Nikolay Pertsev & Boris Pichugin

Stochastic individual-based model of spreading of tuberculosis

Omsk Branch of the Sobolev Institute of Mathematics SB RAS

Contents

Supported by RFBR, grant N 06-01-00127

1. Statement of the problem.

In several papers

1. Perelman M.I., Marchuk G.I., Borisov S.E., Romanyukha A.A. et. al. Tuberculosis epidemiology in Russia: the mathematical model and data analysis // Russ. J. Numer. Anal. Math. Modelling, 2004, vol.19, N.4, pp.305–314.

2. Melnichenko A.O., Romanyukha A.A. A model of tuberculosis epidemiology: estimation of parameters and analysis of factors influencing the dynamics of an epidemic process // Russ. J. Numer. Anal. Math. Modelling, 2008, Vol.23, N.1, pp.1–13.

3. **...**

vas constructed and investigated the mathematical model of spreading of tuberculosis in Russia in the form of nonlinear system of differential equations:

$$dS/dt = -\beta S(B + kB_0) - \mu S + f_S(t),$$

$$dL/dt = (1-p)\beta S(B + kB_0) + \beta_L D + \beta_{L_0} D_0 -$$
$$-L(\gamma + \alpha(B + kB_0) + \mu) + f_L(t),$$

$$dD/dt = p\beta S(B + kB_0) + L(\gamma + \alpha(B + kB_0)) +$$
$$+\beta_D B - \beta_B D - \beta_L D - \varphi_D D - \mu_D D + f_D(t),$$

$$dB/dt = \beta_B D - \beta_D B - \varphi_B B - \mu_B B + f_B(t),$$

$$dD_0/dt = \varphi_D D + \beta_{D_0} B_0 - \beta_{B_0} D_0 - \beta_{L_0} D_0 - \mu_{D_0} D_0 + f_{D_0}(t),$$

$$dB_0/dt = \varphi_B B + \beta_{B_0} D_0 - \beta_{D_0} B_0 - \mu_{B_0} B_0 + f_{B_0}(t)$$

$$+ \text{ initial data.}$$

Here:

- S – the size of the cohort of vaccinated susceptible (uninfected) individuals;
- L – the size of the cohort of latently infected individuals without clinical disease manifestations and patients with primary tuberculosis progressing without clear symptoms;
- D – the size of the cohort of unnotified SSN individuals;
- D_0 – the size of the cohort of notified SSN patients;
- B – the size of the cohort of unnotified SSP individuals;
- B_0 – the size of the cohort of notified SSP patients.

SSP — sputum smear positive,

SSN — sputum smear negative.

In order to give more detailed mathematical description of tuberculosis (TB) we use several characteristics of individuals, such as:

- age or lifespan,
- duration of the period of fast progression,
- durations of various stages of TB, etc.

Distribution lows of these characteristics may be very complicated and differs from exponential lows, which are used in original differential model.

The purpose of our research is:
1) construction the individual-based stochastic model, using basic assumptions of deterministic model of spreading of tuberculosis in Russia;
2) analysis of the results of computer simulation for several sets of the model parameters,
3) attempt to explain the differences between sizes of L, D, D_0, B, B_0 cohorts in several regions of Russia.

2. Brief formulation of the model.

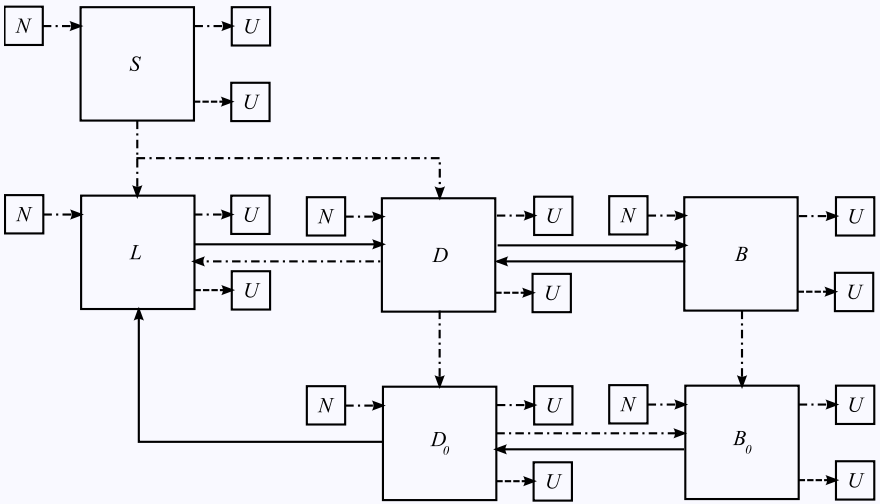
Let X — universal numerable set of individuals, which lived earlier, live present time t and will be in future in some region. We suppose, that for any time $t \geqslant 0$ the set X consists of 8 nonintersecting subsets (*cohorts*)

$$X = U(t) \sqcup S(t) \sqcup L(t) \sqcup D(t) \sqcup B(t) \sqcup D_0(t) \sqcup B_0(t) \sqcup N(t),$$

where $S(t)$, $L(t)$, $D(t)$, $B(t)$, $D_0(t)$, $B_0(t)$ — the above mentioned individuals aged > 16 years;
$U(t)$ — individuals aged > 16 years, don't live in region present time t, but have lived until t;
$N(t)$ — individuals aged ≤ 16 years and individuals, which don't live in region present time t, but will immigrate to region in future.

We use the modification of the basic scheme of transitions of individuals from one cohort to another. These transitions are denoted by various arrows and shows:
1) contacts of individuals,
2) progression of illness and remission of SSP individuals and patients,
3) detection of SSN and SSP individuals,
4) spontaneous recovery and treatment of SSN individuals and patients,
5) natural death and mortality caused by TB,
7) emigration and immigration.

Fig.1. The scheme of transitions of individuals.



Transitions:

unbroken arrows point at peculiar properties of individuals,

dotted arrows indicate natural mortality,

dashdot arrows indicate processes of migration, infection, detection, spontaneous recovery and mortality caused by TB.

Denoted by

$$s(t) = |S(t)|,\ \ l(t) = |L(t)|,\ \ d(t) = |D(t)|,\ \ b(t) = |B(t)|,$$

$$d_0(t) = |D_0(t)|,\ \ b_0(t) = |B_0(t)|$$

— sizes of cohorts at time $t \geqslant 0$,

$$R(t) = S(t) \sqcup L(t) \sqcup D(t) \sqcup B(t) \sqcup D_0(t) \sqcup B_0(t), \quad r(t) = |R(t)|,$$

— set and number of individuals, who live in region at time $t \geqslant 0$,

$$G(t) = D_0(t) \sqcup B_0(t), \quad g(t) = |G(t)|$$

— set and number of detected SSN, SSP patients, who live in region at time $t \geqslant 0$.

Denoted by ℓ_x random variable, equals lifespan of individual x, provided that it's age > 16 years and without any influence of TB.
Suppose, that all ℓ_x are similar distributed:

$$\mathsf{P}(\ell_x > a) = F_\ell(a), \quad x \in X, \ a \geqslant 0,$$

F_ℓ — given function.
If individual x survived to age $16 + \ell_x$, then he dies a natural death and he goes to cohort переходит U by dotted arrow (fig. 1).

For each pair $x \in X$, $i \in \mathbb{Z}_+$ define random variable $\tau_{LD,x,i}$:
if individual x at time $t_i \in T$ turned to cohort L,
then $\tau_{LD,x,i}$ equals time before transition of individual to cohort D by unbroken arrow (fig. 1).
Similar we define random variables $\tau_{DB,x,i}$, $\tau_{BD,x,i}$, $\tau_{D_0L,x,i}$, $\tau_{B_0D_0,x,i}$.
For all $x \in X$, $i \in \mathbb{Z}_+$ and $a \geqslant 0$

$$\begin{aligned}
&\mathsf{P}(\tau_{LD,x,i} > a) = F_{LD}(a),\\
&\mathsf{P}(\tau_{DB,x,i} > a) = F_{DB}(a),\\
&\mathsf{P}(\tau_{BD,x,i} > a) = F_{BD}(a),\\
&\mathsf{P}(\tau_{D_0L,x,i} > a) = F_{D_0L}(a),\\
&\mathsf{P}(\tau_{B_0D_0,x,i} > a) = F_{B_0D_0}(a),
\end{aligned}$$

where F_{LD}, F_{DB}, F_{BD}, F_{D_0L}, $F_{B_0D_0}$ — given functions.
We suppose, that these functions take into account peculiar properties of individuals.

Let us define transitions, which are denoted by dashdot arrows (fig. 1).

We postulate, that in the small time interval $[t; t+h)$, $t \geqslant 0$, $h \to 0+$, may occur events of the following types.

1) With probability $f_S\, p_S\, h + o(h)$, $f_S > 0$, $p_S \in [0; 1]$, to cohort S from cohort N goes individual x, reached age 16 years.

2) With probability $f_S\, (1 - p_S)\, h + o(h)$ to cohort S from cohort N goes immigrant x.

3) With probability $f_L\, p_L\, h + o(h)$ to cohort L from cohort N goes individual x, reached age 16 years.

4) With probability $f_L\, (1 - p_L)\, h + o(h)$, $f_L > 0$, $p_L \in [0; 1]$, to cohort L from cohort N goes immigrant x.

5) With probability $f_D\, h + o(h)$, $f_D > 0$, to cohort D from cohort N goes immigrant x.

6) With probability $f_B\, h + o(h)$, $f_B > 0$, to cohort B from cohort N goes immigrant x.

7) With probability $f_{D_0}\, h + o(h)$, $f_{D_0} > 0$, to cohort D_0 from cohort N goes immigrant x.

8) With probability $f_{B_0}\, h + o(h)$, $f_{B_0} > 0$, to cohort B_0 from cohort N goes immigrant x.

9) With probability $e\, h + o(h)$, $e > 0$, individual $x \in S(t) \sqcup L(t) \sqcup D(t) \sqcup B(t)$ becomes immigrant: he goes to cohort U (note, that we have $s(t) + l(t) + d(t) + b(t)$ events of this type, according to number of individuals in cohorts S, L, D, B).

10) With probability $e_{D_0}\,h + o(h)$, $e_{D_0} > 0$, individual $x \in D_0(t)$ becomes immigrant: he goes to cohort U (we have $d_0(t)$ events of this type).

11) With probability $e_{B_0}\,h + o(h)$, $e_{B_0} > 0$, individual $x \in B_0(t)$ becomes immigrant: he goes to cohort U (we have $b_0(t)$ events of this type).

12) With probability $\nu_D\,h + o(h)$, $\nu_D > 0$, individual $x \in D(t)$ dies due to TB: he goes to cohort U (we have $d(t)$ events of this type).

13) With probability $\nu_B\,h + o(h)$, $\nu_B > 0$, individual $x \in B(t)$ dies due to TB: he goes to cohort U (we have $b(t)$ events of this type).

14) With probability $\nu_{D_0}\,h + o(h)$, $\nu_{D_0} > 0$, individual $x \in D_0(t)$ dies due to TB: he goes to cohort U (we have $d_0(t)$ events of this type).

15) With probability $\nu_{B_0}\,h + o(h)$, $\nu_{B_0} > 0$, individual $x \in B_0(t)$ dies due to TB: he goes to cohort U (we have $b_0(t)$ events of this type).

16) With probability $\beta\,(1 - p_D)\,h + o(h)$, $\beta > 0$, $p_D \in [0; 1]$, contacts the pair of individuals $x \in S(t)$ and $y \in B(t)$; as a result individual x became latently infected: he goes to cohort L (we have $s(t)\,b(t)$ events of this type).

17) With probability $\beta\,(1 - p_D)\,k\,h + o(h)$, $k > 0$, contacts the pair of individuals $x \in S(t)$ and $y \in B_0(t)$; as a result individual x became latently infected: he goes to cohort L (we have $s(t)\,b_0(t)$ events of this type).

18) With probability $\beta\, p_D\, h + o(h)$, contacts the pair of individuals $x \in S(t)$ and $y \in B(t)$; as a result individual x became infected: he goes to cohort D (we have $s(t)\, b(t)$ events of this type).

19) With probability $\beta\, p_D\, k\, h + o(h)$, contacts the pair of individuals $x \in S(t)$ and $y \in B_0(t)$; as a result individual x became infected: he goes to cohort D (we have $s(t)\, b_0(t)$ events of this type).

20) With probability $\alpha\, h + o(h)$, $\alpha > 0$, contacts the pair of individuals $x \in L(t)$ and $y \in B(t)$; as a results we have secondary infected individual x: he goes to cohort D (we have $l(t)\, b(t)$ events of this type).

21) With probability $\alpha\, k\, h + o(h)$, contacts the pair of individuals $x \in L(t)$ and $y \in B_0(t)$; as a results we have secondary infected individual x: he goes to cohort D (we have $l(t)\, b_0(t)$ events of this type).

22) With probability $\varphi_D\, h + o(h)$, $\varphi_D > 0$, individual $x \in D(t)$ became notified: he goes to cohort D_0 (we have $d(t)$ events of this type).

23) With probability $\varphi_B\, h + o(h)$, $\varphi_B > 0$, individual $x \in B(t)$ became notified: he goes to cohort D_0 (we have $b(t)$ events of this type).

24) With probability $\beta_{DL}\, h + o(h)$, $\beta_{DL} > 0$, individual $x \in D(t)$ has spontaneous recovery: he goes to cohort L (we have $d(t)$ events of this type).

25) With probability $\beta_{D_0 B_0}\, h + o(h)$, $\beta_{D_0 B_0} > 0$, individual $x \in D_0(t)$ became spontaneous SSP: he goes to cohort B_0 (we have $d_0(t)$ events of this type).

26) With probability $o(h)$ occur two or more of mentioned above events.

Constants

f_S, f_L, f_D, f_B, f_{D_0}, f_{B_0}, e, e_{D_0}, e_{B_0}, ν_D, ν_B, ν_{D_0}, ν_{B_0}, β, k, α, φ_D, φ_B, β_{DL}, $\beta_{D_0B_0}$ and probabilities

p_S, p_L, p_D

are the model parameters.

Parameters k, p_S, p_L, p_D have no dimension, but other — have dimension 1/year.

3. Results of computer simulation.

The above–mentioned model vas used to work out the simulation algorithm, allows to compute sample realizations of random process $Z(t)$.
Simulation algorithm bears on standard Monte-Carlo methods.
Creation of simulation algorithm required special language for representation stochastic model in computer program.
Simulation algorithm is realized as console application `pm.exe` (Populations Modeller) for platform Win32.

Computer program `pm.exe` have:
1) compiler routine of simulation language,
2) generator of pseudo-random numbers,
3) simple statistical analysis the results of calculation,
4) possibility of simulation the community of individuals consisting $\sim$ millions of individuals;
5) verification of domain for mathematical functions and acceptable values of random variables.
Detailed theoretical description of the model, simulation language and computer program one can see in Internet (`http://iitam.omsk.net.ru/~pichugin/`, `Russian version`).

The purpose of computer simulation:
analysis the dynamics of mathematical expectations $\mathsf{E}r(t)$, $\mathsf{E}g(t)$ during time interval $t \in [0, 200]$ years,
where
$r(t)$ – number of all individuals (cohorts S, L, D, B, D_0, B_0),
$g(t)$ – number of detected patients (cohorts D_0 and B_0).
We study also the distribution lows of random variables
$r(200)$, $g(200)$.
We consider various forms of distribution functions F_ℓ, F_{LD}, F_{DB}, F_{BD}, F_{D_0L}, $F_{B_0D_0}$ and a broad set of the values of the model parameters.
Main idea:
we modify these functions and parameters and compare the results,
it may be interpret as differences between various regions.

Every computer experiment give us 100 realizations of $Z(t)$. Initial number of cohorts are assumed to be

$$s(0) = 272640,\ l(0) = 439027,\ d(0) = 926,$$

$$b(0) = 120,\ d_0(0) = 887,\ b_0(0) = 1032.$$

These values correspond to stationary solution of based differential model, which were calculated from observed data.

Experiment N 1. Distributions F_ℓ, F_{LD}, F_{DB}, F_{BD}, F_{D_0L}, $F_{B_0D_0}$ are taken in the exponential form

$$F_\ell(a) = \exp(-\mu\, a), \quad F_{LD}(a) = \exp(-\beta_{LD}\, a),$$

$$F_{DB}(a) = \exp(-\beta_{DB}\, a), \quad F_{BD}(a) = \exp(-\beta_{BD}\, a),$$

$$F_{D_0L}(a) = \exp(-\beta_{D_0L}\, a), \quad F_{B_0D_0}(a) = \exp(-\beta_{B_0D_0}\, a),$$

where μ, $\beta_{LD}, \beta_{DB}, \beta_{BD}, \beta_{D_0L}, \beta_{B_0D_0}$ — positive constants, (dimension 1/year), $a \geqslant 0$.

Experiments N 2, 3, 4. We use the another distribution lows:

$$F_{\ell}(a) = \exp\left(-\int_{16}^{16+a} (A + P\,\exp(Q\,u))\,du\right),$$

$$F_{LD}(a) = \exp\left(-\int_{0}^{a} \frac{\gamma_1\,du}{1+\gamma_2\,u}\right),$$

$$F_{DB}(a) = \frac{1}{\sqrt{2\,\pi}\,\sigma_{DB}} \int_{\ln(a/m_{DB})}^{\infty} e^{-u^2/(2\,\sigma_{DB}^2)}\,du,$$

$$F_{BD}(a) = \frac{1}{\sqrt{2\,\pi}\,\sigma_{BD}} \int_{\ln(a/m_{BD})}^{\infty} e^{-u^2/(2\,\sigma_{BD}^2)}\,du,$$

$$F_{D_0L}(a) = \max(0, 1 - 0.5\,\beta_{D_0L}\,a),$$

$$F_{B_0D_0}(a) = \max(0, 1 - 0.5\,\beta_{B_0D_0}\,a).$$

Parameters of these functions are taken such, that mathematical expectations of corresponding random variables coincide with mathematical expectations $1/\mu$, $1/\beta_{LD}$, $1/\beta_{DB}$, $1/\beta_{BD}$, $1/\beta_{D_0L}$, $1/\beta_{B_0D_0}$ for exponential distributions,
namely:
$\gamma_2 = \omega\beta_{LD}$, $\gamma_1 = \beta_{LD} + \gamma_2$, $\omega > 0$,
$m_{DB} = \exp(-\sigma_{DB}^2/2)/\beta_{DB}$,
$m_{BD} = \exp(-\sigma_{BD}^2/2)/\beta_{BD}$;
parameters A, P, Q are taken by numerical method.

Table 1

Confidence limits for $\mathsf{E}r(t)$, $\mathsf{E}g(t)$, p–level 0.95

t	$\mathsf{E}r(t)$	$\mathsf{E}g(t)$
Exp. 1		
0	714632.0 ± 0.0	1919.0 ± 0.0
50	714568.5 ± 153.8	1921.7 ± 10.6
100	714627.1 ± 170.8	1913.9 ± 10.2
150	714643.9 ± 186.9	1922.4 ± 11.2
200	714684.8 ± 166.1	1929.7 ± 9.7
Exp. 2		
0	714632.0 ± 0.0	1919.0 ± 0.0
50	704264.4 ± 172.2	1935.6 ± 10.5
100	714935.7 ± 181.1	1914.6 ± 11.3
150	714888.5 ± 160.0	1922.1 ± 11.5
200	714887.1 ± 150.8	1922.6 ± 10.6

t	$\mathsf{E}r(t)$	$\mathsf{E}g(t)$
Exp. 3		
0	714632.0 ± 0.0	1919.0 ± 0.0
50	705976.1 ± 149.7	1499.7 ± 8.9
100	716429.6 ± 179.6	1469.3 ± 7.8
150	716595.4 ± 172.3	1470.2 ± 7.5
200	716505.2 ± 174.2	1477.2 ± 7.7
Exp. 4		
0	714632.0 ± 0.0	1919.0 ± 0.0
50	701505.9 ± 160.7	2753.8 ± 11.5
100	712412.4 ± 147.9	2718.9 ± 13.1
150	712280.8 ± 163.4	2730.3 ± 13.6
200	712384.2 ± 170.8	2727.6 ± 12.1

Experiment N 1. We take the set of parameters, using parameters of based differential model.

Experiment N 2. We try to find the set of parameters, such that

$$\mathbf{E}r(t) \to r(0) = 714632,\ \mathbf{E}g(t) \to g(0) = 1919,\ t \to +\infty.$$

We obtained, that the set of parameters for exp. N 1 don't allows to solve our task. We use the new set of parameters.

Experiment N 3. We take $\sigma_{DB} = 0.1$ instead of $\sigma_{DB} = 0.8325$.
Transition time from cohort D to cohort B becomes practically deterministic and more short.
Individuals more intensive goes to cohort B and it leads to increasing the number of individuals who died due to TB;
and leads accordingly to decreasing of $g(t)$.

Experiment N 4. We take $\omega = 2.5$ instead of $\omega = 0.5$.
The probability of transition from cohort L to cohort D during short period is up.
It leads to increasing the number of individuals of cohort D;
and leads accordingly to increasing of $g(t)$.

Fig.2. Plots of estimates of $\bar{r}(t)$; 1, 2, 3, 4 — experiments.

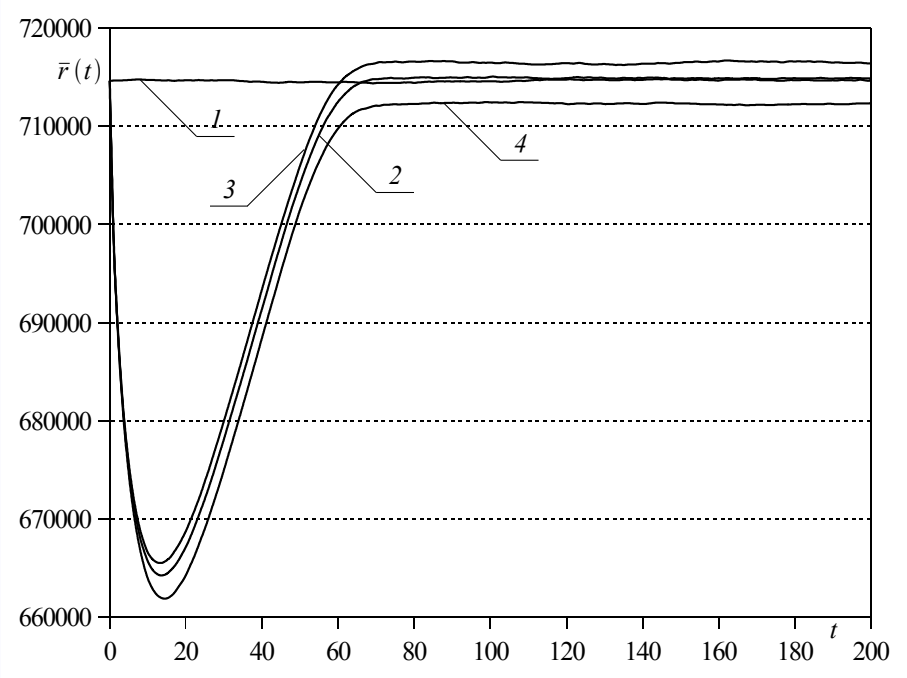
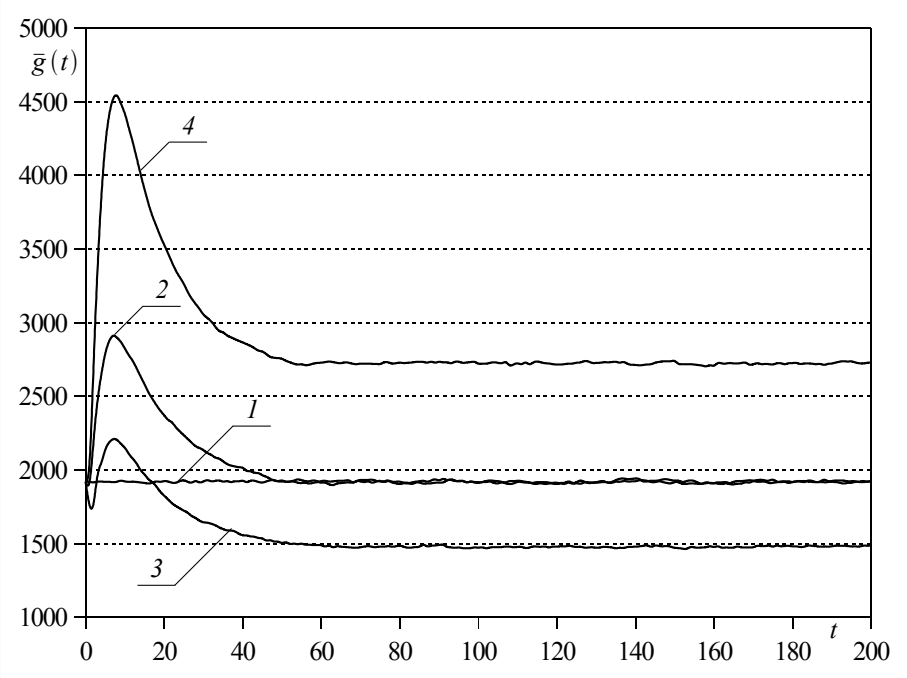


Fig.3. Plots of estimates of $\bar{g}(t)$; 1, 2, 3, 4 — experiments.



From fig. 2 and fig. 3 one can see

a) Experiment N 1: statistical estimates $\bar{r}(t)$, $\bar{g}(t)$ of mathematical expectations $\mathsf{E}r(t)$, $\mathsf{E}g(t)$ practically coincide with initial data (we have stationary random process);

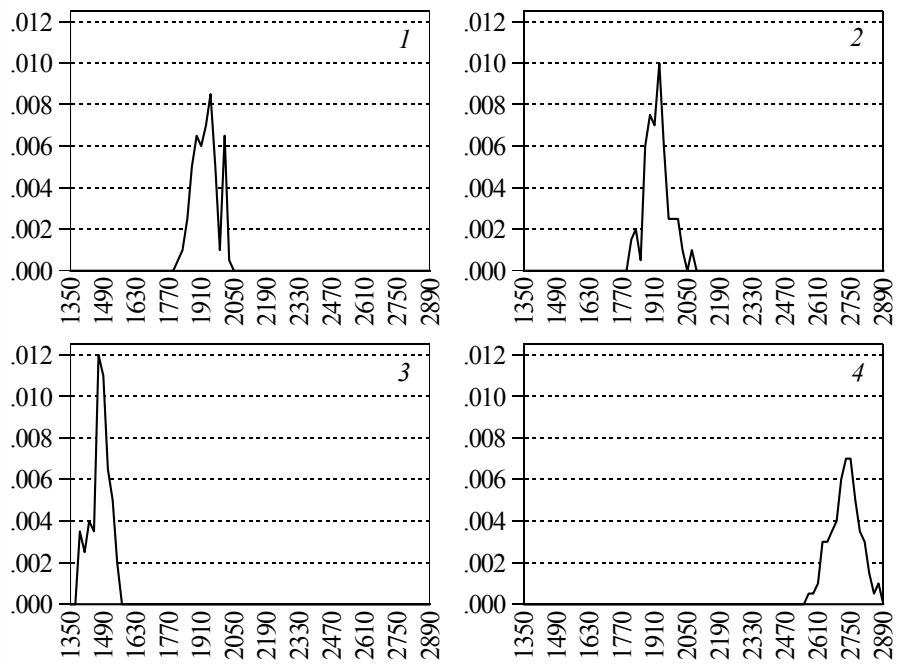
b) Experiment N 2: we have transitional process and $\mathsf{E}r(t)$, $\mathsf{E}g(t)$ tend to $r(0)$, $g(0)$, $t \to +\infty$.

One can see, that mean size of individuals — $\boldsymbol{R(t)}$
differs from one experiment to another (2, 3, 4), but very slightly.

However, mean size of notified SSN and SSP patients — $\boldsymbol{G(t)}$
differs from one experiment to another (2, 3, 4) very and very essentially !

We have, that for each experiment random variable $\boldsymbol{g(200)}$ has normal distribution with different parameters (see fig. 4).

Fig. 4. Histograms of random variable $g(200)$; 1, 2, 3, 4 — experiments.



4.Conclusion.

Our results shows that variation even one of the model parameters leads to the serious changes of distribution lows the sizes of cohorts.
It may be used as helpful information for complex data processing and study the problems of Tuberculosis Epidemiology.

We thanks academician Guri Marchuk and prof. Alexey Romanyukha (Institute of Numerical Mathematics, RAS) for the statement of the studied problem and helpful discussion of partial questions.