

Mathematical modelling of the progression of active tuberculosis: Insights from fluorography data



Konstantin Konstantinovich Avilov^{a, *}, Alexei Alexeevich Romanyukha^a,
Evgeny Mikhailovich Belilovsky^b, Sergey Evgenevich Borisov^b

^a Institute for Numerical Mathematics, Russian Academy of Sciences, 8, Gubkina Str., Moscow, 119333, Russia

^b Moscow Research and Clinical Center for Tuberculosis Control of the Moscow Government Department of Health, 10, Stromynka Str., Moscow, 107076, Russia

ARTICLE INFO

Article history:

Received 31 January 2022

Received in revised form 21 June 2022

Accepted 24 June 2022

Available online 30 June 2022

Handling Editor: Dr HE DAIHAI HE

Keywords:

Natural history of tuberculosis (TB)

Mathematical modeling

Minimal tuberculosis

Model structure

Fluorographic screening data

ABSTRACT

Little is known about the dynamics of the early stages of untreated active pulmonary tuberculosis: unknown are both the rates of progression and the model “scheme”. The “parallel” scheme assumes that infectiousness of tuberculosis cases is effectively pre-defined at the onset of the disease, and the “serial” scheme considers all cases to be non-infectious at the onset, with some of them later becoming infectious.

Our aim was to estimate the progression of the early stages of pulmonary tuberculosis using data from a present-day population. We used the routine notification data from Moscow, Russia, 2013–2018 that contained the results and time of the last fluorographic screening preceding the detection of tuberculosis cases. This provided time limits on the duration of untreated tuberculosis.

Parameters of TB progression under both models were estimated. By the goodness of fit to the data, we could prefer neither the “parallel”, nor the “serial” model, although the latter had a bit worse fit. On the other hand, the observed rise in the fraction of infectious tuberculosis cases with the time since the last screening was explained by the “serial” model in a more plausible way – as gradual progression of some cases to infectiousness. The “parallel” model explained it through less realistic quick removal of non-infectious cases and accumulation of the infectious ones.

The results demonstrate the potential of using such detection data enriched with reassessments of the previous screenings.

© 2022 The Authors. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Tuberculosis (TB) is a communicable disease with a complex natural history characterised by long-lasting latent infection and usually slow disease progression. Observational data on the dynamics of TB infection and early stages of the disease are extremely scarce because the processes are mostly asymptomatic and hard to detect. Yet, quantitative parameters of these

* Corresponding author.

E-mail addresses: avilov@inm.ras.ru, kkavilov@gmail.com (K.K. Avilov).

Peer review under responsibility of KeAi Communications Co., Ltd.

processes are crucial to modelling of TB epidemiology. In this study, we propose a novel approach to quantifying the dynamics of early stages of active pulmonary TB on the basis of modern routine notification data.

The natural history of disease is generally defined as “the course of a disease from its pathological onset or inception to its resolution (either through complete recovery or eventual death)” (Porta (ed.), 2014), especially in the absence of treatment. In case of TB, its natural history usually encompasses all processes from the initial infection with *M. tuberculosis* (MTB, the causative agent of TB), through latent TB infection, and to active TB disease. After the detection (diagnosing) of a TB case and, hence, the start of their treatment, the course of their disease is usually not considered “natural” anymore. For a long time, the natural history of TB was modelled mostly as a two-stage process (latent infection and active disease), but recently it has been realised that the continuous spectrum of TB infection and disease must be subdivided into more stages in order to analyse it correctly and produce better control policies (Barry et al., 2009; Drain et al., 2018; Esmail et al., 2014; Frascella et al., 2021; Houben et al., 2019; Kendall et al., 2021). The definitions and terminology differ from article to article, but the most discussed modification is splitting TB disease into clinical (symptomatic) and subclinical (asymptomatic) TB. Earlier stages of active TB are also recognised: “incipient TB” in (Drain et al., 2018; Kendall et al., 2021), and “minimal TB” in (Frascella et al., 2021) – see Fig. 1.

Because TB cases are usually detected at advanced stages of the disease, little is known about the earlier stages of TB and their dynamics (Kendall et al., 2021). Furthermore, since it is unethical to deprive diagnosed TB patients of therapy, it is not possible to obtain modern data on the natural course of TB. There are some meta-analyses of the data from the pre-chemotherapy era when no effective treatment for TB was available (Ragonnet et al., 2021; Tiemersma et al., 2011), but they do not provide enough information, and their estimates are relevant for patients from the beginning of the 20th century.

Besides that, it is unclear what structure the model of the progression of active TB should have: the stages may be sequential (as in (Frascella et al., 2021), Fig. 1) – we call it “serial model”, or the stages may be alternative to each other – we call it “parallel model” (see details below).

This study takes advantage of the features of Russian anti-TB system and its notification data:

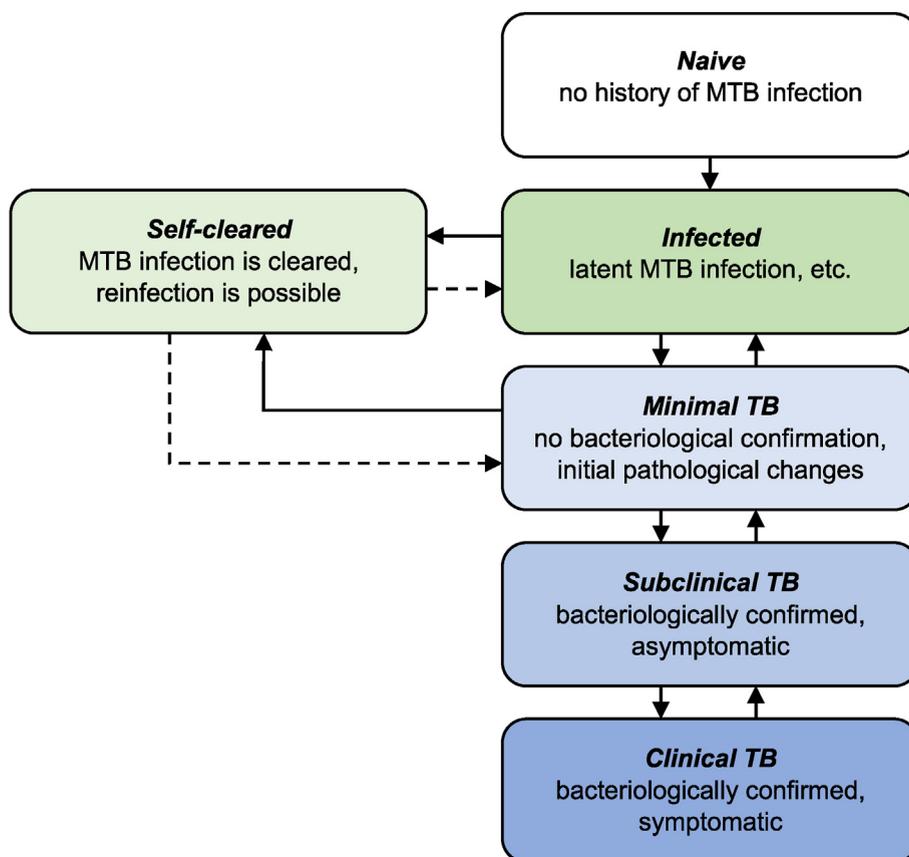


Fig. 1. The model of the natural history of tuberculosis infection and disease by (Frascella et al., 2021). Boxes represent states (stages) of the model. Green are *Mycobacterium tuberculosis* (MTB) infection. Blue denotes active tuberculosis disease. Arrows show possible state transitions. Dashed arrows represent superinfection. Bacteriological confirmation is detection of MTB in sputum. “Symptomatic” usually means “prolonged cough (>2–3 weeks)”, but other subjectively noticeable TB-associated symptoms may be used. TB = tuberculosis.

- Russian general population is relatively regularly actively screened for TB (with fluorography, i.e. chest X-ray) – at least, in big cities;
- by Russian standards, pulmonary TB cases are stratified by bacterioexcretion (i.e., by presence or absence of a significant amount of MTB in sputum);
- in the notification data, available is the information on the *previous* regular screening (time and results); this imposes limits on the duration of the untreated disease that precedes detection.

These features enabled us to build mathematical models of progression of active pulmonary TB with stages defined by bacterioexcretion (T_+ group – excreting MTB; T_- group – not excreting MTB) and estimate their parameters. T_+ group in Russian statistics comprises both subclinical and clinical TB as defined in (Frascella et al., 2021), and T_- group is approximately equivalent to minimal TB from (Frascella et al., 2021) – see Fig. 1. Thus, our study may be interpreted as modelling the interplay between minimal and subclinical/clinical TB.

We compared two “types” of mathematical models describing the progression of TB in terms of bacterioexcretion (infectiousness): “parallel model” with T_- and T_+ forms of TB effectively separated from the onset of the disease (Fig. 2A) (this “scheme” was used by the majority of epidemiologic modelling works, e.g. (Baltussen et al., 2005; Blower et al., 1995; Dye et al., 1998)), and “serial model” that assumes that all TB cases start as T_- , with some of them later progressing to T_+ (Fig. 2B) (it was used by a smaller number of works, e.g. (Avilov et al., 2015; Brogger, 1967; Perelman et al., 2004; Yaesoubi & Cohen, 2013)). There exist recent pathophysiological (Dannenberg, 1999) and empirical (Okada et al., 2012) evidences in favour of the serial model structure, but they are neither conclusive, nor provide quantitative estimations of the parameters of the model.

Our study does not attempt to build a full model of the natural history of TB: our models do not include latent infection and its activation, or TB-induced mortality. They are the models of the progression of active TB disease only.

1.1. Brief information on tuberculosis

Tuberculosis is a communicable disease caused by mycobacteria of *M. tuberculosis* complex (MTB) (Davies et al., 2014). It spreads from human to human mostly by airborne droplets. The most prevalent and the most epidemiologically significant form is pulmonary TB, with pulmonary TB cases excreting or not excreting detectable amounts of the bacilli with breathing, coughing, etc. The feature of TB is that the vast majority of infected individuals (about 90%) never develop active TB disease in their lifetime remaining latently infected with *M. tuberculosis*. On the other hand, if the infection activates and causes active TB disease, it can be lethal: TB is the leading cause of death from a single infectious agent (more than HIV/AIDS), and it is estimated that in 2019 about 10 million people developed active TB worldwide, and 1.4 million died from TB (World Health Organization, 2020). The average duration of untreated active TB is estimated to be 3 years (Tiemersma et al., 2011) or about 1.5 years for smear-positive TB and about 5 years for smear-negative TB (Ragonnet et al., 2021). Case-fatality rate is about 70% for smear-positive TB and about 20% for smear-negative TB in both studies (Ragonnet et al., 2021; Tiemersma et al., 2011).

2. Materials and methods

2.1. Data

We used a register of TB cases detected in Moscow, Russia, 2000–2018, provided by the Moscow Research and Clinical Center for Tuberculosis Control of the Moscow Government Department of Health. The register contains information on many parameters of the individual detected TB cases including bacterioexcretion status at detection, mode of detection (active or

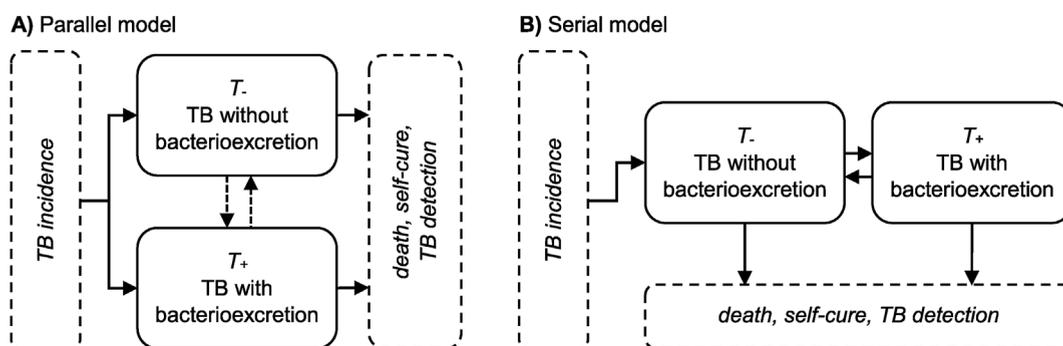


Fig. 2. Two basic structures of mathematical models of the progression of untreated active TB with regard to bacterioexcretion. A) Parallel model. B) Serial model. Transitions between groups T_- and T_+ are possible in both directions in both structures, but in the parallel model, the transitions are very weak to absent. TB = tuberculosis, T_+ = TB with bacterioexcretion, T_- = TB without bacterioexcretion.

passive), and results of the last TB X-ray screening preceding the detection event (based on expert reanalysis of the fluorography pictures; it was possible to obtain them for 68.1% of TB cases of the appropriate type).

It is important to note that it was *not* a cohort or follow-up study. The register contains data only on those persons who were eventually diagnosed with active TB. The true number of screened persons and their data are unknown.

The register works from the following definitions:

- Bacillary pulmonary TB (TB with bacterioexcretion, infectious TB) – pulmonary TB with MTB presence in sputum detected by cultural or bacterioscopic (including sputum smear microscopy) methods. Detection of MTB via polymerase chain reaction (PCR) tests only is not considered as bacterioexcretion. Thus, “bacillary TB” is basically “bacteriologically confirmed TB”, and it comprises both clinical and subclinical TB as defined in (Frascella et al., 2021).
- Non-bacillary pulmonary TB (TB without bacterioexcretion, non-infectious TB) – active pulmonary TB without MTB presence in sputum, as per cultural and bacterioscopic methods (both tests are always performed). In this case, the diagnosis of active TB is made by an expert board of phthisiologists on the basis of a sum of clinical, radiological, and laboratory data. The data include clinical symptoms like prolonged cough, chest X-ray and computed tomography results, PCR and IGRA tests (T-SPOT.TB, etc.), advanced CFP10-ESAT6 skin tests (Diaskintest) and others. In rare uncertain cases, the final diagnosis is made after a presumptive anti-TB treatment. Hence, non-bacillary TB may be viewed as approximately equivalent to “minimal TB” from (Frascella et al., 2021).
- Fluorography screening – regular screening with chest X-ray. The formal requirement for the adult general population of Russia is to be screened once a year, although it is rarely strictly fulfilled. Some groups (healthcare and social service workers, teachers, kindergarten workers, recovered TB patients, etc.) are screened more often. The screening results are “abnormal” if there are any shadows on the X-ray picture that are expertly considered to be suspicious of TB. Otherwise, the result is normal or “healthy”.
- Actively detected TB case – a TB case detected at a screening (mostly fluorography). The patient had no particular impetus to be screened other than general screening schedules.
- Passively detected TB case – a TB case diagnosed when they have presented to a doctor with TB-associated symptoms (i.e., some symptoms or general bad feeling caused by TB have urged the patient to seek medical advice or triggered the clinician's suspicion).

Results of the previous TB screening are provided in the register in categorical form: “never screened”, “previous fluorography picture is not found”, “healthy”, “ill (missed pathologic process in the lungs)”; the time since the previous screening is discrete: “<1 yr”, “1–2 yrs”, “3–5 yrs”, “>5 yrs”.

Main inclusion criteria in the study were: newly detected pulmonary TB, age ≥ 15 years, the result of the previous screening is “healthy”, detection years 2013–2018 (see further details and more patients' characteristics in Appendix A). This resulted in total 5616 cases included in the study, which is 66.7% of all TB cases of the appropriate type detected in Moscow in 2013–2018 and registered in the database. The infection status of the persons in hand at screening was unknown: they could have been either sensitive or latently infected. We used the data only from a narrow time-frame because our computational model required an assumption of constant detection parameters, and there was a significant increase in the fraction of actively detected cases in years before 2013 (see Appendix A, Fig. AF1), which most likely signals a change in the case-detection methodology.

The distribution of the cases with regard to bacterioexcretion and mode of detection is presented in Fig. 3 and in Appendix B.

The vast majority of Russian citizens is immunised with the BCG vaccine at birth, which is believed to reduce active TB in children severely, although its influence on the adult population is uncertain (Abubakar et al., 2013). For this reason, we used the data only from TB cases of at least 15 years old.

As the study used only retrospective depersonalised notification data, ethics approval was not required.

2.2. Modelling principles

Our goal is to construct a mathematical model that describes the progression of active pulmonary TB in terms of stages defined by the presence or absence of bacterioexcretion. So, given the data described above, we could say that, since the detected disease had started sometime between the previous screening with “healthy” result and the moment of detection, the observed growth of T_+ fraction among the detected cases with the time since the previous screening (Fig. 3B) is a directly measured progression rate of the disease because the more time has passed since the screening, the more time the disease had to progress. But it would be wrong because the data is a product of both disease progression and case-detection processes, with case-detection depending upon both the time since the previous screening and the severity of the disease (details below). This creates non-uniformity of the distribution of the disease onset time-points over the “screening-detection interval” and makes the detected cases a skewed sample of the undetected ones. So, we had to create a combined model of the progression of TB and the case-detection process and fit it to the detection data, thus estimating the parameters of both processes.

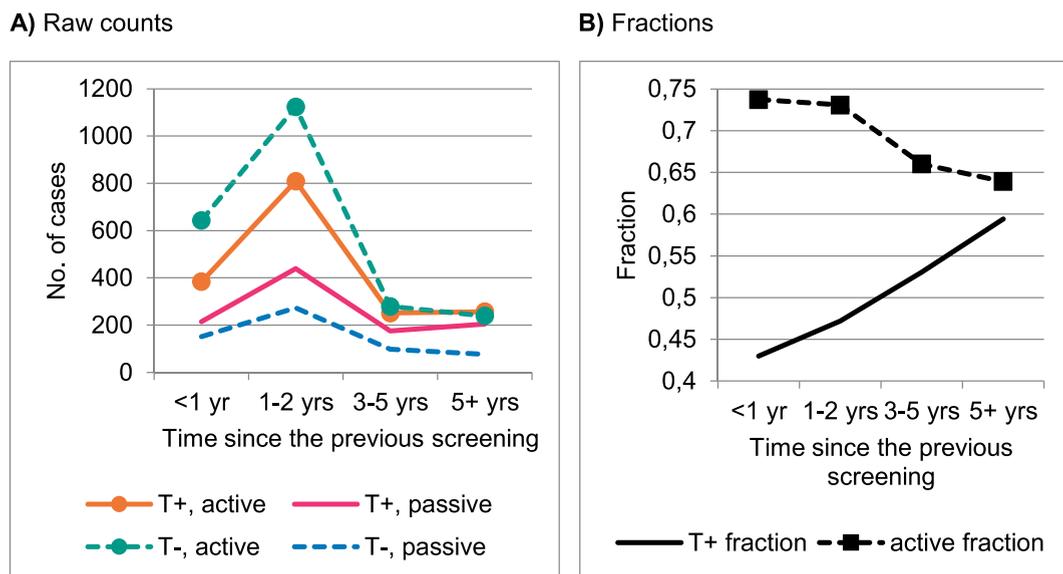


Fig. 3. Data on the newly detected pulmonary TB cases in Moscow, Russia, 2013–2018, 15+ years old, healthy at the last screening before detecting TB. **A)** Raw numbers of cases, stratified by bacillary status at detection and active/passive detection mode. **B)** Fractions of bacillary (T_+) cases and actively detected cases among all detected cases. TB = tuberculosis, T_+ = TB with bacterioexcretion, T_- = TB without bacterioexcretion.

It is possible to construct an agent-based model in calendar time, as we did in our previous work (Avilov et al., 2019), but it proved to be not effective. The new approach is to simplify the assumptions so that an analytical model is feasible. By assuming that case-detection parameters and the force of TB infection do not change with *calendar* time, we have all detected TB cases undergoing the same probabilistic “background process” from their last screening to detection. So, we could equivalently model the probability distribution of a single screened person to develop TB and then be detected at every stage of TB via every detection method, or we could model the fate of a large group of individuals screened at one moment in time and calculate the detection counts they generate. By choosing the latter variant, we modelled a virtual cohort of people who had undergone their most recent X-ray screening with “healthy” result at time $t = 0$ (so t becomes “time since screening” instead of “calendar time”) – we call them “model population”. With t growing, the people undergo new regular screenings (which withdraws them from the model population and implicitly sends them back to $t = 0$), develop active TB, self-cure or die from TB, and are detected as TB cases via the active method (if they are screened again while ill) or the passive method (with constant rates depending on T_-/T_+ state). The model detection fluxes are then fitted to the real ones.

The possibility to merge all real TB cases regardless of their calendar time of detection and to model them as coming from a single virtual cohort is based on the fact that if, in a simulational model, for each possible day of screening we created a separate, yet *identical* cohort and then measured at a given day (or time interval) the incidence generated by these cohorts and the corresponding distribution of the “times since the last screening”, we would obtain the same distribution as generated by a single cohort. Obviously, the property is lost if the “day cohorts” are not identical to each other (e.g., if they experience different case-detection rates).

The general principles of our mathematical model follow the vast majority of published models of TB epidemiology (e.g., (Avilov et al., 2015; Baltussen et al., 2005; Blower et al., 1995; Brogger, 1967; Dye et al., 1998; Murray & Salomon, 1998; Perelman et al., 2004; Waaler et al., 1962; Yaesoubi & Cohen, 2013)): we used a compartmental epidemiological model based on ordinary differential equations; it had no stratification by age or sex in the model population, but clearly differentiated T_- and T_+ TB cases by using two separate groups for them (thus forming the submodel of the progression of active TB). Transitions between model groups were defined by constant or time-dependent *per capita* rates. In spite of the extreme simplicity of the model and possible underrepresentation of the complexity of underlying processes, such models are widely used as analytical and policy-making tools (including the basic SLT model in TB (Waaler et al., 1962) and general epidemiological SEIR model (Anderson & May 1992)). One of our aims was to test the applicability of the serial and parallel structures of models of the progression of active TB under the current modelling paradigm, and so we used the very basic standard approaches to modelling the progression of TB.

Modelling case-detection is more ambiguous. Not so many published TB epidemiological models explicitly include case-detection processes. In this work, we followed a simple *per capita* rate-based detection model that we had used in our previous works (Avilov et al., 2015, 2019). This model assumes that all individuals in the population have equal access to health care, which appears to be a reasonable first approximation for the resident population of a big city like Moscow.

Just like in (Avilov et al., 2019), passive case-detection was modelled by time-independent *per capita* rates which were different for T_- and T_+ cases. This corresponds to the plausible assumption that the longer the patient is ill, the higher is their

cumulative chance to seek medical care because of the symptoms. We expected passive detection to be more effective for T_+ cases who have more severe symptoms on average, which gives T_+ cases more impetus to seek medical help. T_- cases, because of mild symptoms, are expected to be less likely to seek medical care or to trigger the clinician’s suspicion if they contact the ill for some non-TB-related reason.

Active case-detection (screening) could not be modelled by a constant rate because of the aforementioned fluorography regularity rules and minimal time-intervals between screenings. Thus, the fluorographic active screening rate φ_a in the model varied over t , with $\varphi_a(t = 0) = 0$, peaking at $t = 1 \dots 3$ yrs, and likely going down afterwards to crudely model the selection of individuals less compliant to regular screening. In our model, active case-detection is equally effective in detecting T_- and T_+ patients if they undergo screening: we assumed that high-quality fluorography is able to detect both small and big pathologic processes in the lungs with similar sensitivity.

2.3. Mathematical model

We constructed a combined compartmental mathematical model that can be trimmed to either the parallel or serial model by zeroing out certain constants. The scheme of the model is presented in Fig. 4, and the ordinary differential equations that govern the model are in Appendix A.

We used a simplified one-group submodel of TB infection and activation. Normally, it would consist of at least two model groups (S – susceptible, L – latently infected) which allow explicit modelling of infection ($S \rightarrow L$ flux) and activation ($L \rightarrow T$ flux). But since we were not trying to build a full model of the natural history of TB (that starts from the moment of infection) and were focused only on the progression between the stages of active TB disease, the only relevant quantity from the S and L stages was the total active TB incidence. In a full model, the incidence would be a combined result of the starting prevalence of latent infection, the force of infection, the rate of activation, and the rates of re-screening and death. Yet in practice, if the epidemiologic situation is relatively stable and only a minute fraction of individuals develops active TB, all these processes would stack and affect the resulting TB incidence multiplicatively, with re-screening totally dominating all other processes. Thus, we got rid of the unneeded free parameters by merging both susceptible and latently infected individuals into a single group \hat{S} and introducing the effective (stacked) infection-and-activation rate $\hat{\lambda}$ so that the total instantaneous TB incidence became $\hat{\lambda}\hat{S}$. \hat{S} starts at $t = 0$ as an arbitrary very big number, and $\hat{\lambda}$ is always a free parameter that scales the total incidence to the needed value. Individuals from \hat{S} are also withdrawn by screening (at the time-dependent rate $\varphi_a(t)$) and general mortality (μ). So, the function of group \hat{S} is basically to feed incident TB cases to T_- and T_+ at a rate that diminishes with t proportional to the fraction of people having not screened again by time t . Parameter $\hat{\lambda}$ is purely technical, and it has no practical epidemiological interpretation.

The disease progression submodel consists of two groups: T_- for untreated TB cases without bacterioexcretion, and T_+ for ones with bacterioexcretion. b is the fraction of incident cases that develop bacterioexcretion very quickly, effectively instantaneously. μ_- and μ_+ denote combined rates of death and self-cure of T_- and T_+ . Here, self-cure is defined as spontaneous cessation of any active TB disease. The reason to combine death and self-cure of TB cases is that, in practical calculations based on case-detection counts, only the sum of these rates matters. If a few self-cured persons returned back into group \hat{S} , it would change the size of \hat{S} by a vanishingly small amount and similarly negligibly affect incidence $\hat{\lambda}\hat{S}$ (see Appendix A for discussion). From the computational point of view, the effect would be lost in data noise and would only generate numerical instability. So, we chose to stabilise the model by withdrawing the self-cured from the model population and disregarding short-term TB relapse. γ is the rate of progression to bacterioexcretion (moving from T_- to T_+), and δ is the rate of regression to non-bacillary TB (moving from T_+ to T_-).

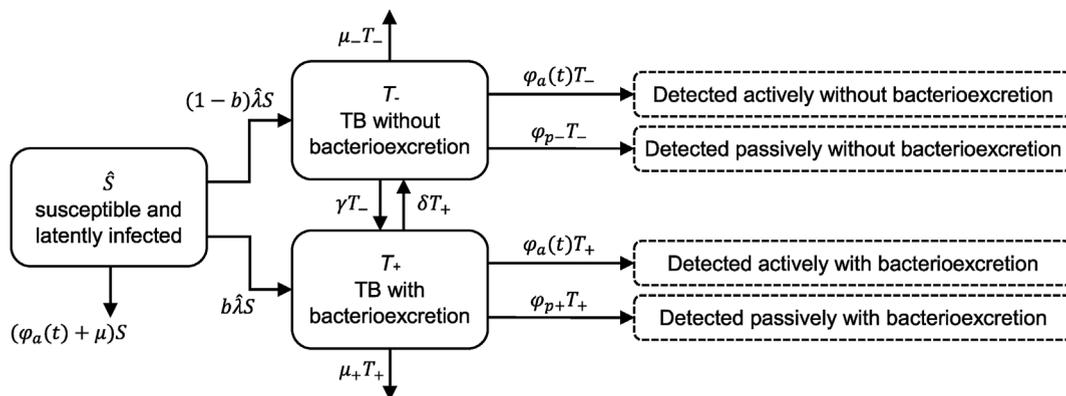


Fig. 4. Scheme of the combined model. Solid boxes are model groups (\hat{S} , T_- , T_+). Dashed boxes are accumulated detection fluxes. Arrows show instantaneous fluxes and labels their intensities. The model population is a virtual cohort of people who underwent their last screening at time $t = 0$ with “healthy” result. $\varphi_a(t)$ is the screening rate which depends upon t , the time since the last screening. Description of other parameters is in Table 1. TB = tuberculosis.

Active case-detection is modelled by the same $\varphi_a(t)$ rate for both T and T_+ . Passive detection rates φ_p and φ_{p+} are time-independent but different for T and T_+ . Time dependence of $\varphi_a(t)$ is modelled as piece-wise linear function, with values at time-nodes being free parameters (see details in [Appendix A](#)).

2.4. Fitting the model to data

The main fitting target is to approximate the real TB detection data ([Fig. 3A](#)) by the four model detection fluxes ([Fig. 4](#)). To do so, the model detection fluxes were integrated in the respective time intervals. For “<1yr”, “1–2yrs”, and “3–5yrs” bins, the limits are obvious: $t \in [0; 1]$, $[1; 3]$, $[3; 5]$ years. The “≥5yrs” bin is ambiguous: it was approximated by $t \in [5; 10]$ years interval because by 10 years after screening almost all individuals are re-screened and the remaining ones are a very selected and non-representative subset of the starting population.

Besides that, a regularization of the fitting process was used in the form of targeting the average duration of untreated TB. Average disease duration D was calculated analytically from the parameters of the TB progression submodel (see [Appendix A](#)). The target value of D was $D_0 = 3$ years (in accordance with ([Tiemersma et al., 2011](#))).

The final target function to be minimized was

$$F = w(D - D_0)^2 + \sum_{i=-,+} \sum_{j=a,p} \sum_{k=1..4} \left(\Phi_{i,j,k}^{model} - \Phi_{i,j,k}^{data} \right)^2,$$

where $\Phi_{i,j,k}$ is the count of cases in the detection bin i, j, k ($i = -, +$ denotes non-bacillary and bacillary TB, $j = a, p$ denotes active and passive detection, and $k = 1 \dots 4$ denotes the time-bin: 1 = “<1yr”, 2 = “1–2yrs”, 3 = “3–5yrs”, 4 = “≥5yrs”), and w is a weight coefficient chosen heuristically to obtain good fit to both detection data and disease duration.

We used a library version of the interior-point minimisation algorithm supplied in the Optimization Toolbox™ of MathWorks Matlab (fmincon function). Because the optimisation process sometimes got stuck in local minima or did not converge properly, we restarted each optimisation task 50 times using random initial values of free parameters (see [Appendix A](#) and code in [Appendix C](#)) and chose the solution with the lowest value of function F . This allowed us to obtain stable and reliable results.

To convert the basic model into the parallel or serial one, we forced certain coefficients to be zero: for the parallel model, we set $\gamma = \delta = 0$; and for the serial model, we set $b = 0$. The combined model is the basic model as is.

2.5. Sensitivity analysis

We estimated the sensitivity of our models to variation of their parameters. In other words, we determined how much the progression parameters can be varied without losing good fit to the real data. To do so, two approaches were used.

First approach: each of the parameters related to the progression of TB ($b, \gamma, \delta, \mu_-, \mu_+$) was varied over a grid within wide limits, with all other parameters except the varied one remaining at the optimal values. Goodness-of-fit for each set of tested parameters was measured by coefficients of determination (R^2) of the real detection data by the model predictions. Since R^2 -s for the absolute numbers of detected TB cases (R^2_{Φ}) were usually very high, a more practically valuable measure was also used: R^2 for the fraction of bacillary cases among the detected ones ($R^2_{\Phi^+}$). The reason for this is that we want to “recreate” with our models not only [Fig. 3A](#), but also [Fig. 3B](#).

Second approach: we varied the same progression parameters and additionally the target disease length D_0 , but all other model parameters (including the detection rates) were re-optimised, i.e., a new best fit to the data was found with all parameters except the varied one being free variables in the optimisation process. Goodness-of-fit was measured just as in the first approach. The second method appears to give more practically meaningful estimations.

The results were presented as graphs of R^2 -s as functions of the varied parameter. In most cases, the optimised value of each parameter is close to the maximum point of R^2 -s. So, by choosing an arbitrary cut-off for R^2 -s for a “still good fit”, it is possible to determine the limits of the varied parameter. Furthermore, if R^2 -s fall off quickly around the optimal value of a parameter, the parameter is called “critical” for the model.

2.6. Adjusted serial model

As it will be shown in the Results, the pure serial model produces an unrealistically low estimation of self-cure and mortality rate in T_+ group μ_+ . At the same time, the sensitivity analysis will show that coefficients μ_- and μ_+ in the serial model can be varied in a wide range with practically no reduction in goodness-of-fit, provided other parameters are re-estimated accordingly. Thus, we can choose any value of μ_+ in the allowed range and declare it as the optimal parameter for the serial model, but there is no definitive argument to prefer one value of μ_+ over all others.

A very rough criterion can be constructed by targeting the total case-fatality rate (CFR): if we assume that all TB cases leaving the model from T are cured (i.e., μ_- represents only self-cure), and all TB cases leaving the model from T_+ in fact die from TB (i.e., μ_+ represents only mortality), then the fraction of those leaving from T_+ (denoted M_+ , see [Appendix A](#) and data in [Appendix B](#)) becomes the total CFR. Another rough estimate is 45% total CFR for untreated pulmonary TB: 70% CFR for

sputum-smear positive TB, 20% CFR for sputum-smear negative TB (Ragonnet et al., 2021; Tiemersma et al., 2011), and approximately 50%/50% split between sputum-smear positive and sputum-smear negative TB among the new cases. So, we can scan the results of the sensitivity analysis of the serial model with re-optimisation with various values of μ_- and μ_+ in the allowed range and choose a set of parameters that has M_+ as close to 45% as possible. This set is shown in the results as “adjusted serial model”.

The “adjusted serial model” can be viewed as a more plausible variant of the serial model. Although the criterion used to choose its parameters is extremely questionable (especially in ascribing self-cure and mortality exclusively to μ_- and μ_+), and so all other “allowed” parameter sets for the serial model can still be regarded as plausible.

Furthermore, the “adjusted serial model” should not be taken as a wider model of the progression of TB with clearly separated and quantified TB mortality and self-cure. It is just an extremely crude method to select a set of parameters of the serial model within the mathematically allowed range and with believable values.

3. Results

3.1. Basic fitting

We fitted three variants of the “basic” model to the data: the parallel, the serial, and the combined one. The fitting results are shown in Table 1 and Fig. 5 together with the results for the adjusted serial model. It can be easily seen that the combined model has effectively converged to the parallel one. The graphs of active detection rate $\varphi_a(t)$ are presented as a part of sensitivity analysis in Appendix A.

3.2. Sensitivity analysis without re-optimisation

The graphs of R_Φ^2 and R_+^2 as functions of the progression parameters are presented in Appendix A. For the parallel model and all varied parameters (b, μ_-, μ_+), the results without re-optimisation were similar: the fit to the absolute detection counts (R_Φ^2) remained good, although the fit to the more practically valuable T_+ fraction ($R_{+1\dots3}^2$) fell quickly when the parameters deviated from their optimal values. For the serial model without re-optimisation, parameters γ, δ , and μ_+ exhibited the same behaviour, although varying μ_- resulted in relatively quickly falling R_Φ^2 while R_+^2 and $R_{+1\dots3}^2$ remained nearly unchanged. Thus, without re-optimisation, significant variation of all parameters appeared to be detrimental to the fit of the models. This means that no parameter is redundant.

3.3. Sensitivity analysis with re-optimisation

With re-optimisation, the results were quite different. The fit of detection counts (measured by R_Φ^2 and $R_{\Phi1\dots3}^2$) was extremely good in all cases for both models.

For the parallel model, coefficient μ_+ was “critical” in the sense that even the smallest deviations from the optimal value did not let the model to fulfil the disease length requirement. Coefficient μ_- was found to be variable in a ~10–15% range from the optimal value, although variations to the low side tended to break the disease length requirement. Coefficient b was

Table 1

Parameters and diagnostic values of four variants of the “basic” mathematical model fitted to the TB detection data from Moscow, Russia, 2013–2018.

Symbol	Description	Parallel model	Serial model	Adjusted serial model	Combined model	Dimension
S_0	S value at $t = 0$	10^{8a}	10^{8a}	10^{8a}	10^{8a}	pers.
μ	General mortality	0.008074 ^a	0.008074 ^a	0.008074 ^a	0.008074 ^a	1/yr
λ	Infection-and-activation rate	0.0000525	0.0000459	0.0000443	0.0000525	1/yr
b	Fraction of incident cases instantly progressing to bacillary TB (T_+)	0.334	0 ^a	0 ^a	0.334	–
γ	T_- to T_+ progression rate	0 ^a	2.98	2.97	0	1/yr
δ	T_+ to T_- regression rate	0 ^a	3.17	2.85	0	1/yr
μ_-	T_- mortality + self-cure rate	0.876	0.644	0.350 ^a	0.876	1/yr
μ_+	T_+ mortality + self-cure rate	0.149	0.00209	0.316	0.149	1/yr
φ_{p-}	T_- passive detection rate	0.155	0.164	0.164	0.155	1/yr
φ_{p+}	T_+ passive detection rate	0.346	0.380	0.380	0.346	1/yr
R_Φ^2	Coefficient of determination for all detection fluxes $\Phi_{ij,k}$	99.79%	99.28%	99.28%	99.79%	–
$R_{\Phi1\dots3}^2$	Coefficient of determination for detection fluxes $\Phi_{ij,k}$ in time-bins $k = 1\dots3$	99.79%	99.51%	99.50%	99.79%	–
R_+^2	Coefficient of determination for the fraction of T_+ cases among detected	95.83%	73.40%	73.55%	95.82%	–
$R_{+1\dots3}^2$	Coefficient of determination for the fraction of T_+ cases in time-bins $k = 1\dots3$	95.35%	94.23%	94.16%	95.40%	–
D	Average TB duration	3.004	3.002	3.000	3.004	yr

^a = fixed value; TB = tuberculosis, T_+ = TB with bacterioexcretion, T_- = TB without bacterioexcretion.

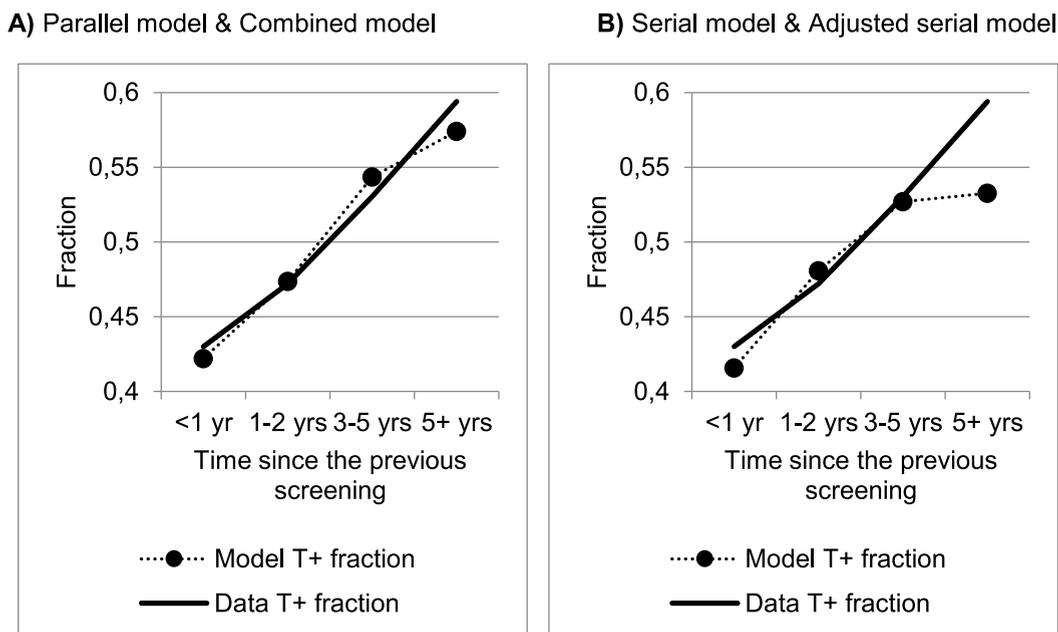


Fig. 5. Comparison of the fractions of bacillary TB cases (T+) in the data and in the models' best fits. **A)** Parallel and combined models (their predictions differ only in the 5th decimal digit, so one graph represents both of them). **B)** Serial model and adjusted serial model (their predictions are practically indistinguishable too).

variable within a ~10–15% range too. Variation of the target disease length D_0 in the 2 ... 4 years range did not significantly affect any goodness-of-fit metrics.

For the serial model, coefficients μ_- and μ_+ could be varied in a wide [0, 0.65 1/yr] range, with their sum remaining nearly constant. Coefficients γ and δ permitted big variation too, and increasing them from their optimal values improved $R_{+1...3}^2$ very slightly and worsened R_+^2 much more strongly. For the practical ranges for γ and δ , see [Appendix A](#). Variation of the target disease length D_0 , just as for the parallel model, did not affect goodness-of-fit metrics.

3.4. Active detection rate

The active detection rate $\varphi_a(t)$ had similar time-profiles for both models and all parameters with reasonably good fit to the data (see [Appendix A](#)). $\varphi_a(t)$ grew significantly from $t = 0$ to $t = 1$ yr, vanished by $t = 5$ yrs, and then grew again to $t = 10$ yrs.

3.5. Adjusted serial model

Among the results of the sensitivity analysis of the serial model with re-optimisation, with μ_- and μ_+ being in the range where the goodness-of-fit metrics are practically indistinguishable from the pure serial model, we chose a set of parameters that had M_+ as close to 45% as possible (see data in [Appendix B](#)). This set has fixed $\mu_- = 0.35$ 1/yr, and estimated $\mu_+ = 0.316$ 1/yr and $M_+ = 45.9\%$. The set is placed in [Table 1](#) as “adjusted serial model”.

The “adjusted” estimation of $\mu_+ = 0.316$ 1/yr looks much more realistic than the pure serial model's $\mu_+ = 0.00209$ 1/yr. So, because of the negligible difference in goodness-of-fit between the pure and the adjusted parameter sets of the serial model, we recommend to consider the adjusted one as the resulting estimate in this study.

4. Discussion

4.1. General

The starting point of this study was the need to estimate the parameters of progression between stages of active pulmonary TB for a modern population. The estimates would be useful in both epidemiologic models of TB and models that estimate case-detection quality from standard TB notification data ([Avilov et al., 2015](#)). Since we worked with the data from Moscow, Russia, we used the definitions of stages of TB that are standard in Russian healthcare. When compared to the recently published classifications of TB stages ([Drain et al., 2018](#); [Frascella et al., 2021](#); [Kendall et al., 2021](#)), Russian “non-bacillary TB” may be roughly matched to “minimal TB”, and Russian “bacillary TB” matches both “subclinical TB” and “clinical TB” from ([Frascella et al., 2021](#)). Under the definitions used in ([Drain et al., 2018](#); [Kendall et al., 2021](#)), “non-bacillary TB”

covers both “incipient TB” and a part of “subclinical TB”, while Russian “bacillary TB” constitutes another part of “subclinical TB” and “clinical TB”. So, our study was aimed at measuring the parameters of progression from minimal TB to subclinical and clinical one. For this reason, we did not try to build a comprehensive model of the natural history of TB infection and TB disease: we did not separately model the activation of TB infection (i.e., transition from a latent infection to an active TB disease) and disregarded the balance between self-cure and death of TB cases (we merged both in a single process of “termination of TB disease”).

Still, as compared to other works quantifying the activation of TB infection (Behr et al., 2018; Drain et al., 2018; Emery et al., 2021; Esmail et al., 2014) or estimating the case-fatality rate and the duration of TB disease (Ragonnet et al., 2021; Tiemersma et al., 2011), our study explores a novel area of quantitative parameters of the early stages of active pulmonary TB, thus contributing to the overarching goal of the comprehensive natural history model.

4.2. Modelling

The controversy of the parallel and serial “schemes” of mathematical models of the progression of TB exists almost as long as these models exist. We compared these approaches on the basis of fit to the modern real data on a relatively big number of cases. Although our definitions of the “early” and “advanced” types of active TB differed from the more prevalent ones (often “infectious” and “non-infectious TB” are defined as “sputum smear positive” and “sputum smear negative TB”), our modelling still added some new knowledge on the problem, especially with regard to the dynamics of the early stages of active TB.

Under our study design, it was not possible to model the progression of TB as a “stand-alone” phenomenon because TB detection data are always strongly affected by case-detection systems and methods. So, we used a combined mathematical model that comprised progression and case-detection sub-models. Thus, all judgements about the progression submodel implicitly included the assumption that the detection submodel is correct, which is unlikely to be absolutely true. The time-dependent submodel of active detection seems to be realistic enough, but the constant-rate passive detection submodel might be an oversimplification.

4.3. Fitting results

Both the parallel and serial models were capable of good fit to the real data in the first three fluorography time-bins, with the serial model being slightly worse in describing the more ambiguous fourth time-bin (“5+ years”).

The following observations can be made:

- The parallel model has a low rate of death and self-cure from T_+ group ($\mu_+ = 0.148$ 1/yr) as compared to T_- group ($\mu_- = 0.876$ 1/yr). In the serial model, the balance of μ_- and μ_+ can be arbitrarily chosen under the empirical condition $\mu_- + \mu_+ \approx 0.66$ 1/yr and with an appropriate change in other parameters. This enabled us to derive the “adjusted serial model” with more epidemiologically sound parameters than the straightforward fitting generated. But, in any case, the parallel model predicts higher overall death and self-cure rate of TB cases.
- The fraction of incident cases who under the parallel model instantly develop bacterioexcretion ($b = 0.334$) is not extreme and, hence, is plausible.
- The T_- to T_+ and T_+ to T_- rates in the serial model (γ , δ) are quite high and approximately equal (≈ 3 1/yr), which may conform to the well-known from the pre-chemotherapy era “wave-like” course of untreated TB.
- The passive detection rates in both models are similar and “reasonable” in that the rate for T_- cases (φ_{p-}) is less than a half of the one for more severe T_+ cases (φ_{p+}). The active detection rates $\varphi_a(t)$ have believable bell-like dynamics on the $t = 0 \dots 5$ yrs interval and are growing after that; the growth can be attributed to the highly selected nature of the patients in the “>5 yrs since the previous screening” bin and uncertain practical time boundaries of the bin. For this reason, we mostly disregarded the goodness-of-fit in the last time-bin when comparing different models.

So, at the first glance, both models look as plausible. Yet, there is a profound difference in mechanisms that the models employ to reproduce the increase in the fraction of bacillary (T_+) cases with the time since the previous screening (Fig. 3B). In the serial model, at low times t almost all active TB cases are T_- (because all they start as T_-), and they gradually percolate into the T_+ state, thus increasing T_+ proportion among the detected ones (until a balance between T_- and T_+ settles). In the parallel model, on the contrary, T_+ cases appear instantly, and only the drastic disbalance of death and self-cure rates (μ_- and μ_+) makes T_- cases to “die out” quickly, while the long-lived T_+ cases remain and so increase their proportion. Moreover, in the parallel model, the average duration of untreated T_+ TB is $1/\mu_+ = 6.7$ years, which appears to be unrealistic.

The sensitivity analysis showed that most parameters in both models can be varied to a certain extent depending on what level of goodness-of-fit is considered acceptable. Both models can easily accommodate to any target disease length D_0 in the 2 ... 4 yrs range with no reduction in goodness-of-fit.

4.4. Which model?

Neither type of modelling of the progression of active TB can be readily dismissed as “completely unplausible” on the basis of our model fitting.

The parallel model is somewhat better in approximation of the last “5+ years” time-bin (Table 1, Fig. 5). The combined model converged to the parallel one when was fitted to the data. On the other hand, the parallel model predicts unrealistic properties of untreated T_+ TB cases.

As for the serial model, there are pathophysiological (Dannenberg, 1999) and empirical (Okada et al., 2012) evidences that in reality TB progresses “from stage to stage”, gradually acquiring new properties like infectiousness. Besides that, the parameters of our adjusted serial model look plausible, and they can be adjusted further, given the discovered “hidden free variable” in the serial model. Yet, the serial model does not fit the last time-bin of the data well, and it predicts the relaxation of the T_+ fraction among the detected TB cases to a steady state, which appears to be absent in the available data.

So, we are inclined to interpret the results of our adjusted serial model as a rough estimation of the dynamics of early TB disease progression, with a caveat that the active case-detection submodel should be revised (to fix the “5+ years” time-bin problems) and this will likely alter the estimates. The parallel model can still be valid as an “instrumental” model that simply disregards a very short T_+ to T_+ progression period.

One possible explanation for the models of both structures fitting the real data equally well is, as we hypothesise, the heterogeneity of the real population the data come from. Earlier it was shown that there is a significant heterogeneity in resistance to TB infection among the staff of Russian TB hospitals (Romanyukha et al., 2009). Extrapolating this to the whole population of Moscow, we may assume that less resistant TB cases progress towards bacterioexcretion quickly after the onset of the disease (effectively following the parallel model), and other, more resistant cases progress slowly (as the serial model commands). So, the real data come from a mixed process. On the other hand, the convergence of the combined mathematical model (that is a mixture of both basic models) to the parallel one may indicate that considering bacterioexcretion status as the only measure of disease progression gives too little information, and inclusion of other patients’ parameters may be beneficial for reverse-engineering the natural history of TB.

4.5. Limitations

The results and conclusions of the study are limited by several factors:

- Because of the mass vaccination with BCG at birth, the average parameters of TB cases in Russia may be different from those in other populations. We included only adult (15+ years old) TB cases into our study, so that the influence of BCG was minimized. Still, some protection from BCG is possible, and so our estimations of the progression rates may be lower than average.
- Our model did not allow for age, sex, and other risk factors of the TB cases. Although many of these parameters were available in the database, we did not use them because the number of cases included into our study is too small to subdivide it by age or sex and carry out similar analysis in each of the subgroups with reasonable statistical stability. This may also limit the ability to extrapolate our findings to other populations with different age and sex structures.
- We used very simple linear ordinary differential equations to describe our model. A more flexible mathematical apparatus (with, at least, non-exponential distribution of time in a model group) could fit the data better.
- The case-detection submodel is likely oversimplified in favour of the simplicity of the differential equations. A more detailed analysis of the real detection process and its features might produce a better detection model.
- The study provides only point estimates of the parameters, but no confidence intervals or other measures of uncertainty. Since our current study is more of a “proof of the concept” of obtaining useful information from the fluorography data, our parameter estimates are not final. The “practical” model of the progression of the early stages of active TB will be much more complex, and it will worth developing the statistical apparatus to estimate the uncertainty of its parameters.

5. Conclusion

By utilising the data on the previous fluorographic screening of detected TB cases, the study demonstrated the possibility of inferring useful information from such type of data. The analysis entitled neither the parallel, nor the serial mathematical model of the progression of active pulmonary TB as a clear “winner”, but for both models the parameters were estimated. Yet, we slightly prefer the serial model of TB progression because it explains the growth of the fraction of bacillary TB cases with time in a more plausible way and, besides that, because of existing empirical evidences in its favour.

We used an oversimplified submodel of TB case detection and only very simple markovian two-group submodels of the progression of TB. More complex models – including those using more stages and more detailed information on the health status at detection and accounting for heterogeneity of patients – will likely better approximate the real data and enable deeper insights into the natural history of TB and modelling of the TB case-detection process.

For these reasons, our current conclusions and parameter estimates should be viewed as preliminary ones only. Despite that, this study revealed the very different mechanisms that the parallel and serial model of the progression of active TB employ to explain the growth of the fraction of bacillary TB with time.

Author contributions

KKA: Conceptualization, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. AAR: Conceptualization, Funding acquisition, Supervision, Writing – review & editing. EMB: Conceptualization. Data curation, Investigation, Writing – review & editing. SEB: Conceptualization, Supervision.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

KKA and AAR were supported by the Moscow Center of Fundamental and Applied Mathematics (Agreement 075-15-2019-1624 with the Ministry of Education and Science of the Russian Federation), <https://mathcenter.ru/en/>. EMB and SEB were supported by research project No. AAAA-A20-120062990040-1 "Development of a program for expanded access to tuberculosis control measures for the population of the city of Moscow: medical, social and economic justification" funded by the Department of Health of the Moscow Government, <https://mosgorzdrav.ru/en-US/index.html>.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.idm.2022.06.007>.

References

- Abubakar, I., Pimpin, L., Ariti, C., Beynon, R., Mangtani, P., Sterne, J., Fine, P., Smith, P., Lipman, M., Elliman, D., Watson, J., Drumright, L., Whiting, P., Vynnycky, E., & Rodrigues, L. C. (2013). Systematic review and meta-analysis of the current evidence on the duration of protection by bacillus Calmette-Guérin vaccination against tuberculosis. *Health Technology Assessment*, 17(37). <https://doi.org/10.3310/hta17370>
- Anderson, R. M., & May, R. M. (1992). *Infectious diseases of humans: Dynamics and control*. Oxford University Press, ISBN 978-0-19-854040-3.
- Avilov, K. K., Romanyukha, A. A., Belilovsky, E. M., & Borisov, S. E. (2019). Comparison of modeling schemes for natural course of pulmonary tuberculosis. *Mathematical Biology and Bioinformatics*, 14(2), 570–587. <https://doi.org/10.17537/2019.14.570> (in Russian).
- Avilov, K. K., Romanyukha, A. A., Borisov, S. E., Belilovsky, E. M., Nechaeva, O. B., & Karkach, A. S. (2015). An approach to estimating tuberculosis incidence and case detection rate from routine notification data. *International Journal of Tuberculosis & Lung Disease*, 19(3), 288–294. <https://doi.org/10.5588/ijtld.14.0317>
- Baltussen, R., Floyd, K., & Dye, C. (2005). Cost effectiveness analysis of strategies for tuberculosis control in developing countries. *BMJ*, 331(7529), 1364. <https://doi.org/10.1136/bmj.38645.660093.68>
- Barry, C. E., Boshoff, H. I., Dartois, V., Dick, T., Ehrh, S., Flynn, J., Schnappinger, D., Wilkinson, R. J., & Young, D. (2009). The spectrum of latent tuberculosis: Rethinking the biology and intervention strategies. *Nature Reviews Microbiology*, 7, 845–855. <https://doi.org/10.1038/nrmicro2236>
- Behr, M. A., Edelstein, P. H., & Ramakrishnan, L. (2018). Revisiting the timetable of tuberculosis. *BMJ*, 362. <https://doi.org/10.1136/bmj.k2738>. Article k2738.
- Blower, S. M., McLean, A. R., Porco, T. C., Small, P. M., Hopewell, P. C., Sanchez, M. A., & Moss, A. R. (1995). The intrinsic transmission dynamics of tuberculosis epidemics. *Nature Medicine*, 1(8), 815–821. <https://doi.org/10.1038/nm0895-815>
- Brogger, S. (1967). Systems analysis in tuberculosis control: A model. *American Review of Respiratory Disease*, 95(3), 419–434. <https://doi.org/10.1164/arrd.1967.95.3.419>
- Dannenberg, A. M., Jr. (1999). Pathophysiology: Basic aspects. In D. Schlossberg (Ed.), *Tuberculosis and non-tuberculous mycobacterial infections* (4th ed., pp. 17–47). WB Saunders.
- Davies, P. D. O., Gordon, S. B., & Davies, G. (Eds.). (2014). *Clinical tuberculosis* (5th ed.). CRC Press, ISBN 978-1444154344.
- Drain, P. K., Bajema, K. L., Dowdy, D., Dheda, K., Naidoo, K., Schumacher, S. G., Ma, S., Meermeier, E., Lewinsohn, D. M., & Sherman, D. R. (2018). Incipient and subclinical tuberculosis: A clinical review of early stages and progression of infection. *Clinical Microbiology Reviews*, 31. <https://doi.org/10.1128/CMR.00021-18>. Article e00021-18.
- Dye, C., Garnett, G. P., Sleeman, K., & Williams, B. G. (1998). Prospects for worldwide tuberculosis control under the WHO DOTS strategy. Directly observed short-course therapy. *Lancet*, 352(9144), 1886–1891. [https://doi.org/10.1016/S0140-6736\(98\)03199-7](https://doi.org/10.1016/S0140-6736(98)03199-7)
- Emery, J. C., Richards, A. S., Dale, K. D., McQuaid, C. F., White, R. G., Denholm, J. T., & Houben, R. M. G. J. (2021). Self-clearance of Mycobacterium tuberculosis infection: Implications for lifetime risk and population at-risk of tuberculosis disease. *Proceedings of the Royal Society B: Biological Sciences*, 288. <https://doi.org/10.1098/rspb.2020.1635>, 1943. Article 20201635.
- Esmail, H., Barry, C. E., Young, D. B., & Wilkinson, R. J. (2014). The ongoing challenge of latent tuberculosis. *Philosophical Transactions of the Royal Society B*, 369. <https://doi.org/10.1098/rstb.2013.0437>. Article 20130437.
- Frascella, B., Richards, A. S., Sossen, B., Emery, J. C., Odone, A., Law, I., Onozaki, I., Esmail, H., & Houben, R. M. G. J. (2021). Subclinical tuberculosis disease – a review and analysis of prevalence surveys to inform definitions, burden, associations, and screening methodology. *Clinical Infectious Diseases*, 73(3). <https://doi.org/10.1093/cid/ciaa1402>. Article e830–e841.
- Houben, R. M. G. J., Esmail, H., Emery, J. C., Joslyn, L. R., McQuaid, C. F., Menzies, N. A., Sanz, J., Shrestha, S., White, R. G., Yang, C., & Cobelens, F. (2019). Spotting the old foe-revisiting the case definition for TB. *Lancet Respiratory Medicine*, 7(3), 199–201. [https://doi.org/10.1016/S2213-2600\(19\)30038-4](https://doi.org/10.1016/S2213-2600(19)30038-4)
- Kendall, E. A., Shrestha, S., & Dowdy, D. W. (2021). The epidemiological importance of subclinical tuberculosis. A critical reappraisal. *American Journal of Respiratory and Critical Care Medicine*, 203(2), 168–174. <https://doi.org/10.1164/rccm.202006-2394PP>
- Murray, C. J. L., & Salomon, J. A. (1998). Modeling the impact of global tuberculosis control strategies. *Proceedings of the National Academy of Sciences of the United States of America*, 95(23), 13881–13886. <https://doi.org/10.1073/pnas.95.23.13881>

- Okada, K., Onozaki, I., Yamada, N., Yoshiyama, T., Miura, T., Saint, S., Peou, S., & Mao, T. E. (2012). Epidemiological impact of mass tuberculosis screening: A 2-year follow-up after a national prevalence survey. *International Journal of Tuberculosis & Lung Disease*, 16(12), 1619–1624. <https://doi.org/10.5588/ijtld.12.0201>
- Perelman, M. I., Marchuk, G. I., Borisov, S. E., Kazennykh, B. Y., Avilov, K. K., Karkach, A. S., & Romanyukha, A. A. (2004). Tuberculosis epidemiology in Russia: The mathematical model and data analysis. *Russian Journal of Numerical Analysis and Mathematical Modelling*, 19(4), 305–314. <https://doi.org/10.1515/rnam.2004.19.4.305>
- Porta, M. (Ed.). (2014). *A dictionary of epidemiology* (6th ed.). Oxford University Press, ISBN 978-0-19-997672-0.
- Ragonnet, R., Flegg, J. A., Brilleman, S. L., Tiemersma, E. W., Melsew, Y. A., McBryde, E. S., & Trauer, J. M. (2021). Revisiting the natural history of pulmonary tuberculosis: A bayesian estimation of natural recovery and mortality rates. *Clinical Infectious Diseases*, 73(1), e88–e96. <https://doi.org/10.1093/cid/ciaa602>
- Romanyukha, A. A., Litvinov, I. S., & Sannikova, T. E. (2009). Modelling the dynamics of specific immunity under protracted antigenic load. *Russian Journal of Numerical Analysis and Mathematical Modelling*, 24(4), 361–375. <https://doi.org/10.1515/RJNAMM.2009.022>
- Tiemersma, E. W., van der Werf, M. J., Borgdorff, M. W., Williams, B. G., & Nagelkerke, N. J. D. (2011). Natural history of tuberculosis: Duration and fatality of untreated pulmonary tuberculosis in HIV-negative patients: A systematic review. *PLoS One*, 6(4), Article e17601. <https://doi.org/10.1371/journal.pone.0017601>
- Waalder, 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 and the Nation's Health*, 52(6), 1002–1013. <https://doi.org/10.2105/ajph.52.6.1002>
- World Health Organization. (2020). *Global tuberculosis report 2020*. World Health Organization, ISBN 978-92-4-001314-8.
- Yaesoubi, R., & Cohen, T. (2013). Identifying dynamic tuberculosis casefinding policies for HIV/TB coepidemics. *Proceedings of the National Academy of Sciences of the United States of America*, 110(23), 9457–9462. <https://doi.org/10.1073/pnas.1218770110>