative variables		Qualitative variables	
Sign	**Description**	**Sign**	**Description**
T	an index of a time period	pv(t)	percentage of cured in period t
T	a number of periods under consideration	pbcp(t)	percentage of MBT+ in period t
A(t)	a number of active TB cases in a period t	kkbkp(t)	a coefficient of contact with MBT+ in period t
B(t)	a number of cured TB cases in a period t	pvvbkp(t)	percentage of newly identified MBT+ in period t
BCP(t)	a number of MBT+TB cases in a period t		percentage of newly identified MBT- in period t
BCM(t)	a number of MBT-TB cases TB in a period t	pu(t)	percentage of deaths from tuberculosis in period t
KBKP(t)	a number of contacts with MBT+TB in a period t	pvvbkm(t)	percentage of other changes in a period t
VVBKP(t)	a number of newly Identified MBT+ in a period t		
VVBCM(t)	a number of newly Identified MBTs- in a period t		
UT(t)	a number of deaths from tuberculosis in a period t		
At(t)	other Changes in a period t		

The epidemiological process is described by the following system of equations, determining the number of the active TB cases:

$$A(t+1) = A(t)-B(t)-UT(t)+BBBCP(t)+VVBKM(t)+PI(t) \quad t=0,..,T \qquad (1)$$

Calculation of quantitative variables of the current period:

$$B(t) = pv(t)*A(t)/100 \quad t=0,..,T \tag{2}$$

$$BCP(t) = pbcp(t)*A(t)/100 \quad t=0,..,T \tag{3}$$

$$BCM(t) = A(t)- BCP(t) \quad t=0,..,T \tag{4}$$

$$BCP(t) = BCP(t)*BCP(t) \quad t=0,..,T \tag{5}$$

$$VVBKP(t) = pvvbkp(t)*KBKP(t)/100 \quad t=0,..,T \tag{6}$$

$$VVBKM(t) = pvvbkm(t)*KBKP(t)/100 \quad t=0,..,T \tag{7}$$

$$UT(t) = pu(t)*A(t)/100 \quad t=0,..,T \tag{8}$$

$$PI(t) = ppi(t)*A(t)/100 \quad t=0,..,T \tag{9}$$

Thus, 7 qualitative indicators were used in the model and the increase or not in the number of patients with active tuberculosis in the next period in comparison with the previous one depends on their values. To assess the situation as a whole, it is advisable to have a single indicator that evaluates the qualitative characteristics in the aggregate, which we call the Coefficient of General Dynamics or abbreviated COD. This indicator should have the following properties, if it is strictly less than 1, then the number of active decreases, if strictly more, then increases. The value of this indicator is calculated within our model by the equation:

$$COD = 1-pv/100+pbkp*kkbkp(pvvbkp+pvvbkm)/10000-pu/100+ppi/100 \tag{10}$$

From this Ex. (10) it follows how each of the qualitative parameters affects the development of the epidemiological situation. In addition, it is possible to calculate what the value of each of them should be when fixing the others so that the COD is less than one. For example, the Minimum Percentage of Cure (MPV) in this case will be calculated as follows:

$$\mathrm{MPV = pbkp*kkbkp*(pvvbkp+pvvbkm)/100-pu+ppi} \tag{11}$$

The code and the MPV in this form rather simplistically describe the development of the situation, but are meaningfully interpreted and are quite suitable for approximate calculations. Based on the available data, 16 parameters of the development of the epidemic situation for tuberculosis until 2030, including the COD and the MPV, were calculated for all districts of the region and the region as a whole. In 2022, the authors had the opportunity to compare the calculated values of the model (forecast) with the actual ones for 2018-2021, registered by means of state statistical observation [20,21,22,23], which was done to verify the relevance of the model. The model was verified by comparing the calculated values with the actual values obtained in the period 2018-2021. To determine the statistical reliability of the differences, the Student's coefficient t was calculated [24]. The accuracy of the forecast can influence the quality of the decisions made [25]. The level of reliability of the model for each forecast period is verified by calculating the indicator of the clarity of the ratio of actual values with predicted values, the indicators of forecast accuracy are calculated. The average absolute percentage error ("Mean Absolute Percentage Error"- MAPE) is calculated using the equation (12) [26].

$$\text{MAPE} = \frac{1}{h} * \sum_{j=1}^{h} \frac{|e_{T+j}|}{y_{T+j}} \quad (12)$$

The reliability of the forecast is estimated through the coefficient of symmetric mean absolute percentage error ("Symmetric MAPE" – SMAPE) according to the Ex. (13) proposed by Makridakis, S. and Hibon, M. 1979 [27] taking into account the possibilities and limitations of the method [28]:

$$S\text{MAPE} = \frac{1}{h} * \sum_{j=1}^{h} \frac{2*|e_{T+j}|}{y_{T+j}+\widehat{y_{T+j}}} \quad (13)$$

The estimation of the root mean square deviation (RMSD) is applied according to the Ex. (14) [29]:

$$RMSD = \sqrt{\frac{1}{N} * \sum_{t=1}^{N} (y(t) - ME)^2} \quad (14)$$

2. Results

The results of the model analysis are shown in Table 3. The method of graphical analysis is applied to compare the predicted trend with the actually registered values in the period 2018-2021.

Table 3 Comparison of the calculated values (forecast) of the model with the actual (fact)

Name	2019 Forecast	Fact/Fore-cast, %	2020 Forecast	Fact/Fore-cast, %	2021 Forecast	Fact/Fore-cast, %
Number of Active TB cases	77146	4,84	6931	-11,51	6770	-20,92
Percentage of cured, %	227,54	-5,56	27,21	45,87	26,93	29,56
Transferred to clinical cured group	11968	-0,97	1886	29,06	1823	2,47
Percentage of MBT + in active TB, %	336,82	14,34	36,81	17,63	36,74	18,40
Number of Contacted with MBT+ TB	22631	19,99	2551	4,12	2487	-6,31
Number of Contacted with MBT- TB	22,91	-25,09	2,92	1,03	2,93	16,38
Contacted with MBT+, abs	7668	-10,05	7450	5,10	7280	9,29
Share of MBT - TB cases , %	63,18	-8,36	63,19	-10,27	63,26	-10,69
Total Number of MBT - TB cases	4515	-3,99	4380	-20,62	4283	-29,40
Number of New MBT+TB cases	925	28,97	900	0,44	878	-3,30
Number of New MBT- TB cases	1418	-28,63	1390	-39,78	1371	-34,94
Mortality in TB patients, %	5,95	-17,65	5,89	-6,62	5,88	-0,85
Number of dead from TB	425	-13,41	408	-17,16	398	-21,61

Most of the deviations are due to the fact that the actual values of characteristics such as the Absolute number of deaths from tuberculosis, the total number of active contingents, the MBT+ Contingents first identified by MBT were lower than predicted, which indicates both a positive trend in the epidemic situation and the effectiveness of anti-tuberculosis work in the region. The results of assessing the quality of the forecast using the method are shown in Table 4, from which it follows.

Table 4. Results of the forecast quality assessment by model parameters

Name	Forecast for 2019	Forecast for 2020	Forecast for 2021	Overall forecast 2019-2021
MAPE, %	13,14	15,14	14,87	14,38
SMAPE, %	13,54	15,24	15,68	14,82
SD, %	16,08	17,76	20,52	18,12

3. Conclusion

The proposed mathematical simulation dynamic model, based on the use of AI, made it possible to identify, reliably calculate and graphically display trends in the movement of values of the studied characteristics of the epidemic process of tuberculosis in each of the 63 municipalities and in the Sverdlovsk region as a whole.

Comparison of the calculated values (forecast) of the model made in 2018 with the actual values of 2019-2021 revealed a reliable coincidence of the trend of movement of the studied parameters in the region. If we assume that the quality indicators will remain at the level of 2018, then in the next 20 years the epidemiological situation on tuberculosis in the region will develop as follows.

The forecast results obtained with the help of a simulation model are quite suitable for practical use in operational resource planning for measures to counteract the spread of tuberculosis at the regional level.

References

[1] Global Tuberculosis Report 2018. WHO. [Internet]. URL: https://apps.who.int/iris/bitstream/handle/10665/274453/9789241565646-eng.pdf (12 /04/ 2022)].

[2] Global Tuberculosis Report 2019. WHO. [Internet]. URL: https://apps.who.int/iris/bitstream/handle/10665/329368/9789241565714-eng.pdf (12 /04/ 2022).

[3] Global Tuberculosis Report 2017. WHO. [Internet]. URL: https://www.who.int/teams/global-tuberculosis-programme/tb-reports (12 /04/ 2022).

[4] Global Tuberculosis Report 2020. WHO. [Internet]. URL: https://apps.who.int/iris/bitstream/handle/10665/336069/9789240013131-eng.pdf (12 /04/ 2022)..

[5] Global Tuberculosis Report 2018. WHO. [Internet]. URL: https://apps.who.int/iris/bitstream/handle/10665/274453/9789241565646-eng.pdf (12/04/2022.).

[6] Global Tuberculosis Report 2016. WHO. [Internet]. URL: https://apps.who.int/iris/bitstream/handle/10665/250441/9789241565394-eng.pdf?sequence=1&isAllowed=y (Дата обращения 12 апреля 2021 г.).

[7] Syunyakova D.A. Features of the epidemiology of tuberculosis in the world and in Russia in the period 2015-2020. Analytical survey. *Social'nye aspekty zdorov'a naselenia / Social aspects of population health* [serial online] 2021; 67(3):11. Available from: http://vestnik.mednet.ru/content/view/1273/30/lang,ru/. DOI: 10.21045/2071-5021-2021-67-3-11. (In Rus).

[8] WHO releases new global lists of high-burden countries for TB, HIV-associated TB and drug-resistant TB 17 June 2021, WHO. [Интернет]. URL:https://www.who.int/news/item/17-06-2021-who-releases-new-global-lists-of-high-burden-countries-for-tb-hiv-associated-tb-and-drug-resistant-tb (Дата обращения 17 июня 2022 г.)]

[9] Nechaeva O.B., Podymova A.S. (2018). Vliyanie VICH-infekcii na demograficheskuyu situaciyu v Rossii. Medicinskij al'yans, 1, 6-15.(In Russian)

[10] Astrelin A.M. Tendencii zabolevaemosti, rasprostranennosti i smertnosti ot VICH-infekcii i tuberkuleza v regionah Rossii v XXI veke. Demograficheskoe obozrenie. 2020. TOM 7, №4:82-107.(In Russian)

[11] Starodubov V.I., Kadyrov F.N., Obuhova O.V., Bazarova I.N., Endovickaya YU.V., Nesvetajlo N.YA. «Vliyanie koronavirusa COVID-19 na situaciyu v rossijskom zdravoohranenii» Analiticheskij doklad: Pod obshchej redakciej akademika RAN Starodubova V.I. "Central'nyj nauchno-issledovatel'skij institut organizacii i informatizacii zdravoohraneniya" Ministerstva zdravoohraneniya Rossijskoj Federacii, Moskva 2020. 45 p. . [Internet]. URL: https://mednet.ru/images/materials/news/doklad_cniioiz_po_COVID-19-2020_04_26.pdf Data obrashcheniya 17 iyunya 2022 g.)(In Russian)

[12] Obshchenacional'nyj plan dejstvij, obespechivayushchih vosstanovlenie zanyatosti i dohodov naseleniya, rost ekonomiki i dolgosrochnye strukturnye izmeneniya v ekonomike (odobren Pravitel'stvom RF 23.09.2020, protokol № 36, razdel VII) (N P 13-60855 ot 02.10.2020) . [Internet]. URL: http://www.consultant.ru/document/cons_doc_LAW_333667/ (Data obrashcheniya 17 iyunya 2022 g.)(In Russian)

[13] Waaler, H.T., Geser, A., Andersen, S. The use of mathematical models in the study of the epidemiology of tuberculosis. Am. J. publ. Health. 1962. 52. 1002–1013.

[14] Brogger S. Systems analysis in tuberculosis control: A model. American Review of Respiratory Disease. 1967. 95(3). 419–434

[15] Styblo K, Bumgarner J.R. Tuberculosis can be controlled with existing technologies: evidence. The Hague: Tuberculosis Surveillance Research Unit Progress Report; 1991. p. 60–72.

[16] Vynnycky E., Fine P.E.M. The natural history of tuberculosis: the implications of age dependent risks of disease and the role of reinfection. Epidemiol. Infect. 1997. 119.183–201.

[17] Avilov K.K., Romanyuha A.A. Matematicheskie modeli rasprostraneniya i kontrolya tuberkuleza // Matematicheskaya biologiya i bioinformatika. 2007. T. 2, №2.PP. 188-318. (http://www.matbio.org/downloads/ Avilov2007(2_188).pdf)[Mathematical models of the spread and control of tuberculosis] (In Russian)

[18] Metodika analiza epidemicheskoj situacii po tuberkulezu: metod. rekomendacii / NII ftiziopul'monologii MMA im. I. M. Sechenova; razrab. M. V. SHilova. - M., 2007. - 56 p. [Methodology for analyzing the epidemic situation of tuberculosis: methodical guidelines] (In Russian)

[19] Son I. M., Leonov S. A., Sel'covskij P. P. i dr. Ocenka epidemicheskoj situacii po tuberkulezu i analiz deyatel'nosti protivotuberkuleznyh uchrezhdenij. - M.: CNIIOIZ, 2007. - 64 p. [Assessment of the epidemic situation of tuberculosis and analysis of the activities of anti-tuberculosis institutions](In Russian).

[20] Prikaz Federal'noj sluzhby gosudarstvennoj statistiki ot 31 dekabrya 2010 g. № 483 «Ob utverzhdenii statisticheskogo instrumentariya dlya organizacii Minzdravsocrazvitiya Rossii federal'nogo statisticheskogo nablyudeniya za deyatel'nost'yu uchrezhdenij sistemy zdravoohraneniya», [Order of the Federal State Statistics Service No. 483 dated December 31, 2010 "On Approval of statistical tools for the Organization of the Ministry of Health and Social Development of Russia of Federal statistical monitoring of the Activities of Healthcare Institutions"](In Russian).

[21] Prikaz Rosstata ot 28.01.2009 № 12 (red. ot 20.01.2017) «Ob utverzhdenii statisticheskogo instrumentariya dlya organizacii Minzdravsocrazvitiya Rossii federal'nogo statisticheskogo nablyudeniya v sfere zdravoohraneniya»,[Order of the Federal State Statistics Service No. 483 dated December 31, 2010 "On Approval of statistical tools for the Organization of the Ministry of Health and Social Development of Russia of Federal statistical monitoring of the Activities of Healthcare Institutions"] (In Russian).

[22] Prikaz Ministerstva Zdravoohraneniya Rossijskoj Federacii ot 13 avgusta 2003 g. № 410 «Ob utverzhdenii uchetnoj formy № 089/u-tub «Izveshchenie o bol'nom s vpervye v zhizni ustanovlennym diagnozom aktivnogo tuberkuleza, s recidivom tuberkuleza»Order of the Ministry of Health of the Russian Federation No. 410 dated August 13, 2003 "On approval of the registration Form No. 089/u-tub ["Notification of a patient with a diagnosis of active tuberculosis for the first time in his life, with a relapse of tuberculosis”](In Russian).

[23] Ob utverzhdenii Pravil vedenija Federal'nogo registra lic, inficirovannyh virusom immunodeficita cheloveka, i Federal'nogo registra lic, bol'nyh tuberkulezom [On the approval of the Rules for maintaining the Federal Register of Persons Infected with the Human Immunodeficiency Virus and the Federal Register of Persons with Tuberculosis.]. Postanovlenie Pravitel'stva Rossijskoj Federacii ot 8 aprelja 2017 g. № 426 [Resolution of the Government of the Russian Federation No. 426 dated April 8, 2017] (In Russian).]

[24] Glanc S. Mediko-biologicheskaya statistika. Per. s angl. — M., Praktika, 1998. — 459 p.[Glantz S.. Biomedical statistics. Trans. from English.](In Russian).

[25] Ahmar, Ansari. (2020). Forecast Error Calculation with Mean Squared Error (MSE) and Mean Absolute Percentage Error (MAPE). JINAV: Journal of Information and Visualization. 1. 10.35877/454RI.jinav303.

[26] Myttenaere, Arnaud & Golden, Boris & Le Grand, Benedicte & Rossi, Fabrice. (2015). Using the Mean Absolute Percentage Error for Regression Models. Neurocomputing. 192. 10.1016/j.neucom.2015.12.114.

[27] Makridakis, Spyros & Hibon, Michele. (1979). Accuracy of Forecasting: An Empirical Investigation. Journal of the Royal Statistical Society. Series A (General). 142. 10.2307/2345077.

[28] Goodwin, Paul & Lawton, Richard. (1999). On the asymmetry of the symmetric MAPE. International Journal of Forecasting. 15. 405-408. 10.1016/S0169-2070(99)00007-2.

[29] Patrignani et al. (Particle Data Group).39. STATISTICS.— B: Review of Particle Physics // Chin. Phys. C.— 2016.— Vol.40.— P.100001.—doi:10.1088/1674-1137/40/10/100001. [Internet]. URL:https://pdg.lbl.gov/2016/reviews/rpp2016-rev-statistics.pdf (17/06/2022)