Bi-logistic model for disease dynamics caused by *Mycobacterium tuberculosis* in Russia

Anastasia I. Lavrova[1,2], Eugene B. Postnikov[3],
Olga A. Manicheva[2] and Boris I. Vishnevsky[2]

[1]Saint-Petersburg State University, Medical Faculty, Universitetskaya emb. 7/9, Saint-Petersburg, Russia
[2]Saint-Petersburg State Research Institute of Phthisiopulmonology, Lygovsky avenue 2-4, Saint-Petersburg, Russia
[3]Department of Theoretical Physics, Kursk State University, Radishcheva street 33, Kursk, Russia

AIL, 0000-0002-8969-535X

In this work, we explore epidemiological dynamics by the example of tuberculosis in Russian Federation. It has been shown that the epidemiological dynamics correlates linearly with the virulence of *Mycobacterium tuberculosis* during the period 1987–2012. To construct an appropriate model, we have analysed (using LogLet decomposition method) epidemiological World Health Organization (WHO) data (period 1980–2014) and obtained, as result of their integration, a curve approximated by a bi-logistic function. This fact allows a subdivision of the whole population into parts, each of them satisfies the Verhulst-like models with different constant virulences introduced into each subsystem separately. Such a subdivision could be interconnected with the heterogeneous structure of mycobacterial population that has a high ability of adaptation to the host and strong mutability.

1. Introduction

Tuberculosis remains a main problem for public health resulting in new TB cases and deaths [1] due to high development of multi- and extensively drug resistant (MDR and XDR) *Mycobacterium tuberculosis* (*MbT*). The scale of the tuberculosis epidemic with multi-drug resistance (MDR-TB) is huge; according to the World Health Organization (WHO) report, 450 000 patients with MDR-TB were registered in 2012 year. Also, according to the WHO, in 92 countries were registered the cases with extensively drug-resistant TB (XDR-TB) that is the most hard form of

disease leading to the lethal outcome. As was reported by the Department of Public Health and Communications of Ministry of Healthcare of the Russian Federation, the incidences with sensitive form of tuberculosis decreased in 2015 comparing with 2014, but MDR-TB cases were observed to rise among patients for those whose tuberculosis was detected for the first time. Thus, MDR-TB incidence increased from 4.6 to 5.2 per 100 thousand of population, and the proportion of patients with MDR-TB among persons with bacterial excretion at pulmonary tuberculosis was 47.3%. At the same time, the bacterial load of respiratory material of patients with MDR/XDR-TB does not differ from that in the drug sensitivity tuberculosis, indicating the viability of the pathogen, even with a wide range of its resistance to anti-TB drugs and high epidemiological risks of such strains [2].

Mathematical modelling of tuberculosis development and expansion could be helpful at the diagnosis and prognoses at the TB treatment on the population-level (for review, see [3]). The basis for most models recently developed is the SIR model [4] which includes different types of subpopulations (susceptible, infected and recovered) and describes transmission dynamics between them. There are relatively simple models taking into account uninfected, latent TB and active TB persons and more detailed models that include subdivisions to TB strains: drug sensitive and drug resistant allowing to describe different ways for the infection and the recovery in population in different countries [3,5–8].

To describe a trend in tuberculosis mortality using data for a long period of time, Holloway *et al.* [9] have proposed an approach based on the data fitting by logistic curves without introduction of any detailed mathematical model. Such technique allowed to characterize the dynamics of tuberculosis decline and to show that social aspects in TB prevention, namely, nutrition, hygiene, education about TB, etc., have an impact on reducing the incidence of tuberculosis in a population in most countries and could be a serious strategy in the fight against the disease.

At the same time, a not so long-term dynamics, which has non-monotone character, is still a very challenging problem. One of the key questions here is not only social but microbiological factors affecting such dynamics as well as their combination (for review, see [10,11]). In fact, the key parameter, a knowledge of which allows the conclusion, whether or not an epidemic outbreak occurs, is the basic reproduction number $\mathcal{R}_0$ [12]. It should be noted that its variation can be closely connected with the actual instant virulence of a pathogen [13,14] which defines its ability of adaptation to a host organism and higher probability to cause a disease. It evolves in dependence on the sensibility or resistance of macroorganism to infection. Therefore, in the scale of epidemic outbreak, this microorganism's property could be an important factor, which can influence the disease rate, its maximal strength and decrease [15]. In particular, an extended SIR model [14] of malaria developed recently considers the virulence as a phenomenological parameter included in the contact rate and the death rate induced by this disease. Moreover, this problem can be stated as a problem of pathogens evolution with respect to their virulence [16].

Thus, the primary goal of our work is to reveal a possible time-dependent functional composition which adequately approximates actual time dynamics of tuberculosis outbreak in Russia during last three decades reflected in WHO data; to discuss a model, which provides such dynamics, as well as possible control factors, which govern the multimodal TB epidemic activity in Russia detected over several last decades; as well as to discuss an origin of the same, qualitatively noted for *M. tuberculosis* in laboratory studies of the same period.

2. Model and method

First of all, we need to determine a class of functional dependencies, which are suitable for the data fitting. As a most natural candidate we consider logistic curves, which can be obtained as certain approximate solutions of the the three-compartmental Kermack–McKendrick SIR model [4]

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -kSI, \tag{2.1}$$

$$\frac{\mathrm{d}I}{\mathrm{d}t} = kSI - \tau^{-1}I \tag{2.2}$$

and

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \tau^{-1}I, \tag{2.3}$$

where S and I mean the number of susceptible and infected persons, respectively. The constant k is the parameter determined by the infection properties, e.g. by virulence of *MbT* strains, and social (a contact rate) reasons.

The variable R, which can be defined as

$$R(t) = \tau^{-1} \int_0^t I(t') \, dt' \tag{2.4}$$

can be interpreted as proportional to the cumulative number of persons suffering from the infection during the time interval $[0, t]$. Note that we do not connect this variable with a number of recovered/removed persons, as has been done in the original model interpretation, but rather as a generalized number of persons who cannot be further considered as susceptible to *primary* TBC infection.

In fact, we will use below the data (WHO and laboratory data) obtained for patients with tuberculosis detected for the first time, which usually is accompanied by active bacterial excretion to environment. Therefore, the main goal of the hospital's treatment is to stop an excretion and then the patient is transferred to out-patient treatment. After cessation of bacterial excretion, the patient is either completely cured—in this case, complete decomposition of the inflammation foci or fibrosis of the destruction areas occur in the lungs—or passes into the state of the chronic process, which is characterized by the unclosed inflammation foci that can result in the tuberculosis relapse. The relapse is not considered as first time detected. Whence, the parameter τ is attributed to the average duration of an active phase of tuberculosis, as it is conventionally interpreted while using SIR model and its generalizations for TB modelling [7,8,12]. Its combination with the parameter k, described above, gives the basic reproduction number $\mathcal{R}_0 = k\tau$, the value of which $\mathcal{R}_0 > 1$ corresponds to an emergence of the illness outbreak.

The system (2.1)–(2.3) is supplied with the conservation law

$$S + I + R = \text{const.} \tag{2.5}$$

that corresponds to the assumed constant population size during the time range considered in this study. Another invariant directly follows from the mathematical structure of (2.1)–(2.3) as [17] dividing equation (2.1) by equation (2.3) and then integrating the resulting equation.

$$\log\left(\frac{S}{S_0}\right) + k\tau R = 0, \tag{2.6}$$

where $R_0 = R(0) = 0$ at the very start of an epidemic.

As a result, the system (2.1)–(2.3) with (2.5)–(2.6) substituted, reduces to one nonlinear ordinary differential equation

$$\frac{dR}{dt} = \tau^{-1}[1 - R - S_0 \, e^{-k\tau R}]. \tag{2.7}$$

Following Kermack & McKendrick [4], it is possible to introduce the expansion

$$e^{-k\tau R} \approx 1 - k\tau R + \frac{(k\tau R)^2}{2},$$

which is sufficiently accurate for realistic medically observed values of the product $k\tau$ (see, for example, the discussion of applicability of such simplification to modelling plague and influenza outbreaks in [17]).

However, in contrast to the original [4], where an arbitrary $S_0 = S(0)$ is considered that is used now as a conventional approach [18], we use the normalization const. $= 1$ in equation (2.5) and take into account that $I_0 \ll S_0$, i.e. $S_0 \cong 1$ for this norm. As a result, equation (2.7) reduces to Verhulst's logistic equation

$$\frac{dR}{dt} = \tau^{-1}(k\tau - 1)R\left[1 - \frac{(k\tau)^2 R}{2(k\tau - 1)}\right]. \tag{2.8}$$

Such approach has several advantages useful for the analysis of realistic medical data. It operates with lesser number of adjustment constants that results in better stability of numerical algorithms; the solution of equation (2.8) is a relatively simple logistic function, in contrast to the conventional case of an arbitrary S_0 [4,17], which requires the hyperbolic tangent function.

At the same time, this reduced representation keeps the basic features, which are required to reproduce an epidemic outbreak: equation (2.8) has two stationary points: the unstable $R_{s1} = 0$ and the stable $R_{s2} = 2(k\tau - 1)/(k\tau)^2$. As a result, R takes its values within the interval $R \in [R(0) > 0, R_{s2}]$ and monotonously grows since $dR/dt \geq 0$ within this interval.
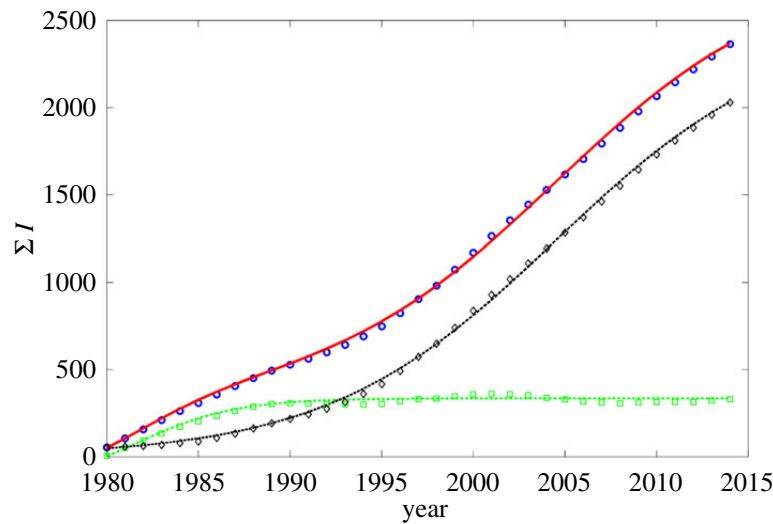
Figure 1. The dynamics of the cumulative incidence of tuberculosis per 100 000 for Russia accordingly to WHO/Europe (circles) and its bi-logistic approximation (red curve). The black curve/diamond markers and green curve/square markers show the decomposition into partial components of the approximation and the data. The parameters fitted: $d_1 = -190$, $a_1 = 12$, $b_1 = 1981.5$, $\kappa_1 = 525$; $d_2 = 0$, $a_2 = 27.5$, $\kappa_2 = 2480$, $b_2 = 2004.5$.

In particular, the second derivative

$$\frac{\mathrm{d}^2 R}{\mathrm{d}t^2} = \tau^{-1}(k\tau - 1)\frac{\mathrm{d}R}{\mathrm{d}t}\left[1 - \frac{(k\tau)^2 R}{(k\tau - 1)}\right]$$

changes its sign when $R = R_{s2}/2$, which implies that R curve has an inflection point and the number of actively infected persons I has a unique maximum there due to its definition as a derivative of R, see (2.3). Thus, a shape of outbreak will be described.

We need to note that one needs to take a non-zero (although it can be arbitrary small) initial condition $R_0 = R(0)$ to solve (2.8). It is a consequence of the simplification applied to the exact equation (2.7). However, this fact does not prevent usage of the logistic curve

$$R = \frac{\kappa_i}{1 + \exp(-\log(81)(t - b)/a)}, \tag{2.9}$$

which is the solution of equation (2.8) for the fitting epidemic curves, since the exact starting time for a widespread illness may be indefinite from medical statistics. Here the parameters are

$$a = \frac{\log(81)}{\tau^{-1}(k\tau - 1)}, \quad \kappa^{-1} = \frac{(k\tau)^2 R_0}{2(k\tau - 1)} \quad \text{and} \quad b = \frac{a \log((\kappa/R_0) - 1)}{\log(81)},$$

and the factor $\log(81)$ is used to keep correspondence to the work [19], which is supplied by the data processing software, which we will use for the primary fitting our data.

Thus, we will use further the logistic function as main building blocks for the analysis of recorded medical statistics. But these data (see blue circles in figure 1; their origin will be described in detail below) differ from a single pure logistic growth. For this reason, we will apply the so-called LogLet transform proposed in the work [19] for the decomposition of complex non-stationary growth processes into a series of saturated partial components. Such an approach was proposed in the work cited for a decomposition of the data representing some economic process, such as a sequential replacement of product models on a market. This situation has a certain similarity to the case of change of prevailing microbial strains presented among a population within the concept of so-called 'Evolution of microbial markets' [20].

If the full cumulative number of people suffering from the infection will be considered as composed by several subgroups subjected to different k_i, τ_i, then the total evolution of $R(t)$ can be represented as a sum of solutions (2.9) supplied with indices i denoting individual logistic components,

$$R = \sum_i R_i = \sum_i \left[\frac{\kappa_i}{1 + \exp(-\log(81)(t - b_i)/a_i)} + d_i\right], \tag{2.10}$$

where d_i is determined by the choice of initial background epidemic level mentioned above.

Note also that the standard basic solutions (2.9) allow the coordinate alteration, which gives a straight line in semi-logarithmic coordinates

$$\log(FR_i(t)) = \frac{\log(81)}{a_i}(t - b_i),$$

which can be obtained denoting $F_i(t) = R_i(t)/\kappa_i$ in the fraction above and expressing the logarithmic term in the denominator there as

$$FR_i = \frac{F_i(t)}{1 - F_i(t)}.$$

Being applied to the full processed data $R = \sum_i R_i$, this representation as a fraction is called the Fisher–Pry transform. If inflection points of all individual components are sufficiently separated in time, the multimodal growth curve will be transformed into a set of linear sub-intervals, each of which correspond to the individual logistic component of equation (2.10) [19].

3. Data and results of their processing

All data explored were taken from the European health for all database (HFA-DB) presented on WHO/Europe's portal [21]: incidence of tuberculosis per 100 000 for Russia within the time range 1980–2014. They are shown in figure 1 as blue circles. To evaluate their LogLet decomposition, the two-stage numerical procedure was used. At the first step, we applied the specialized online software `Loglet Lab 3.0` [22] to obtain first estimations for all coefficients in the series (2.10). The input data were represented as the cumulative sum

$$R_j = \sum_n^j I_n,$$

replacing the continual integral representation (2.4), where the index j denotes a time point of the current time t_j.

At the second step, the obtained coefficients d_i, a_i, b_i k_i were additionally adjusted pointwise to the processed data. The final values obtained at this stage are given in the caption to figure 1, which shows the function (2.10) with these parameters as the red solid curve. One can note quite accurate visual reproduction of the data dynamics that is confirmed by the statistical analysis: the correlation coefficient between actual and fitted data is equal to 98.8%.

Figure 1 demonstrates also two individual components of (2.10): green and black dashed lines correspond to indices $i = 1$ and $i = 2$, respectively, as well as the decomposition of the processed data into the partial processes obtained as difference of the total R_j and the complementary partial fitted components (they are shown as markers).

One can see that these partial processes actually resemble logistic growth with saturation (the correlations between the component's data and their fits are 98.47% and 99.97% from these components.) This conclusion is supported by the Fisher–Pry transform too (figure 2): both components are fitted by straight lines. Moreover, this representation allows clear distinguishing between the time ranges for the prevalence of each partial logistic process.

This time-separated linear components show the coexistence time period between components is quite small. This behaviour may originate from the fact that tuberculosis is a persistent infection widely represented in population but subjected to mutations of MbT that correspond to the larger values of the parameters a_2 and κ_2 in comparison with a_1 and κ_1 characterizing 'old traditional' tuberculosis. The former is introduced into the population relatively recently and is more aggressive, while the dynamics of previously prevailing strain is practically saturated, see the green curve in figure 1. Some medical evidence of these conclusions will be given in the next section.

Thus, the results of the direct bi-logistic fitting medically recorded data argue that we deal with a replacement of the main infection agent rather than with the competitive process.

Finally, figure 3 presents the dynamics model incidences obtained via the definition (2.3) with the substituted fitting function (2.10). One can see that it catches the principal bimodal character of the actual data. Owing to the construction of the model, these feature can be explained as a presence of two sequential outbreaks. The first one practically decays within the considered time interval, while the second one only started before 1980 and reached its maximum around 2005.
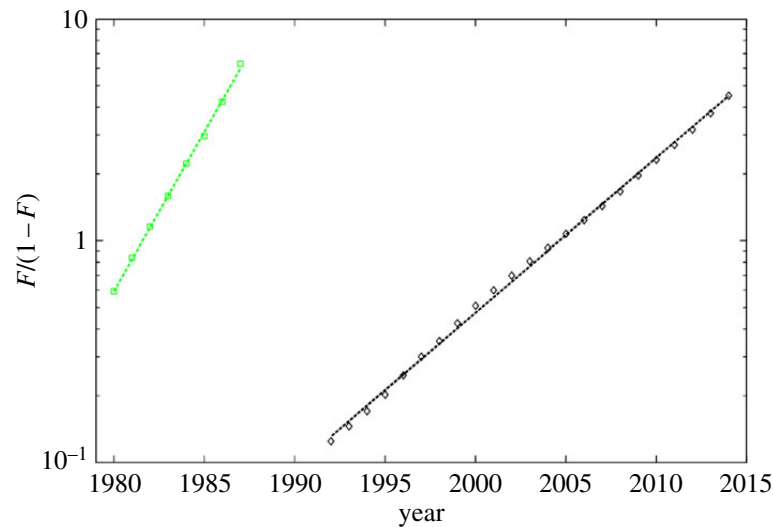
Figure 2. The Fisher–Pry LogLet transform of the data presenting cumulative incidence of tuberculosis per 100 000 for Russia in semi-logarithmic scale. The notation of partial components is the same as in figure 1.
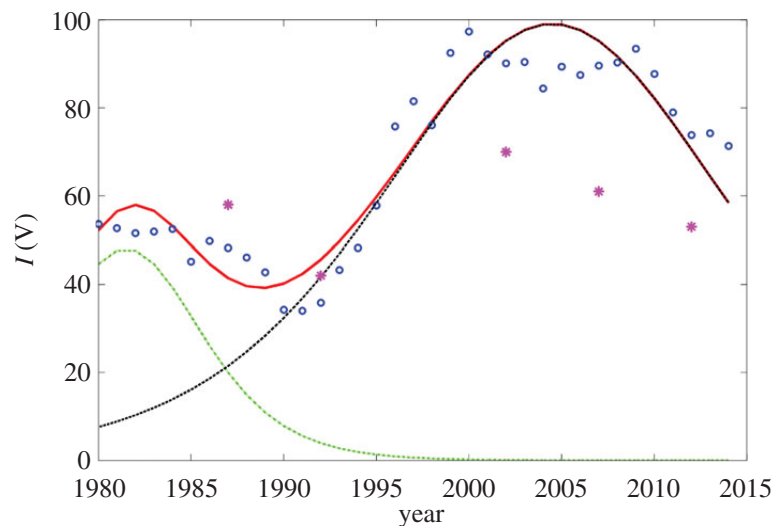
Figure 3. The dynamics of the incidence of tuberculosis per 100 000 for Russia: the actual data accordingly to WHO/Europe (circles) and their bi-logistic model (solid red curve) as well as its two partial components (dashed green and black curves). The asterisks show the known experimental virulence data.

4. Discussion and conclusion

The data processing of TB incidences evaluated based on the bi-logistic combination of the simplified SIR models shows that this models reproduces the observed time dynamics with sufficiently high accuracy. It reveals the detailed background of the non-monotonic dynamics as a coexistence of two SIR processes with a time lag and different basic reproduction numbers $\mathcal{R}_0^{(1)} = k_1 \tau_1 = 1.0019$ and $\mathcal{R}_0^{(2)} = k_2 \tau_2 = 1.0099$. That both values are close to unity corresponds to a persistent character of TBC among the population, but their slight overcoming of unit value leads to long-time acting outbreaks. $\mathcal{R}_0^{(2)} > \mathcal{R}_0^{(1)}$ and figure 3 shows that the second outbreak, which started in the late 1970s (see black dashed line in figure 3) and presents up to date is more strong (compare green and black dashed curves) in comparison with that one, which reached its maximum in early 1980s (see green line).

It should be pointed out that such transient behaviour is not unique among the time courses of infectious diseases revealed by mathematical modelling. For example, a study [23] of the classic example of localized epidemic outbreak, the plague in Eyam village in 1665–1666, shows the picture quite similar to our figure 1, which has been interpreted as a two-phase process, where the second phase corresponded to the more deadlier (i.e. more virulent) form of the disease. This also allows the discussion of the change in virulence data of TB obtained in laboratory investigations (they are shown in figure 3 as asterisks).

It should be noted that experimental data on the correlation between strains with high and low virulence were obtained using various models, both in animals (subcutaneous, intracranial infection of

guinea pigs and also intravenous infection of mice) and on the THP-1 cell line. Each value (asterisks in figure 3) is the percentage of strains with high virulence with respect to the total number of strains of the pathogen investigated during this period; these data have been obtained from patients of the clinics of the St Petersburg Research Institute of Phthisiopulmonology in 1987–2012.

Recently, it is argued [13] that the pathogen virulence has a multifaceted interconnection with the values of the basic reproduction number $\mathcal{R}_0$ since, in particular, the growing virulence can result in slower recovery from infection and increased probability of a contact transmission. In the case of tuberculosis, the low effectiveness of therapy (39.1%) in patients with multiple and broad drug resistance (MDR and XDR) of the pathogen leads to the same consequences (slower recovery from infection and increased probability of contact transmission) [24].

Moreover, in addition to the classic coevolution model of host–virus interaction [25] (see also [10]), which connects virulence and $\mathcal{R}_0$ by a nonlinear functional dependence and assumes maximizing virulence as maximizing the basic reproduction number, modern studies consider competitive population dynamics and coevolution of virulent agents in host populations composed by several groups [26].

It should be noted that tuberculosis is the result of the dynamic and multiple-factor interaction of the host and the pathogen with specific genetic and phenotype characteristics for each of them. At the same time, the virulence of a particular strain population isolated from a specific host (macroorganism) during disease is determined by the realization of the pathogenic potential of the cells of this population, their adaptation to the internal environment of the given host. The host–pathogen relationships are realized at different stages: from molecular (genes, proteins, receptors, etc.) to population level. Each level has its own spatial volume and duration of interaction of macro- and microorganisms, while each subsequent level is characterized by the extension of time and space. For example, adhesion of *M. tuberculosis* on the surface of a human cell caused by the interaction of surface molecules of the cell membrane of the pathogen and receptors of the epithelium of the respiratory tract or macrophages is carried out simultaneously and in a volume determined by the molecular level. Bacteria phagocytosis, inflammation and then destruction in the lung or other tissue and finally systemic changes in macro-organism can last over a long period of time (days–months). Coevolution, or interaction between the population of strains of the pathogen circulating in a certain region and the host population, characterized by certain socio-, geo- and ethnic characteristics, continues for decades; *Homo sapiens–Mycobacterium tuberculosis* relationship lasts a thousand years. Of course, the host–pathogen relationships are realized at all levels simultaneously, where all levels interact in a mutual manner. Hence, we would like to clarify that the definition of 'virulence' is more correct to apply for the host–pathogen relationship only at the level of the macroorganism/pathogen, but for the population level, which is studied by the methods of molecular epidemiology, it would be better to suggest the concept of pathogen virulent potential of phylogenetic line/subpopulation/cluster. It should be emphasized that the host–pathogen relationship in tuberculosis is affected by the geno- and phenotypic heterogeneity of the populations of both players of this process [27,28].

This is our case, when two subpopulations are considered susceptible to two possible groups of mycobacteria characterized by different virulence reflected in $\mathcal{R}_0^{(1)}$ and $\mathcal{R}_0^{(2)}$ and extinction of the first group influence in the mid-1990s. The second of these groups may be a cluster of strains of *M. tuberculosis* belonging to the Beijing genetic line, one of the most representative lines on the territory of the Russian Federation. In the years 2004–2006, their population reached 44.4% in the Pskov region [29], 34.0% in central Russia [30], 50.0% in Tuva republic [31], 71.0% in Eastern Siberia [32]. In general, members of this family are distinguished by a high transmission [33], association with multi-drug resistance [34,35] and the ability to cause disease, where it should be noted that the variability of pathogenic characteristics are associated with genetic heterogeneity of strains within bacterial subpopulations [36]. The most successful clone (subline) of the Beijing family on the territory of Russia is B0/W148 [37,38]. For example, in the territory of the Republic of Karelia has been observed an insignificant decrease of the strains fraction of the Beijing family from 67.0% in 2006 to 55.1% in 2012, while the fraction of clonal cluster B0/W148 has increased from 22.9% to 34.9%, respectively. In St Petersburg (the neighbouring region), these ratios remained unchanged. At the same time, the strains of the B0/W148 cluster were 18.0% (2006) and 37.6% (2012) of the number of drug-resistant strains isolated in the territory Karelia. A high degree of association of this subline with drug resistance has been obtained also for strains in St Petersburg [37,39].

Note, respectively, that the ratio of these subgroups (black and green lines in figures 1–3) presented in the whole population changes with time that gives time-varying data for the laboratory determined virulence [40]. Thus, such subdivision opens the question for more detailed characterization and

classification of pathogens and subjects of their action in microbiological studies that can be considered as a future perspective challenge.

References

1. World Health Organization. 2016 *Global tuberculosis report*. Geneva.

2. Yablonskii PK, Vizel AA, Galkin VB, Shulgina MV. 2015 Tuberculosis in Russia: its history and its status today. *Am. J. Resp. Crit. Care* **191**, 372–376. (doi:10.1164/rccm.201305-0926OE)

3. Zwerling A, Shrestha S, Dowdy DW. 2015 Mathematical modelling and tuberculosis: advances in diagnostics and novel therapies. *Adv. Med.* **2015**, 1–10. (doi:10.1155/2015/907267)

4. Kermack WA, McKendrick AG. 1927 A contribution to the mathematical theory of epidemics. *Proc. R. Soc. Lond. A* **115**, 700–721. (doi:10.1098/rspa.1927.0118)

5. Porco TC. 1998 Quantifying the intrinsic transmission dynamics of tuberculosis. *Theor. Popul. Biol.* **54**, 117–132. (doi:10.1006/tpbi.1998.1366)

6. Feng Z, Castillo-Chavez C, Capurro AF. 2000 A model for tuberculosis with exogenous reinfection. *Theor. Popul. Biol.* **57**, 235–247. (doi:10.1006/tpbi.2000.1451)

7. Shilova MV, Dye C. 2001 The resurgence of tuberculosis in Russia. *Phil. Trans. R. Soc. Lond. B* **356**, 1069–1075. (doi:10.1098/rstb.2001.0895)

8. Gomes MGM, Franco AO, Gomes MC, Medley GF. 2004 The reinfection threshold promotes variability in tuberculosis epidemiology and vaccine efficacy. *Proc. R. Soc. Lond. B* **271**, 617–623. (doi:10.1098/rspb.2003.2606)

9. Holloway KL, Staub K, Rühli F, Henneberg M. 2014 Lessons from history of socioeconomic improvements: a new approach to treating multi-drug-resistant tuberculosis. *J. Biosoc. Sci.* **46**, 600–620. (doi:10.1017/S0021932013000527)

10. Anderson RM, May RM. 1992 *Infectious diseases of humans: dynamics and control*. Oxford, UK: Oxford University Press.

11. Yoshikura H. 2012 On case-fatality rate: review and hypothesis. *Jpn. J. Infect. Dis.* **65**, 279–288. (doi:10.7883/yoken.65.279)

12. Roberts MG. 2007 The pluses and minuses of R_0. *J. R. Soc. Interface* **4**, 949–961. (doi:10.1098/rsif.2007.1031)

13. Porco TC, Lloyd-Smith JO, Gross KL, Galvani AP. 2005 The effect of treatment on pathogen virulence. *J. Theor. Biol.* **233**, 91–102. (doi:10.1016/j.jtbi.2004.09.009)

14. Alizon S, Hurford A, Mideo N, van Baalen M. 2008 Virulence evolution and the trade-off hypothesis: history, current state of affairs and the future. *J. Evol. Biol.* **22**, 245–259. (doi:10.1111/j.1420-9101.2008.01658.x)

15. Mackinnon MJ, Read AF. 2004 Virulence in malaria: an evolutionary viewpoint. *Phil. Trans. R. Soc. Lond. B* **359**, 965–986. (doi:10.1098/rstb.2003.1414)

16. Bolker BM, Nanda A, Shah D. 2010 Transient virulence of emerging pathogens. *J. R. Soc. Interface* **7**, 811–822. (doi:10.1098/rsif.2009.0384)

17. Murray JD. 2012 *Mathematical biology: I. An introduction*. Berlin, Germany: Springer.

18. Frauenthal JC. 2012 *Mathematical modeling in epidemiology*. Berlin, Germany: Springer Science & Business Media.

19. Meyer PS, Yung JW, Ausubel JH. 1999 A primer on logistic growth and substitution: the mathematics of the Loglet Lab Software. *Technol. Forecast. Soc.* **61**, 247–271. (doi:10.1016/S0040-1625(99)00021-9)

20. Werner GDA et al. 2014 Evolution of microbial markets. *Proc. Natl Acad. Sci USA* **111**, 1237–1244. (doi:10.1073/pnas.1315980111)

21. European health for all database (HFA-DB) WHO/Europe. See http://data.euro.who.int/hfadb/.

22. Loglet Lab 3.0 Javascript Edition. See http://lizardinthesun.com/loglet/.

23. Massada E, Coutinho FAB, Burattini MN, Lopez LF. 2004 The Eyam plague revisited: did the village isolation change transmission from fleas to pulmonary?. *Med. Hypotheses* **63**, 911–915. (doi:10.1016/j.mehy.2004.03.028)

24. Sterlikov SA, Testov VV, Vasilieva IA. 2016 Treatment results of patients with multiple and extensively drug-resistance registered in 2012 in the Russian Federation and in the world. *Tuberk. Bolezn. Legk.* **94**, 22–27. (in Russian). (doi:10.21292/2075-1230-2016-94-11-22-28)

25. Anderson RM, May RM. 1982 Coevolution of hosts and parasites. *Parasitology* **85**, 411–426. (doi:10.1017/S0031182000055360)

26. Williams PD. 2011 New insights into virulence evolution in multigroup hosts. *Am. Nat.* **179**, 228–239. (doi:10.1086/663690)

27. Brites D, Gagneux S. 2015 Co-evolution of *Mycobacterium tuberculosis* and *Homo sapiens*. *Immunol. Rev.* **264**, 6–24. (doi:10.1111/imr.12264)

28. Gagneux S. 2012 Host–pathogen coevolution in human. *Phil. Trans. R. Soc. B* **367**, 850–859. (doi:10.1098/rstb.2011.0316)

29. Mokrousov I, Vyazovaya A, Otten T, Zhuravlev V, Pavlova E. 2012 *Mycobacterium tuberculosis* population in Northwestern Russia: an update from Russian-EU/Latvian border region. *PLoS ONE* **7**, e41318. (doi:10.1371/journal.pone.0041318)

30. Ivanov IY, Stepanshina VN, Lipin MT, Korobova OV, Shemiakin IG. 2004 Spoligotypes of clinical *Mycobacterium tuberculosis* strains isolated in patients with tuberculosis in the Central Region of Russia. *Probl. Tuberk. Bolezn. Legk.* **4**, 23–27. (full text in Russian; PubMed: 15171039).

31. Matrakshin AG et al. 2004 Genotypic characteristics of *Mycobacterium tuberculosis* strains from the Republic of Tyva. *Probl. Tuberk. Bolezn. Legk.* **3**, 37–40. (full text in Russian; PubMed: 15338899).

32. Medvedeva TV, Ogarkov OB, Nekipelov OM, Ushakov IV, Koziakova ES, Skvortsova RG. 2004 MIRU-VNTR genotyping of *Mycobacterium tuberculosis* strains from East Siberian: Beijing family versus Kilimanjaro family. *Mol. Gen. Mikrobiol. Virusol.* **4**, 33–38. (full text in Russian; PubMed: 15597570) .

33. Hu Y, Ma X, Graviss EA, Wang W, Jiang W, Xu B. 2011 A major subgroup of Beijing family *Mycobacterium tuberculosis* is associated with multidrug resistance and increased transmissibility. *Epidemiol. Infect.* **139**, 130–138. (doi:10.1017/S0950268810000890)

34. Hanekom M, Gey van Pittius NC, McEvoy C, Victor TC, Van Helden PD, Warren RM. 2011 *Mycobacterium tuberculosis* Beijing genotype: a template for success. *Tuberculosis* **91**, 510–523. (doi:10.1016/j.tube.2011.07.005)

35. European Concerted Action on New Generation Genetic Markers and Techniques for the Epidemiology and Control of Tuberculosis. 2006 Beijing/W genotype *Mycobacterium tuberculosis* and drug resistance. *Emerg. Infect. Dis.* **12**, 736–743.

36. Hanekom M et al. 2007 A recently evolved sublineage of the *Mycobacterium tuberculosis* Beijing strain family is associated with an increased ability to spread and cause disease. *J. Clin. Microbiol.* **45**, 1483–1490. (doi:10.1128/JCM.02191-06)

37. Mokrousov I. 2013 Insights into the origin, emergence, and current spread of a successful Russian clone of *Mycobacterium tuberculosis*. *Clin. Microbiol. Rev.* **26**, 342–360. (doi:10.1128/CMR.00087-12)

38. Mokrousov I. 2017 Emerging clones of *Mycobacterium tuberculosis* in Russia and former Soviet Union countries: Beijing genotype and beyond. *Int. J. Mycobacteriol* **5**, 69–70. (doi:10.1016/j.ijmyco.2016.08.005)

39. Mokrousov I et al. 2015 Trends in molecular epidemiology of drug-resistant tuberculosis in Republic of Karelia, Russian Federation. *BMC Microbiol.* **15**, 279. (doi:10.1186/s12866-015-0613-3)

40. Vishnevsky BI, Manicheva OA, Yablonsky PK. 2014 Virulence of *Mycobacterium tuberculosis*. *Russ. J. Inf. Imm.* **4**, 319–330. (in Russian). (doi:10.15789/2220-7619-2014-4-319-330)