Creation of a Forecasting Information System in Epidemiology based on Mathematical Modeling

Aliya Takuadina
Doctoral Student, Department of Informatics and Information Security, L.N. Gumilyov Eurasian National University, Kazakhstan.
E-mail: alyoka.01@mail.ru

Aliya Kintonova
PhD in Technical Sciences, Associate Professor, Department of Informatics and Information Security, L.N. Gumilyov Eurasian National University, Kazakhstan.
E-mail: aliya_kint@mail.ru

Nurlan Tabriz
Doctor of Medicine, Professor, Medical University of Karaganda, Kazakhstan.
E-mail: nctabriz@mail.ru

Anna Knaus
PhD in Medical Sciences, Medical University of Karaganda, Kazakhstan.
E-mail: knaus@mail.ru

Natalya Demidchik
PhD in Technical Sciences, Associate Professor, Department of Computer Engineering, L.N. Gumilyov Eurasian National University, Kazakhstan.
E-mail: demidchik_n_n@mail.ru

Received March 30, 2021; Accepted July 25, 2021
ISSN: 1735-188X
DOI: 10.14704/WEB/V18I2/WEB18328

Abstract

The article is devoted to the problem of using mathematical methods for forecasting the tuberculosis epidemic in Kazakhstan using the example of the Karaganda region. The introduction of the article reflects the problem of forecasting tuberculosis in Kazakhstan. The main part of the article reflects the results of the analysis of mathematical methods for forecasting and data processing technology, describes the mathematical analysis of epidemiological indicators made with the SPSS statistical program, describes the factors affecting the incidence among contact persons, and calculates the correlation coefficient. The article shows the importance of mathematical modeling and the importance of developing a specific mathematical model that describes the spread of infection among the population.

Keywords

Forecasting Information System, Statistical Data Processing Software, Mathematical Methods for Forecasting, Technology in Medicine, Mathematical Analysis of Epidemiological Indicators.

Introduction

Humanity has experienced tuberculosis throughout its history, but only in the last few centuries tuberculosis has become a disease that, in some countries, kills more people than any other infectious disease (in the nineteenth century in industrialized Europe and now in sub-Saharan Africa, it is the most common cause of death). In the middle of the 20th century, the discovery of the first effective tuberculosis treatment stimulated research aimed at enhancing the awareness and the formalization of the principles of tuberculosis transmission and development, as well as at optimizing tuberculosis control programs at the population level.

Tuberculosis (TB) is an infectious disease transmitted by inhalation of bacteria exhaled by an infected person (i.e., by airborne droplets). Mycobacterium tuberculosis is a tuberculosis causative agent. A third of the world's population is infected with TB but only one in ten of them becomes ill.

Symptoms of the disease vary depending on which organs are infected. In the case of pulmonary TB, cough with sputum production (sometimes with blood), shortness of breath, and chest pain are common symptoms. People with HIV or suffering from other health conditions that weaken the immune system (such as diabetes) have a much higher risk of the disease development (Myers, Rogers, Cox, Flahault, & Hay, 2000). MDR-TB is resistant to two of the most potent anti-TB drugs. This phenomenon occurs as a result of insufficient or improper treatment and/or unsatisfactory control of airborne infections in medical institutions and places with a high concentration of a large number of people. XDR-TB is resistant to most of important first and second-line drugs. The chances of recovery from it are very small. Tuberculosis can occur in everyone, but most often, this disease affects young people in the eastern part of Europe, as well as migrants and the elderly in Western European countries. It is closely linked to poverty, migration, places of imprisonment and social marginalization (World Health Organization, 2017).

In the Republic of Kazakhstan, the fight against tuberculosis remains a strategic task and is a priority in the activities of the Ministry of Health of the Republic of Kazakhstan.

Therefore, the study of the emergence, spread and measures to fight epidemics are of great scientific and practical importance. The creation of an intelligent system related to the tasks of epidemiology will help to implement supercomputer technologies into medical and social practice.

In order to solve the problem of preventing the tuberculosis epidemic, a possible outbreak predicted in time plays an important role. Therefore, it is important to create a forecasting system based on the use of mathematical models. Forecasting is the most accurate description of the future taking into account all available information including historical data and knowledge of any future events that may affect forecasts.

This article briefly describes ways to solve this problem with the use of mathematical methods for predicting the tuberculosis epidemic in Kazakhstan using the example of the Karaganda region. The main tool for such research is mathematical models of the spread, development, treatment, and prevention of tuberculosis.

Materials and Methods

Forecasting should be an integral part of management decision-making, as it can play an important role in many areas of human activity. Forecasting systems require the development of experience in identifying forecasting problems, applying a number of forecasting methods, choosing suitable methods for each problem, evaluating and refining forecasting methods over time.

There are the following forecasting stages*:*

1. The determination of what to predict (in the early stages of a forecasting project, one needs to decide what to predict.).
2. Forecasting data and methods (Suitable forecasting methods largely depend on what data is available). For example, forecasting time series: The simplest methods for forecasting time series use only information about the predicted variable and do not try to detect factors that affect its behavior. The time series models used for forecasting include decomposition models, exponential smoothing models, and ARIMA models. When forecasting time series data, the goal is to evaluate how the sequence of observations will continue in the future.).
3. Some case studies. For example, one needs to develop a model for forecasting weekly passenger traffic on major domestic routes for one of Australia's leading airlines. The airline needs forecasts of the number of passengers for each main domestic route and

for each travel class (economy class, business class, and first class). The airline has provided weekly traffic data for the previous six years.

4. The main steps in a forecasting task (step 1: problem definition; step 2: collection of information; step 3: preliminary (research) analysis; step 4: choice and selection of models.) (Hyndman & Athanasopoulos, 2018).

The main tool for such research is mathematical models of the spread, development, treatment, and prevention of tuberculosis (Avilov & Romanyukha, 2007).

The World Health Organization (WHO) classifies Kazakhstan as a country with a high priority for tuberculosis (TB) and a high burden of multidrug-resistant tuberculosis (MDR-TB).

TB services work in every region of Kazakhstan (Figure 1). Regional Tuberculosis Dispensaries (RTD) are specialized diagnostic and treatment-and-prophylactic healthcare organizations coordinating the activities of TB services in regions, providing organizational, methodological, practical and advisory TB care to medical organizations and the population of regions.

RTDs perform the following functions:

- Organizational and methodological guidance related to the provision of TB care to the population.
- Advisory, practical, organizational and methodological assistance in TB control for medical organizations and medical services of the Ministry of Justice of the Republic of Kazakhstan, the Ministry of Defense of the Republic of Kazakhstan and the Ministry of Internal Affairs of the Republic of Kazakhstan, as well as educational institutions of the Ministry of Education and Science of the Republic of Kazakhstan.
- Diagnosis of tuberculosis based on standard studies (bacteriological ones such as microscopy, microbial culture and Drug Susceptibility Testing (DST)), clinical and radiographic ones and introduction of tuberculin Mantoux 2 TEsamples, as well as additional methods (instrumental and radiological ones).
- Chemotherapy of patients within standard and individual treatment regimens under the direct supervision of a medical professional.
- Treatment of drug-sensitive and drug-resistant tuberculosis patients.
- Isolation of patients with an infectious form of tuberculosis, those evading treatment, and with a chronic form of tuberculosis.

- Diagnosis and treatment of concomitant pathologies with the involvement of specialized professionals.
- Diagnosis and implementation of measures to reduce and timely eliminate adverse reactions to anti-TB drugs, etc.

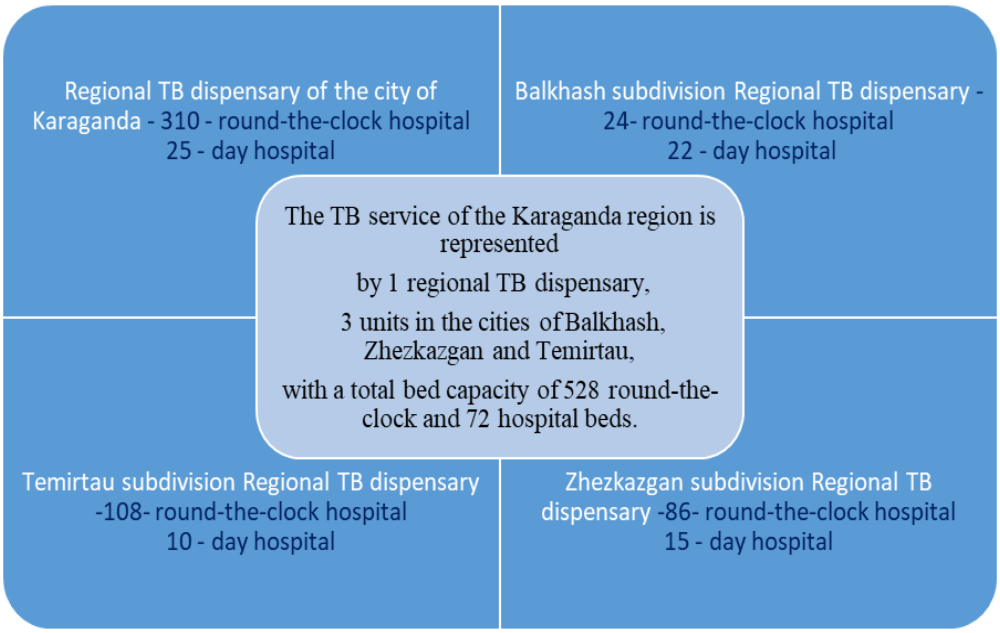

Figure 1 Structure of the TB service of the Karaganda region

The tuberculosis epidemic, as a rule, is accompanied by quantitative and qualitative changes in the nature of this disease in various regions. It is necessary to make a forecast of the spread of the epidemic in a particular region in order to develop an action plan to identify and treat patients in this region. The article considers the Karaganda region of the Republic of Kazakhstan. Mathematical modeling, namely the development of a specific mathematical model that describes the spread of infection, is one of the most effective methods for predicting the spread of the epidemic (Laporte, Barinas, Chang, & Libman, 1996).

The generally accepted epidemiological analysis of the spread and control of tuberculosis is based on the division of the population into a number of classes (groups). The main classes are as follows (Waaler, Geser, & Andersen, 1962):

1. Uninfected (susceptible to infection) individuals (S) are those in whom the causative agents of tuberculosis have not yet penetrated.
2. Latently infected individuals (L) are those in whom the causative agents of tuberculosis are present and are in equilibrium with the immune system. Such individuals are characterized by the absence of any external manifestations of the

disease or an increased risk of death. In some studies, latent infection is considered as the earliest stage of the disease and implies the need to identify and treat latent infected individuals.

3. Sufferers (T) are individuals with clinical manifestations of tuberculosis caused by extensive damage to tissues as a result of the activity of mycobacteria in their organisms. The border between the state of latent infection and the disease is not strictly defined and largely depends on the diagnostic methods used. Currently, patients with active tuberculosis are mainly individuals with radiologically proven tuberculous tissue changes (mainly for tuberculosis of the respiratory organs) or with detected bacterial excretion.

The underlying epidemiological processes linking these three groups are as follows:

1. Increase of people in the S group due to youth (or newborns).
2. Infection of individuals from the S group with mycobacteria excreted by patients from the T group. The specific risk of infection (per unit of time) is called the strength of the infection and is designated as λ. The form of the dependence of λ on the number of groups, time, and other parameters is determined by the chosen model of population mixing (usually λ is proportional to the number or share of infectious individuals in the population). It should be noted that the strength of the infection λ is determined by three classes of real processes such as pathogen excretion by infectious individuals (i.e., the mass character of pathogen excretion), the frequency and proximity of contacts between individuals in the population in question, and the susceptibility of uninfected individuals to infection (i.e., the probability with which the pathogen entering the body overcomes the primary immune response).
3. Development of the disease. In individuals from group L, the disease can develop both due to endogenous activation (the parameters of this process are the property of infected individuals) and as a result of exogenous superinfection (in this case, the process parameters depend both on the properties of the infected individuals and on the strength of the infection λ). Modeling the rapid development of the disease in people infected recently and for the first time with the use of only the S, L and T groups is usually carried out by dividing infected people into those whose disease develops rapidly (transition from S to T) and those who become infected (i.e. those who enter the L group).
4. Additional tuberculosis deaths affect only the T group.
5. Spontaneous self-healing or drug cure from active tuberculosis transfer individuals from the T group to the L group (although in some models, drug-cured individuals are transferred to the S group).

Results and Discussion

The processes under consideration are modeled with the use of nonlinear systems of ordinary differential equations whose coefficients characterize the features of a population and the development of the disease. In order to refine the model for a specific population, a qualitative assessment of the model parameters (or their combinations) is necessary (Kabanikhin, 2009). Mathematical modeling of dynamic processes is based on the use of the law of conservation of mass:

$$\{\dot{y} = F(y(t),\ t, q,\ u)\ y(t_0) = y_0, \qquad (1)$$

Where $y(t)$ is $mass$ of experimental data, then $\frac{y(t)}{V} = f(t)$ is function of experimental data (in epidemiology, it is the number of patients by year), t is time of drug application, q is vector of parameters characterizing the process under consideration (in epidemiology, it is mortality, influx of individuals, the rate of development of the disease, etc.), u is input data (in pharmacokinetics, it is the method of drug application, i.e., intravenously, intramuscularly, orally, etc.) (World Health Organization, 2009).

These parameters can be estimated (or sometimes uniquely determined) with the use of some additional information about biological processes (drug concentration, viruses, infected patients, etc.). The problem of determining biological parameters with the use of such additional information is called the inverse problem and is incorrect in the general case (Takuadina, 2019; Kabanikhin, 2008).

The Consolidated Action Plan to Prevent and Control M/XDR-TB in the WHO European Region 2011–2015 was designed to respond to the worrisome problem of M/XDR-TB and to deploy a comprehensive response for the prevention and control of M/XDR-TB for all 53 Member States of the European Region and partners including Kazakhstan (World Health Organization, 2009). Over the past 5 years, the incidence of TB has decreased at an average rate of 4.3% per year, which is the fastest rate of decline in the world. Despite this, in high-priority countries for TB, the number of new cases is almost 8 times higher than in other countries in the Region. In 2015, for every fourth TB patient, treatment did not lead to a successful result, which is one of the highest rates in the world (World Health Organization, 2017). The Consolidated Action Plan to fight TB in the Republic of Kazakhstan for 2014-2020 is a logical continuation of TB control projects implemented in Kazakhstan, and a continuation of the Consolidated Action Plan to Prevent and Control M/XDR-TB in the WHO European Region 2011–2015. The strategies included in this document comply with the guidelines laid down in the Salamatty Kazakhstan State Health

Care Development Program for 2011-2015. This document has been developed in accordance with the provisions of the Roadmap for Prevention and Control of M/XDR-TB developed and published by the WHO Regional Office for Europe in 2012. The country is determined to achieve the goal of halving the prevalence of tuberculosis over the next 7 years until 2020:

- The long-term strategic vision of this Consolidated Plan is to make Kazakhstan a country free of tuberculosis.
- The overall goal is to reduce TB morbidity and mortality.
- To achieve, by 2020, the following indicators:
  - Morbidity rate of 55/100,000 of the population (81.7/100,000 in 2012).
  - Mortality rate reduction to 5.8/100,000 people (8.0/100,000 in 2012).
  - Treatment coverage for 95% of patients with M/XDR-TB.

Also in 2015, the United Nations adopted the Sustainable Development Goals (SDGs) to implement until 2030. One of these goals is to eradicate the global TB epidemic. The WHO Strategy for the Elimination of Tuberculosis 2016–2035, endorsed by the World Health Assembly in 2014, calls for reducing the number of deaths from TB by 90% and the TB morbidity rate by 80% till 2030 compared to 2015.

The number of cases of tuberculosis for the period from 2003 to 2018 is shown in the dynamics by year (Figure 2), as well as mortality in the Karaganda region (Figure 3).

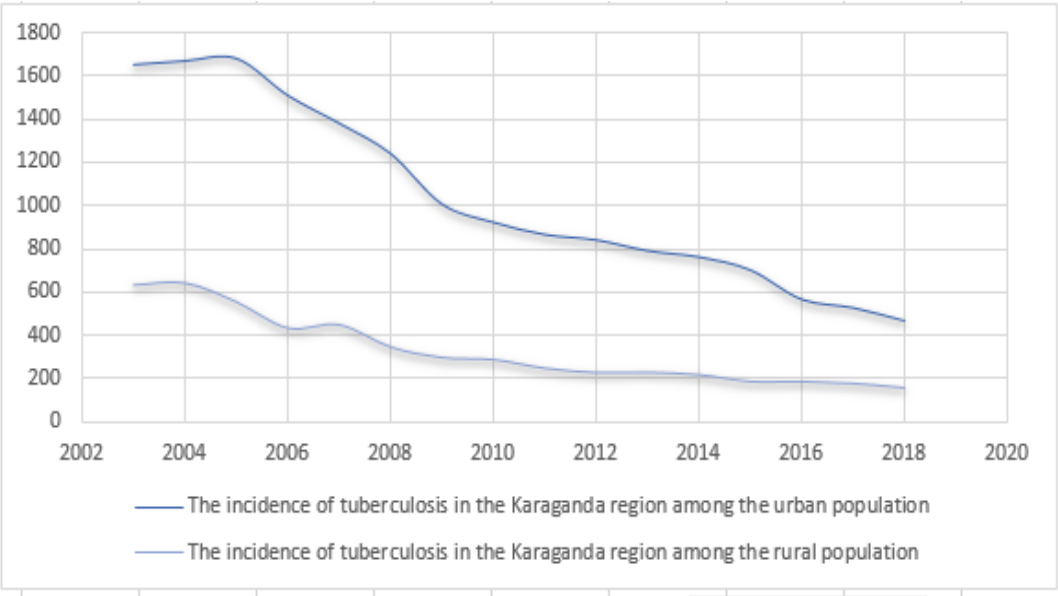

Figure 2 The incidence of tuberculosis in the Karaganda region among urban and rural populations

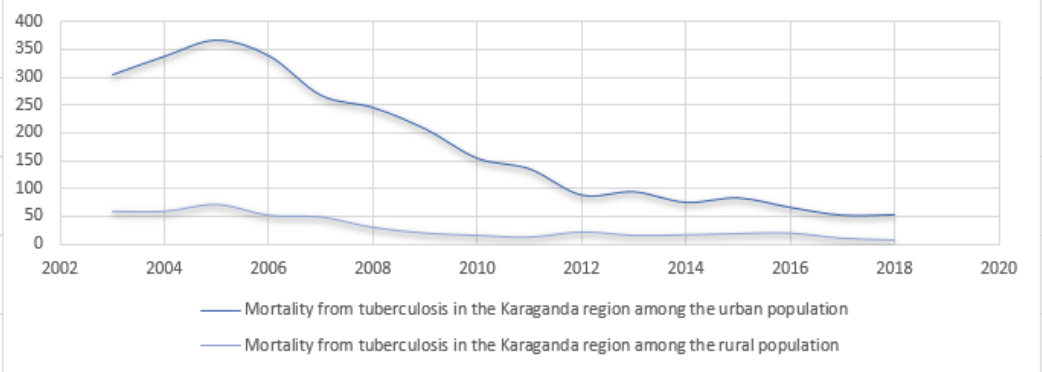

Figure 3 Mortality from tuberculosis in the Karaganda region among the urban and rural populations

The dynamics of the development of active tuberculosis (Figure 4), as well as mortality (Figure 5) among the children and adolescents of the Karaganda region from 2003 to 2018, are presented below:

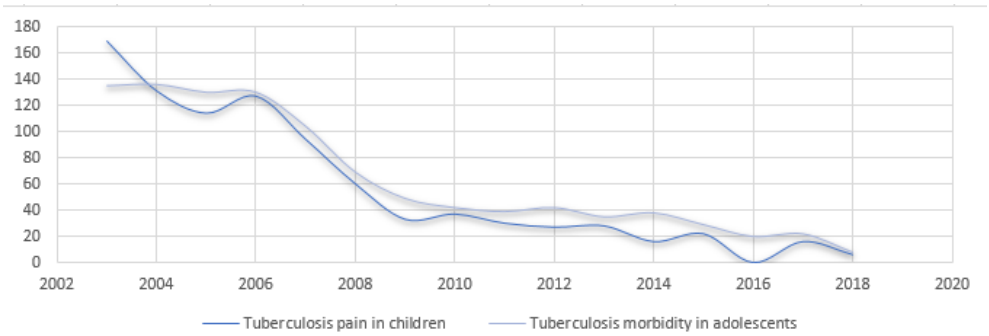

Figure 4 Tuberculosis morbidity in the Karaganda region among children and adolescents

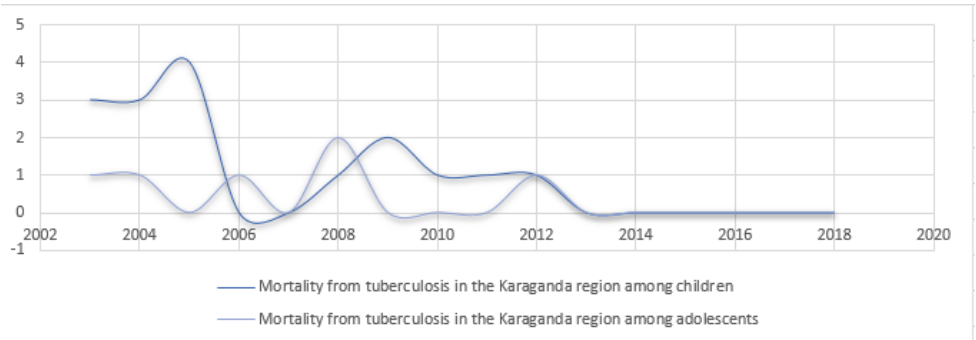

Figure 5 Mortality from tuberculosis in the Karaganda region among children and adolescents

Although there is a decrease in the incidence rate in the region as a whole, tuberculosis has acquired new qualities. Patients often have drug intolerance and drug resistance to the causative agent of tuberculosis (Figure 6).

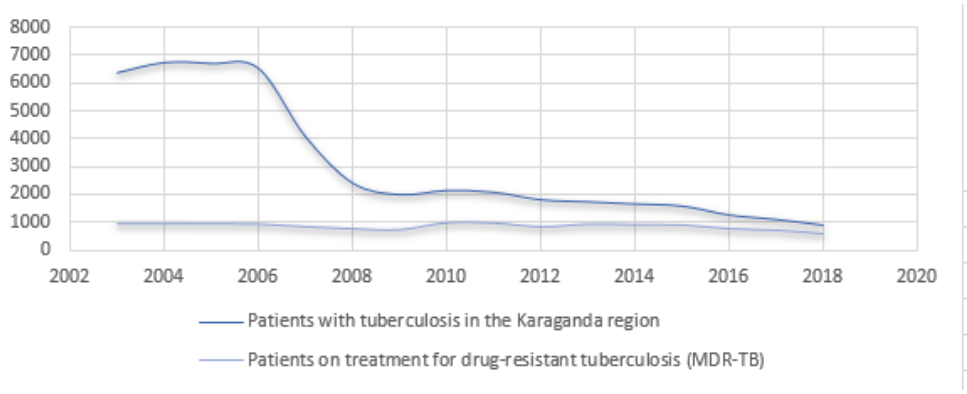

Figure 6: Patients with tuberculosis and patients on treatment for MDR-TB in the Karaganda region

Tuberculosis morbidity depends on many factors in the foci of infection including preventive measures, the timing of the detection of the disease, as well as the effectiveness of treatment of adult patients, which makes the task of early diagnosis one of the most pressing problems of modern phthisiology (Manfredi & D'Onofrio, 2013). The development of the methodology and organizational technologies for the integrated assessment of anti-TB activities as a basis for optimizing the management of TB services is also relevant (Temime, Hejblum, Setbon, & Valleron, 2008; Chubb & Jacobsen, 2010).

Below there are calculations, with the use of mathematical methods, of the correlation coefficient between the epidemiological indicators of tuberculosis morbidity from infection foci and the effectiveness of preventive measures that have been taken in Kazakhstan for last 16 years. The clarification of the correlation relationship between indicators will contribute to the development of a methodology and evaluation of the anti-TB activity of the TB service. Mathematical analysis of epidemiological indicators was performed with SPSS. SPSS is a statistical data processing program designed for applied research and is a universal system that supports the process of data analysis at any level and designed to implement the complete sequence of data analysis steps: from viewing data, creating tables and calculating descriptive statistics to complex statistical analysis (Nasledov, 2005). The analysis was performed in several stages such as data collection, selection of parameters, building a matrix, assessing the adequacy of the model and forecasting epidemiological indicators. In order to identify factors affecting the morbidity among contact people, the Spearman's correlation coefficient of the general epidemiological indicators for tuberculosis was calculated. Since the coefficient is a non-parametric analysis method, verification of the normal distribution is not required. For ordinal (rank) variables or variables whose distribution differs significantly from the normal one, the Spearman's correlation coefficient can be used. The calculation of Spearman's rank correlation coefficient includes the following steps:

1. To assign a serial number (rank) to each of the signs in ascending or descending order.
2. To determine the difference in ranks of each pair of matched values (*d*).
3. To square each difference and summarize the results.
4. To calculate the rank correlation coefficient with the following formula:

$$\rho = 1 - \frac{6 \cdot \sum d^2}{n(n^2 - 1)}$$

5. To determine the statistical significance of the coefficient with the use of the t-criterion calculated with the following formula:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

This coefficient can take values between -1 and +1, and if the value is closer to 1, then this means that there is a strong relationship, and if it is closer to 0, then it is a weak one. The relationship strength is also characterized by the absolute value of the correlation coefficient. The indicators of the tuberculosis morbidity from bacterial excretion foci and the indicators of the general morbidity among adults, as well as organizational measures carried out in the country to reduce the burden of tuberculosis, were chosen as samples when determining the correlation relationship. These indicators for statistical analysis were taken from the statistical overview on tuberculosis in the Karaganda region of the Republic of Kazakhstan for 16 years, from 2003 to 2018.

When calculating the Spearman's correlation coefficient between two variables (people with tuberculosis and people who died from all forms of tuberculosis), a strong direct relationship was revealed ($\rho = 0.9$) ($p < 0.01$). A relatively strong relationship was revealed between people with tuberculosis and people on MDR-TB treatment ($\rho = 0.6$) ($p < 0.01$). Therefore, the following conclusions can be drawn:

1. Although tuberculosis is a treatable disease, in most cases it is fatal;
2. As mentioned earlier, unfortunately, in Kazakhstan, the morbidity rate of MDR-TB is steadily increasing (Table 1).

Table 1 Morbidity Rate of MDR-TB in Kazakhstan

Correlations						
			People with tuberculosis (nr. of persons)	People on treatment for drug-resistant tuberculosis (MDR-TB)	People cured of tuberculosis (nr. of persons)	People who died from all forms of tuberculosis (nr. of persons)
ρ-Spearman	People with tuberculosis (nr. of persons)	Correlation coefficient	1,000	,638**	-,152	,979**
		Value (two-sided)	.	,008	,574	,000
		N	16	16	16	16
	People on treatment for drug-resistant tuberculosis (MDR-TB)	Correlation coefficient	,638**	1,000	-,052	,547*
		Value (two-sided)	,008	.	,849	,028
		N	16	16	16	16
	People cured of tuberculosis (nr. of persons)	Correlation coefficient	-,152	-,052	1,000	-,065
		Value (two-sided)	,574	,849	.	,811
		N	16	16	16	16
	People who died from all forms of tuberculosis (nr. of persons)	Correlation coefficient	,979**	,547*	-,065	1,000
		Value (two-sided)	,000	,028	,811	.
		N	16	16	16	16
**. The correlation is significant at the level of 0.01 (two-way).						
*. The correlation is significant at the level of 0.05 (two-way).						

1) Identifiability Analysis Framework

1. First, a priori identifiability has to be carried out (with neglect of data type and noise, i.e. measurements are carried out in ideal situations). If a problem is ultimately a priori identifiable, this does not mean that it is practically identifiable. But if it is a priori unidentifiable, then it is practically unidentifiable.

The definition of a priori identifiability is formulated as follows (Carson & Cobelli, 2008):

Definition 1. Model (1) is called a priori identifiable if its parameters $q = [q_1, \dots, q_s]$ can be uniquely determined from the input data u (t) and the measurement data y (t).

Methods for the a priori identifiability analysis of dynamical systems are as follows:

- Transfer function method (for linear models)
- Method based on Taylor series expansion (for linear and non-linear models).
- Method based on differential algebra theory.

- Graph-based method (for linear models). In the case of an unidentifiable model, this method finds a change in variables that convert an unidentifiable model into an identifiable one.
- Modification of the method based on differential algebra theory (for non-linear models). In the case of an unidentifiable model, this method finds a change in variables that convert an unidentifiable model into an identifiable one.
- A method based on series expansion with the use of Lie derivative (for non-linear models).
- Tabular identification (for non-linear models).
- A method based on implicit function theorem (for non-linear models).

The article examines and investigates one of the approaches to the analysis of a priori identifiability of a specific mathematical model of the spread of the tuberculosis epidemic (Miao, Xia, Perelson, & We, 2011):

$$\{\frac{dS}{dt} = \Pi - \mu S - p\lambda S - (1-p)\lambda S, \frac{dL}{dt} = p\lambda S - \mu L - \delta L, \frac{dT}{dt} = (1-p)\lambda S + \delta L - (\mu + \mu_T)T, S(0) = S_0, L(0) = L_0, T(0) = T_0. \quad (2)$$

In model (2), the entire population is divided into uninfected individuals (S), latently infected individuals (L), and infected sufferers (T). Model parameters are as follows: Π is the influx of youth into the model population; $\lambda = \beta T$, where β is transmissibility parameter; p is the share of recently infected, whose disease is developed within 1 year; δ is the constant of disease development rate; $1/\mu$ is average life expectancy; μ_T is the tuberculosis mortality. Therefore, this results in a set of parameters $(p, \beta, \delta, \mu_T) \in R^4$.

This method is based on the use of the expansion of a function in a Taylor series and differs from other numerous methods by the fact that it is applicable both to linear systems of differential equations and to non-linear ones.

The expansion of the function $y_i(t, q)$ in a Taylor series at the point $t_0 = 0$ is as follows:

$$y_i(t) = y_i(t_0) + t\dot{y}_i(t_0) + \frac{t^2}{2!}\ddot{y}_i(t_0) + \frac{t^3}{3!}\dddot{y}_i(t_0)$$

It will be denoted as follows:

$$y_i^{(k)}(t_0, q) = \varphi_k^i; k = 1, 2 \ldots (3)$$

This method is based on replacing the left-hand sides in the equation (3) with expressions obtained from the original system of ordinary differential equations (2). After such a replacement, a system of non-linear algebraic equations is obtained. The question of identifiability of a mathematical model is reduced to an analysis of the solvability of an algebraic system (Takuadina, 2019). After a series of calculations, the conclusion was made that the system had a unique solution, which means that the model in question is identifiable.

Then, a researcher chooses the path of analysis: either sensitivity analysis method, or practical identifiability is used.

2. It is necessary to continue the identifiability analysis and when knowing the sensitivity, to choose an algorithm that will work well for sure. Sensitivity analysis is used to assess the identifiability of unknown q parameters of the model for ODE (ordinary differential equation) system:

$$\{\dot{x}(t) = f(t, x(t), q), t \in (0, T);\ y(t) = h(x(t), q), x(0) = x_0,$$

Where $x(t) \in R^M$ denotes the vector of functions (M denotes the number of equations), $q \in R^L$ denotes vector of parameters (L denotes number of parameters), $y(t) \in R^P$ denotes a measurement function ($P \leq M$ denotes the number of measured equations), $t$denotes time. Sensitivity analysis methods do not require actual experimental data, although they may require the number of measurements and the times when these measurements were taken (Carson, & Cobelli, 2008; Cacuci, 2003). In order to study the model with the use of sensitivity analysis methods, it is necessary to know the values of parameters that can be known from the literature or from statistical information.

Sensitivity analysis methods are based on the study of the sensitivity matrix. In order to construct it, let us assume that $t_1 \leq t_2 \leq \cdots \leq t_K$ are moments of time at which the vector function will be measured $y(t)$. Then the coefficients of the sensitivity matrix for a given vector of q^* parameters are calculated with the following formula

$$s_{ij} = \frac{\partial y_i(t_k, q^*)}{\partial q_j},$$

Wherey_i $(i = 1, \ldots . P)$ denotes i-component of the vector of measurement equations and q_j $(j = 1, \ldots, L)$ denotes j-component of the vector of parameters.

Therefore, the sensitivity matrix is defined as follows:

$$S_{P\cdot K\times L} = (s_{11}(t_1) \cdots s_{1L}(t_1) \vdots \ddots \vdots s_{P1}(t_1) \cdots s_{PL}(t_1) \vdots \vdots \vdots s_{11}(t_K) \cdots s_{1L}(t_K) \vdots \ddots \vdots s_{P1}(t_K) \cdots s_{PL}(t_K)).$$

There are sensitivity analysis methods such as orthogonal method and eigenvalue method (Saltelli, Chan, & Scott, 2000).

3. An analysis of practical identifiability enables estimating model parameters with acceptable accuracy from noisy experimental data. The problem is reduced to finding the functional:

$$J(q) = \min_q \sum_{k=1}^{K} |y_1(t_k; q) - V{f_1}^k|^2$$

The bottom line is that it is necessary to set the coefficients randomly, then, if the solutions obtained are far from measurements, it is necessary to minimize and approximate them in a quadratic sense. Some methods that solve the minimization problem are used for this.

The most effective forecasting methods (methods for solving optimization problems) include stochastic methods (genetic algorithms, Monte Carlo method, annealing method, etc.) and deterministic methods **(**Levenberg-Marquardt method, Landweber iteration method, conjugate gradient method, etc.).

- Deterministic methods. They are represented by gradient methods developed in the works of S.I. Kabanikhin, A.L. Karchevsky, A.F. Latypov, Y.V. Nikulichev, etc. The gradient descent method was used to solve the problem of optimization design of a turbine blade. The problem of determining the structure of the borehole area according to the results of induction logging was solved by M.I. Epov, I.N. Yeltsov, and others as optimization with the use of the deterministic flexible polyhedron method (Nelder-Mead method). Classical gradient descent methods are suitable for solving problems with continuous functionals and a small number of variable parameters. Nelder-Mead methods do not reliably look for global extrema of functionals. Gradient type methods are relatively simple to implement and do not require large computational costs. Nevertheless, they are used only for single-purpose optimization problems, have differentiability requirements for the objective function, require a good initial approximation, and work unstably on a large number of parameters (that is, they do not guarantee to find the global minimum of the problem).

  The main idea of the methods is to go in the direction of the steepest descent, and this direction is set by the anti-gradient - ∇F:

$$\vec{x}^{[j+1]} = \vec{x}^{[j]} - \lambda^{[j]} \nabla F(\vec{x}^{[j]})$$

Where $\lambda^{[j]}$ is chosen as

- A constant, then the method may diverge.
- A fractional step, that is, the step length in the descent process is divided by a certain number.
- The steepest descent: $\lambda^{[j]} = argmin_{\lambda} F(\vec{x}^{[j]}))$.

Then the following is chosen: $v_i^{[j]} = -\frac{\partial F}{\partial x_i}$, where all derivatives are calculated with $x_i = x_i^{[j]}$, and step length is reduced $\lambda^{[j]}$as the minimum of the function F is approached. For analytic functions F and small values f_i, Taylor series expansion $F(\lambda^{[j]})$ enables choosing the optimal step size.

$$\lambda^{[j]} = \frac{\sum_{k=1}^{n} \left(\frac{\partial F}{\partial x_k}\right)^2}{\sum_{k=1}^{n} \sum_{h=1}^{n} \frac{\partial^2 F}{\partial x_k dx_h} \frac{\partial F}{\partial x_k} \frac{\partial F}{\partial x_h}},$$

Where all derivatives are calculated with $x_i = x_i^{[j]}$. A parabolic interpolation of the function $F(\lambda^{[j]})$may be more convenient.

The algorithm is as follows:

1. Initial approximation and calculation accuracy $\vec{x}^0, \epsilon$ are set;
2. The following is calculated:

$$\vec{x}^{[j+1]} = \vec{x}^{[j]} - \lambda^{[j]} \nabla F(\vec{x}^{[j]}), \text{ где } \lambda^{[j]} = argmin_{\lambda} F(\vec{x}^{[j]} - \lambda^{[j]} \nabla F(\vec{x}^{[j]})).$$

3. The stop condition is checked:

- If $\left|\vec{x}^{[j+1]} - \vec{x}^{[j]}\right| > \epsilon$, then $jjjj = j + 1$ and go to step 2.
- Else $\vec{x} = \vec{x}^{[j+1]}$and stop (Conte, & de Boor, 1981).

- Lately, stochastic methods have become more widespread than deterministic ones since they do not require the smoothness of target functionals, they have the ability not to stop at local extrema, and are used to solve multicriteria problems. Among stochastic methods, the most appropriate method for optimization design is an evolutionary genetic algorithm. Genetic algorithms, for example, are so popular because they are very imaginative and therefore easy to understand. In fact, it is not

difficult and extremely interesting to present a solution to a problem, as a real biological process of developing a population of living creatures with certain properties. Such a process searches for the optimal solution, while simultaneously analyzing the many current sets of parameters, called individuals that have evolved over many generations in accordance with the prescribed rules. The need for a large number of calculations of the target functionals is compensated by the possibility of parallel calculation. The genetic algorithm has been successfully used to solve the problem of optimization design (Vajda, 1989). Stochastic methods suggest the probabilistic nature of the algorithm used. Simulated annealing algorithm and other algorithms based on the Monte Carlo method do not require the smoothness of the functional and depend insignificantly on the initial approximation. However, most often they are used in single-purpose discrete optimization. The use of the simulated annealing method in continuous optimization problems was carried out by T. Rogalsky and N. Kim. The annealing algorithm also has its own prototype in the real world (this is understandable from the name itself). It came from physics. The process imitated by this algorithm is similar to the formation of a crystalline structure of a substance with minimum energy during cooling and solidification, when particles move randomly at high temperatures, gradually slow down and freeze in places with the lowest energy. In the case of a mathematical problem, the role of particles of matter is performed by parameters, and the role of energy is a function that characterizes the optimality of the set of these parameters. When modeling the process of metal annealing, one searches for such a point or a set of points at which a minimum of some numerical function $F(\underline{x})$ is reached where $\underline{x}=(x_1, \dots, x_m)\in X$. The solution is reached by sequential calculation of the points. $\underline{x_0}, \underline{x_1}, \dots$, are spaces X; every point starting from $\underline{x_1}$, "claims" to bring the solution better than the previous ones. An algorithm accepts a point $\underline{x_0}$ as initial data. At each step, the algorithm calculates a new point and lowers the value understood as "temperature." The algorithm stops when it reaches a zero temperature point. The point $\underline{x_{i+1}}$, according to the algorithm, is based on the current point $\underline{x_i}$ as follows. The operator A is used for the point $\underline{x_i}$. This operator randomly modifies the associated point that results in a new point $\underline{x^*}$. The point $\underline{x^*}$ becomes the point $\underline{x_{i+1}}$ with the probability $P(\underline{x^*}, \underline{x_{i+1}})$ that is calculated according to the Gibbs distribution:

$$P\left(\underline{x_i}\right) = \{1, F\left(\underline{x^*}\right) - F(\underline{x_i}) < 0 \;\; exp\; exp\left(-\frac{F\left(\underline{x^*}\right) - F(\underline{x_i})}{Q_i}\right), F\left(\underline{x^*}\right) - F(\underline{x_i}) \geq 0$$

Here $Q_i > 0$ are the elements of an arbitrary decreasing function converging to a zero positive sequence that is the analog of the falling temperature in a crystal. The rate of

decrease and the law of decrease can be set at the request of the creator of the algorithm. Simulated annealing algorithm is similar to gradient descent but, due to the randomness of the choice of an intermediate point, it should fall into local minima less often than gradient descent. Simulated annealing algorithm does not guarantee that the minimum of the function is found, however, with the correct policy for generating a random point in the space X, usually, an improvement in the initial approximation occurs.

In addition, stochastic methods include swarm intelligence, ant colony algorithm, etc. These methods have undeniable advantages such as the ability to use them to solve multicriteria problems, the lack of requirements for the smoothness of target functionals, and the ability not to stop at local extrema.

2) Analysis of the Mathematical Model of the Spread of the Epidemic of Tuberculosis Using the Example of the City of Karaganda

For analysis, settlements in the Karaganda region (the urban population is indicated in orange and the rural population is indicated in blue) where the incidence is highest are chosen. In the figures below (Figure 7, Figure 8, and Figure 9), over the past three years, the city of Karaganda is also represented.

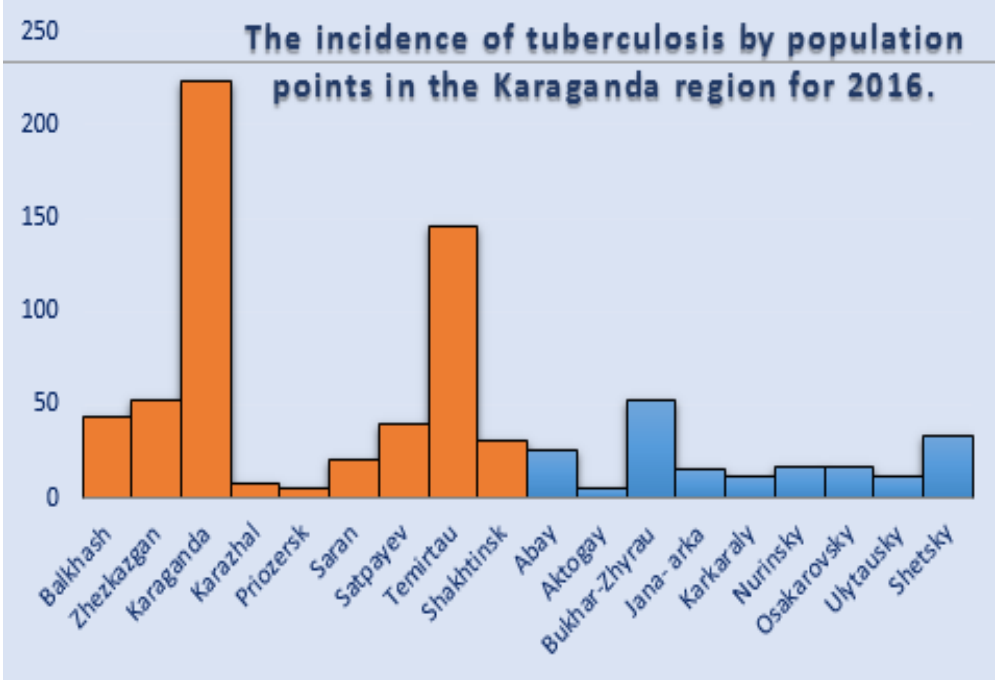

Figure 7 Incidence of Tuberculosis for 2016

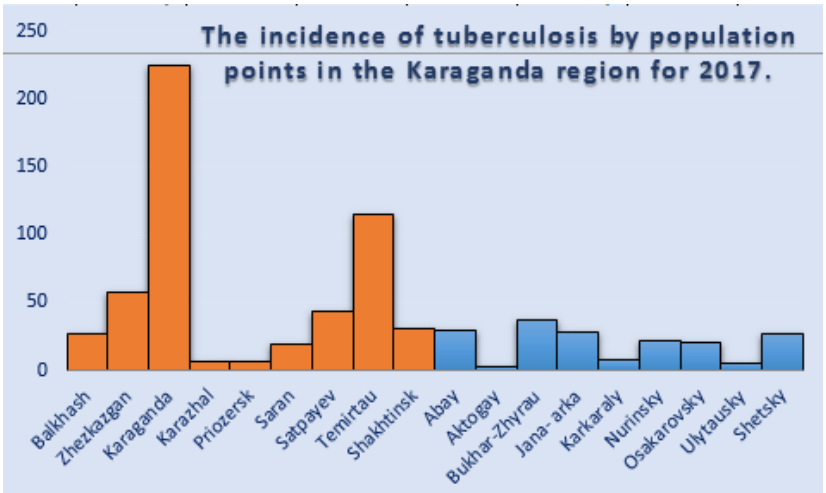

Figure 8 Incidence of tuberculosis for 2017

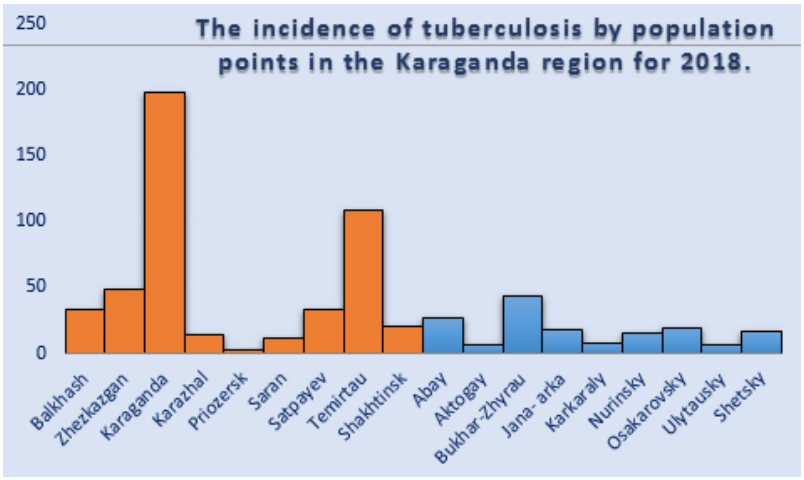

Figure 9 Incidence of tuberculosis for 2018

A prerequisite for the model providing reliable predictions is that the assumptions, on which it is based, correspond to reality. However, this correspondence is always limited since all models are simplifications of reality. An important advantage of mathematical modeling is that it requires transparency and accuracy in relation to assumptions that enables testing the awareness of the epidemiology of the disease by comparing the results of the model and the observed patterns. Models can also help in making decisions by making predictions about important issues such as interventional changes in the spread of the disease.

A mathematical model of the spread of an infectious disease in a population describes the transmission of a pathogen depending on the nature of contacts among infectious and susceptible people, the latent period from infection to infection, the duration of infectivity, the degree of acquired immunity after infection, etc. After all these factors are formulated

in the model, we can make predictions about the number of people who are expected to be infected during the epidemic, the duration of the epidemic, and the peak incidence. Indeed, we can predict the entire epidemic curve providing us with the expected number of cases at any moment (Kabanikhin, Krivorotko, Ermolenko, Kashtanova, & Latyshenko, 2017).

It is necessary to make a forecast of the spread of the epidemic in a specific region in order to create an action plan for the identification and treatment of patients in this region. Mathematical modeling, namely the development of a specific mathematical model that describes the spread of infection among the population, is one of the most effective methods for predicting the spread of the epidemic.

Below the mathematical model of tuberculosis transmission in highly endemic regions of the Asia-Pacific region are considered. This model was proposed by some scientists (Huppert & Katriel, 2013; Trauer, Denholm, & McBryde, 2014) but with some correction:

$$\{\frac{dS}{dt} = l\pi N + \varphi T + \varphi_m T_m - (\lambda_d + \lambda_{dm} + \mu)S, \frac{dL_A}{dt} = \lambda_d(S + L_B + L_{Bm}) - (\varepsilon + k + \mu)L_A, \frac{dL_{Am}}{dt} = \lambda_{dm}(S + L_B + L_{Bm}) - (\varepsilon + k + \mu)L_{Am}, \frac{dL_B}{dt} = kL_A + \gamma I - (\lambda_d + \lambda_{dm} + \nu + \mu)L_B, \frac{dL_{Bm}}{dt} = kL_{Am} + \gamma I_m - (\lambda_d + \lambda_{dm} + \nu + \mu)L_{Bm}, \frac{dI}{dt} = \varepsilon L_A + \nu L_B + (1 - \eta)\omega T - (\gamma + \delta + \mu_i)I, \frac{dI_m}{dt} = \varepsilon L_{Am} + \nu L_{Bm} + \eta\omega T - (\gamma + \delta_m + \mu_i)I_m, \frac{dT}{dt} = \delta I - (\varphi + \omega + \mu_t)T, 0\ \frac{dT_m}{dt} = \delta_m I_m - (\varphi_m + \omega + \mu_t)T_m, S(0) = S_0, L_A(0) = L_{A_0}, L_{Am}(0) = L_{Am_0}, L_B(0) = L_{B_0}, L_{Bm}(0) = L_{Bm_0}, I(0) = I_0, I_m(0) = I_{m_0}, T(0) = T_0, T_m(0) = T_{m_0}. (4)$$

Here,

$$\lambda_d = \chi\beta\rho(I + oT)/N, \lambda_{dm} = \chi\beta_m\rho(I_m + oT_m)/N.$$

In a Cauchy problem (4) the population is divided into vaccinated sensitive individuals (usually children under 14) (S), latent infected individuals, individuals with MDR-TB strain (index m), with a fast (L_A, L_{Am}) and slow (L_B, L_{Bm}) development of an active form of the disease, infected patients, individuals on treatment (T, T_m), and those untreated (I, I_m). $N = S + L_A + L_{Am} + L_B + L_{Bm} + I + I_m + T + T_m$ denotes the whole population.

The notation used to describe mathematical models is standardized as far as possible. Variable models (group numbers) are denoted in capital Latin letters (S, T, etc.), the lowering indices indicate the physical properties of the variable, and the raising indices indicate the impact and external factors. The rate constants of various transitions are denoted by lowercase Greek letters, whilst the probabilities and proportions are indicated by lowercase Latin letters (Table 2).

Table 2 Constants and coefficients

Symbol	Description	Unit of measurement	Value
Π	Influx of youth into the model population	person/year	It depends on the type of population
N	Total population	person	It depends on the type of population
μ	Tuberculosis mortality	person/year	0.016
φ	TB treatment rate	person/year	2
φ_m	MDR-TB treatment rate	person/year	0.5
$\boldsymbol{\varepsilon}$	Rate of early progression of the disease	year	0.129
k	Rate of transition to late disease progression	year	0.821
γ	Spontaneous self-healing frequency	year	0.63
$\boldsymbol{v}$	Rate of development of the active form of the disease with endogenous activation	year	0.075
η	likelihood of developing MDR-TB in treatment	-	0.035
ω	Re-infection rate	person/year	0.25
$\boldsymbol{\delta}$	Frequency of detection of persons with active tuberculosis	person/year	0.72
$\boldsymbol{\delta_m}$	Frequency of detection of individuals with active MDR-TB	person/year	0.035
μ_i	Tuberculosis mortality without treatment	year	0.37
μ_t	Tuberculosis mortality during treatment	year	0.5 μ_i
β	Contagiousness parameter	-	It depends on the type of population
β_m	Organism index with MDR-TB	-	0.7 β
$\boldsymbol{\chi}$	Partial immunity parameter	-	0.49
$\boldsymbol{\rho}$	Infection fraction	-	0.7
$\boldsymbol{o}$	Modification of infectivity treatment	-	0.6
l	BCG vaccination rate	-	0.65
π	Birth rate with sensitivity analysis	person/year	0.025

The model was also applied to the population of the Siberian Federal District (Trauer, Denholm, & McBryde, 2014). Since BCG (Bacillus Calmette–Guérin) vaccination of infants in Russia, as well as in Kazakhstan, is compulsory for citizens whose incidence exceeds 80 cases per 100,000 people, the majority of the population was vaccinated at preschool age. Therefore, in the model (2) there is no division of the population into

sensitive unvaccinated (S_A) and vaccinated persons (S_B) (Takuadina & Imangaliyev, 2019).

The vector of exact parameters is selected $\Theta:=(\varepsilon, k, \nu, \delta, \delta_m, \rho\chi, o)$ on the basis of Table 1. When solving the direct Cauchy problem (4) with parameters Θ with the use of the Runge–Kutta fourth-order approximation method, synthetic data $\Omega^k:=(L_{A_k}, L_{Am_k}, L_{B_k}, L_{Bm_k}, I_k, I_{m_k})$ are obtained and are evenly distributed over the segment (0, T). In order to get as close as possible to real conditions, suppose that the data are known only for functions $S(t), T(t), T_m(t)$ since collecting accurate statistics on the number of latently infected people in real life is almost impossible.

3) Numerical Solution of a Direct Problem for the City of Karaganda

System parameters: number of equations - 9, time interval - 30 years. The spaced point in time is 5 years. The initial data for 2004are entered into the program. The forecast is until 2034 (Figure 10).

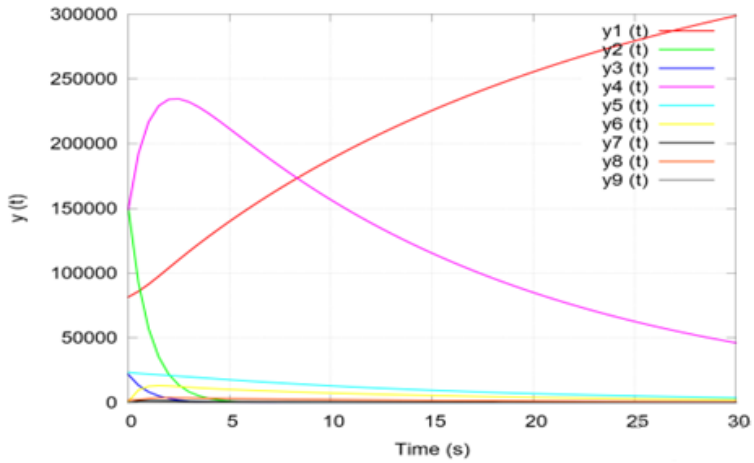

Figure 10 Forecast until 2034

Each curve corresponds to an equation in the mathematical model:

y1(t)=$S(t)$ denotes vaccinated sensitive people (81502 persons).

y2(t)=$L_A(t)$ denotes latently infected people with the fast development of an active form of the disease (150116 persons).

y3(t)=$L_{Am}(t)$ denotes latently infected people with MDR-TB strain with the fast development of an active form of the disease (21681 people).

y4(t)=$L_B(t)$ denotes latently infected people with a slow development of an active form of the disease (149303 persons).

y5(t)=$L_{Bm}(t)$ denotes latently infected people with MDR-TB strain with the slow development of an active form of the disease (23302 persons).
y6(t)=$I(t)$ denotes untreated infected people (757 persons).
y7(t)=$I_m(t)$ denotes untreated infected people with MDR-TB strain (89 persons),
y8(t)=$T(t)$ denotes infected people on treatment (1894 persons).
y9(t)=$T_m(t)$ denotes infected people with MDR-TB strain on treatment (223 persons).

Conclusion

The World Health Organization classifies Kazakhstan as a country with high priority for tuberculosis. The article shows that this problem exists. The problem of tuberculosis with multiple and extensive drug resistance is very important. The solution to the problem of preventing an outbreak of tuberculosis in Kazakhstan is to use a forecast system for an outbreak of tuberculosis. Intelligent forecasting system should use forecasting methods based on mathematical models.

The stages of building a system for forecasting tuberculosis are as follows: determining the subject of forecasting, researching the field of forecasting, collecting forecast data, analyzing existing mathematical models, choosing the right models for building a forecasting system, evaluating and refining forecasting methods over time. Mathematical modeling, namely the development of a specific mathematical model that describes the spread of infection among the population, is one of the most effective methods for predicting the spread of the epidemic.

In order to build a forecasting system, it is necessary to analyze mathematical forecasting methods, perform mathematical analysis of epidemiological indicators with the use of the statistical program SPSS. It is also necessary when constructing a system for forecasting the epidemic of tuberculosis, to describe the factors affecting the incidence rate and calculate the relationship between epidemiological indicators. Subsequently, the construction of a mathematical model of epidemiological processes is carried outon the basis of collected statistics on the incidence and population in the selected territory (Karaganda region, Republic of Kazakhstan).

References

Arauz, W.M.S., Cedeño, G.I., Chávez, S.S., Pérez, A.V., & Gámez, M.R. (2017). Microgrid with a 3.4 kWp photovoltaic system in the Universidad Técnica de Manabí. *International Journal of Physical Sciences and Engineering*, *1*(2), 11-20. https://doi.org/10.21744/ijpse.v1i2.34

Avilov, K.K., & Romanyukha, A.A. (2007) Mathematical models of the spread and control over tuberculosis (review), *Mathematical Biology & Bioinformatics*, *2*(2), 188–318.

Cacuci, D.G. (2003). *Sensitivity and Uncertainty Analysis: Theory*, Chapman Hall/CRC, Boca Raton, FL.

Carson, E., & Cobelli, C. (2008). *Introduction to Modelling in Physiology and Medicine*, Academic Press, San Diego.

Chubb, M., & Jacobsen, K. (2010). Mathematical Modeling and the Epidemiological Research Process. *European Journal of Epidemiology*, *25*(1), 13-19.

Conte, S.D., & De Boor, C. (1981). *Elementary Numerical Analysis: An Algorithmic Approach*, 3rd ed. McGraw-Hill.
https://www-uxsup.csx.cam.ac.uk/courses/moved.Arithmetic/na2.pdf

Huppert, A., & Katriel, G. (2013). Mathematical modelling and prediction in infectious disease epidemiology. *Clinical Microbiology and Infection*, *19*(11), 999-1005.

Hyndman, R.J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice*.
https://otexts.com/fpp2/

Kabanikhin, S.I. (2008). Definitions and examples of inverse and ill-posed problems. *Journal of Inverse and Ill-posed Problems, 16*, 317-357.

Kabanikhin, S.I. (2009). *Inverse and ill-posed problems*. Siberian science publishing, Novosibirsk.

Kabanikhin, S.I., Krivorotko, O.I., Ermolenko, D.V., Kashtanova, V.N., & Latyshenko, V.A. (2017). Inverse problems of immunology and epidemiology. *Eurasian Journal of Mathematical and Computer Applications*, *5*(2), 14-35.

Laporte, R.E., Barinas, E., Chang, Y.F., & Libman, I. (1996). Global epidemiology and public health in the 21st century applications of new technology. *Annals of epidemiology*, *6*(2), 162-167.

Manfredi, P., & D'Onofrio, A. (2013). *Modeling the Interplay between Human Behavior and the Spread of Infectious Diseases*. Springer Science & Business Media.

Miao, H., Xia, X., Perelson, A.S., & Wu, H. (2011). On identifiability of nonlinear ODE models and applications in viral dynamics. *SIAM review*, *53*(1), 3-39.

Myers, M.F., Rogers, D.J., Cox, J., Flahault, A., & Hay, S.I. (2000). Forecasting disease risk for increased epidemic preparedness in public health. *Advances in parasitology*, *47*, 309-330.

Nasledov, A.D. (2005). SPSS: computer data analysis in psychology and social sciences, Piter, St. Petersburg.

Rinartha, K., & Suryasa, W. (2017). Comparative study for better result on query suggestion of article searching with MySQL pattern matching and Jaccard similarity. *In 5th International Conference on Cyber and IT Service Management (CITSM)*, 1-4.

Saltelli, A., Chan, K., & Scott, M. (2000). *Sensitivity analysis, Wiley Series in Probability and Statistics*. John Wiley and Sons, New York.

Sumba, E.F.S., Sumba, A.V.S., Castillo, G.A.L., & Rodríguez, J.A.P. (2020). Impact of distributed generation in the electrical system of Ecuador. *International Journal of Physical Sciences and Engineering*, *4*(1), 1-10.
https://doi.org/10.29332/ijpse.v4n1.389

Supardi, I.W., Suarbawa, K.N., Wibawa, I.M.S., Padmika, M., Mahayani, N.W.D., & Saputra, I.W.W. (2018). Utilizing turbine ventilator as supplier of electrical energy resources for home information. *International Journal of Physical Sciences and Engineering*, *2*(2), 1-10. https://doi.org/10.29332/ijpse.v2n2.127

Takuadina, A.I. (2019). A priori identifiability of a mathematical model describing the epidemiology of tuberculosis' in Prospects for the Development of Modern Science. *Proceedings of the IV International Scientific-Practical Conference, Regional Academy of Management*, Gwangju, Korea.

Takuadina, A.I., & Imangaliyev, S.I. (2019). Determination of a mathematical model for the spread of the tuberculosis epidemic for Kazakhstan. *Bulletin of Almaty University of Energy and Communications*, *1*(44), 32-37.

Temime, L., Hejblum, G., Setbon, M., & Valleron, A.J. (2008). The rising impact of mathematical modelling in epidemiology: antibiotic resistance research as a case study. *Epidemiology & Infection*, *136*(3), 289-298.

Trauer, J.M., Denholm, J.T., & McBryde, E.S. (2014). Construction of a mathematical model for tuberculosis transmission in highly endemic regions of the Asia-Pacific. *Journal of theoretical biology*, *358*, 74-84.

Vajda, S. (1989). Identifiability of first order reaction systems. *Reaction Kinetics and Catalysis Letters*, *11*(1), 39-43.

Waaler, H.T., Geser, A., & Andersen, S. (1962). The use of mathematical models in the study of the epidemiology of tuberculosis. *American Journal of Public Health*, *52*(6), 1002–1013.

Widana, I.K., Sumetri, N.W., Sutapa, I.K., & Suryasa W. (2020). Anthropometric measures for better cardiovascular and musculoskeletal health. *Computer Applications in Engineering Education*, 1–12.

World Health Organization. Regional Office for Europe. (2009). *Roadmap for the prevention and control of drug-resistant tuberculosis*, Copenhagen.

World Health Organization. Regional Office for Europe. (2017). *Bulletin of the World Health Organization.*

Lytvyn, V., Vysotska, V., Osypov, M., Slyusarchuk, O., & Slyusarchuk, Y. (2019). Development of intellectual system for data de-duplication and distribution in cloud storage. *Webology*, *16*(2), 1-42.