Epidemics 21 (2017) 39–47

Contents lists available at ScienceDirect

Epidemics

journal homepage: www.elsevier.com/locate/epidemics

# Optimally capturing latency dynamics in models of tuberculosis transmission

Romain Ragonnet[a,b,*], James M. Trauer[a,c,d], Nick Scott[b,c], Michael T. Meehan[e], Justin T. Denholm[a,d,f], Emma S. McBryde[a,e]

[a] Faculty of Medicine, Dentistry and Health Sciences, University of Melbourne, Australia
[b] Burnet Institute, Australia
[c] School of Population Health and Preventive Medicine, Monash University, Australia
[d] Victorian Tuberculosis Program, Melbourne, Australia
[e] Australian Institute of Tropical Health and Medicine, James Cook University, Australia
[f] Royal Melbourne Hospital, Melbourne, Australia

ARTICLE INFO

*Article history:*
Received 13 February 2017
Received in revised form 30 May 2017
Accepted 14 June 2017
Available online 16 June 2017

*Keywords:*
Tuberculosis latency
Mathematical modelling
Risk of disease activation
Parameter estimation

ABSTRACT

Although different structures are used in modern tuberculosis (TB) models to simulate TB latency, it remains unclear whether they are all capable of reproducing the particular activation dynamics empirically observed. We aimed to determine which of these structures replicate the dynamics of progression accurately. We reviewed 88 TB-modelling articles and classified them according to the latency structure employed. We then fitted these different models to the activation dynamics observed from 1352 infected contacts diagnosed in Victoria (Australia) and Amsterdam (Netherlands) to obtain parameter estimates. Six different model structures were identified, of which only those incorporating two latency compartments were capable of reproducing the activation dynamics empirically observed. We found important differences in parameter estimates by age. We also observed marked differences between our estimates and the parameter values used in many previous models. In particular, when two successive latency phases are considered, the first period should have a duration that is much shorter than that used in previous studies. In conclusion, structures incorporating two latency compartments and age-stratification should be employed to accurately replicate the dynamics of TB latency. We provide a catalogue of parameter values and an approach to parameter estimation from empiric data for calibration of future TB-models.



## 1. Introduction

Tuberculosis (TB) is a major health issue with 10.4 million active cases and 1.8 million deaths worldwide in 2015 (WHO, 2016). Furthermore, around one quarter of the world's population is estimated to be infected with TB (Houben and Dodd, 2016), representing a huge reservoir of potential disease. Accordingly, fully understanding latent TB infection is crucial for assessing the future epidemic trajectory and designing effective TB control policies. Despite this, much reinfection occurs in high incidence cohorts, hampering accurate estimation of latency dynamics. Therefore insights into the activation dynamics following a single infection episode of *Mycobacterium tuberculosis* provided by recent studies in very low transmission settings are particularly valuable (Trauer et al., 2016a; Sloot et al., 2014). These works provide detailed information on patterns of activation, highlighting that most active cases occur within the first few months of infection.

Mathematical modelling has informed TB control programs by simulating interventions, or by explaining the mechanisms underlying observed epidemiological trends (Vynnycky and Fine, 1997; Gomes et al., 2004; Castillo-Chavez and Feng, 1997; Abu-Raddad et al., 2009; Cohen et al., 2008; Dye, 2012; Menzies et al., 2012; Trauer et al., 2016b), yet little is known about whether such modelling has been able to capture latency dynamics accurately. In the past, TB models have been constructed to capture the lifelong probability of disease and, although some models allowed for marked differences between the early and late dynamics of infection (Dowdy et al., 2013; Lin et al., 2011; Aparicio and Castillo-Chavez, 2009), estimates for the associated parameters have not been fit closely to longitudinal data. Despite this, it has been shown

* Corresponding author at: 85 Commercial Road, Burnet Institute, Melbourne, VIC 3003, Australia.
*E-mail address:* romain.ragonnet@burnet.edu.au (R. Ragonnet).

http://dx.doi.org/10.1016/j.epidem.2017.06.002

that when modelling infectious diseases, it is critical to employ appropriate distributions of latent periods (Wearing et al., 2005). Focusing on emerging infectious diseases, Wallinga and Lipsitch further demonstrated that capturing the mean of the generation times is not sufficient to characterise transmission accurately, as the shape of the distribution of the generation intervals also plays a critical role in infection dynamics (Wallinga and Lipsitch, 2007). Although TB is an ancient disease, its epidemiology is continuously evolving. In particular, changes in TB epidemiology in response to emerging phenomena, such as introduction of drug-resistant forms of TB or stronger control programs, are likely to affect the shape of the generation time distribution. Therefore, the recent detailed characterisation of TB activation dynamics represent a valuable opportunity to review and improve modelling practices for the simulation of TB latency.

Compartmental dynamic transmission models – the most common type of TB mathematical model – simulate TB latency with various levels of complexity. While some modellers employ a single latency compartment that precedes the active disease compartment (Colijn et al., 2008; Blower and Chou, 2004), others incorporate a second latency compartment in order to capture two different rates of progression from latent infection to active disease (Hill et al., 2012; Cohen et al., 2006; Trauer et al., 2014). When two latency compartments are incorporated, they can either be positioned in series or in parallel, involving different underlying assumptions regarding the progression pathways to active disease. First, the serial structure implies that newly infected individuals remain at high risk of disease during the initial phase and then, if TB activation has not occurred, they transition to another compartment where their risk of developing active TB is reduced. By contrast, with a parallel compartmental structure, the underlying assumption is that a proportion of infected individuals belong to a high risk category, while the remainder are at lower risk of TB disease. While TB modelling has been used extensively for over 40 years, it remains unclear which of these structures are best adapted to the natural history of TB.

In this study, we aim to determine the most appropriate model structures to simulate TB latency and provide estimates for the parameters associated with these structures across different age categories. We use the distribution of the estimated times from infection to TB activation in 1352 infected contacts of individuals with active pulmonary TB from Victoria (Australia) and Amsterdam (Netherlands) to calibrate the latency structures of different candidate models to the dynamics observed in the data.

## 2. Methods

### 2.1. Literature review

Our search was based on the literature review of mathematical and economic TB modelling articles provided by the TB Modelling and Analysis Consortium, available online at http://tb-mac.org/Resources/Resource/4 (see Appendix in Supplementary file for more details). From this database we identified all 88 publications reporting the use of a deterministic compartmental transmission dynamic model. All selected papers were reviewed independently by two authors (RR, JMT) who classified the manuscripts according to the structure used to model TB latency. These two independent investigations led to the same classification which is presented in the Appendix in Supplementary file (Table S1).

### 2.2. Analytical solution

For each latency structure found in the literature, we associated a basic dynamic model comprised of the latency structure in combination with compartments representing susceptibility to infection and active disease. We then found analytical solutions for the TB activation dynamics corresponding to each model. Namely, considering that individuals were infected at time $t = 0$, we determined the proportion $I(t)$ of infected individuals that had developed active TB after each time $t$ ($t \geq 0$). Analytical expressions are also presented for the total proportion of infected individuals progressing to active disease, obtained by calculating the limit of $I(t)$ as $t$ approaches plus infinity. The detailed method used to obtain the analytic solutions is described in the Appendix in Supplementary file.

### 2.3. Data used to calibrate the models

The models described above were calibrated to individual data on close contacts of individuals with active pulmonary TB notified in the Australian state of Victoria from January 2005 to December 2013. These data are derived from a very low endemic setting and were described in detail by Trauer et al. They consist of 613 infected contacts of whom 67 (10.9%) developed active TB during the study period (Trauer et al., 2016a). To enhance our dataset, we also used the published data on close contacts of pulmonary TB patients from Amsterdam (Netherlands) notified between 2002 and 2011, as reported by Sloot et al. (Sloot et al., 2014). These data include 739 infected individuals, of whom 71 (9.6%) developed active TB. The detailed approaches used to determine both dates of infection and activation in individuals in the two studies are presented in the respective manuscripts. The activation times measured in these data were used to calibrate the different models. In order to validate our approach involving merging of two datasets, we present a comparison of the estimates obtained from the separate fittings to the two datasets (Appendix in Supplementary file Section 8.2). The approach used to extract data from Sloot and colleagues' article is described in detail in the Appendix in Supplementary file, along with a validation analysis of the extraction method while the distribution of the times to activation measured in the two datasets (Victoria and Amsterdam) is presented in Fig. S7 (Appendix in Supplementary file).

Trauer et al. also proposed an imputation method which takes into account the censorship for migration, death, and preventive treatment (Trauer et al., 2016a). We used this approach, which is associated with higher estimates concerning the risk of TB activation, in a supplementary analysis.

### 2.4. Model fitting

Model fitting to data was made by building the survival likelihood defined as follows. For a given model associated with a given set of parameters, $\theta$, we obtain an analytical survival function $S_\theta(t)$ which represents the probability that activation has not occurred yet at time $t$ given that infection occurred at $t = 0$. This function is associated with a hazard function $\lambda_\theta(t)$ defined by $\lambda_\theta(t) = -S'_\theta(t)/S_\theta(t)$, characterising the chance that progression to active TB occurs at precisely time $t$, given survival up to that time. Then, for each infected case $i$ of our dataset, for whom $t_i$ designates the time of either TB activation or end of follow-up, we define an individual likelihood component by $\mathcal{L}_{\theta,i} = S_\theta(t_i)$ if the case was not known to develop active TB; and $\mathcal{L}_{\theta,i} = \lambda_\theta(t_i) \times S_\theta(t_i)$ if the case effectively activated TB at time $t_i$. Finally, we aim to maximise the multi-dimensional likelihood obtained by multiplying all the individual likelihood components together: $L_\theta = \prod_i \mathcal{L}_{\theta,i}$. This problem is equivalent to maximising the following log-likelihood that we define as the fitting score: $FS_\theta = \sum_i \log(\mathcal{L}_{\theta,i})$.

Another fitting method was used for validation and when the data did not allow for the survival likelihood to be utilised. Specifically, a least squares optimisation was performed to minimise the

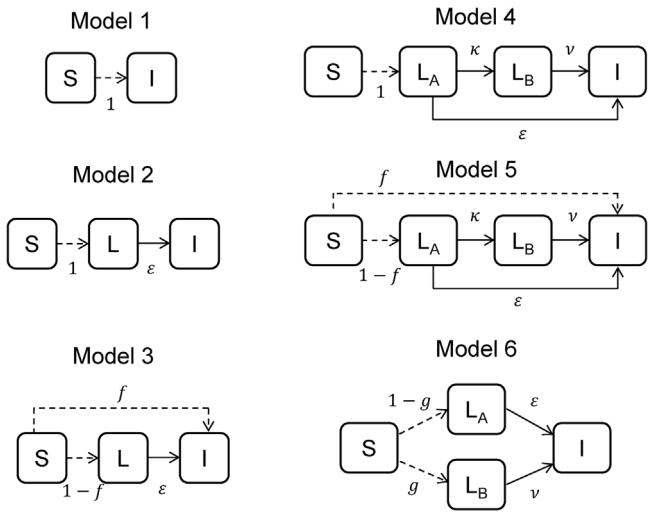


**Fig. 1.** Representation of the different model structures. Solid lines represent progressive transitions between compartments that are parameterised with rates, while dashed lines represent instantaneous transitions between compartments that are associated with proportions. The flows related to natural death are not represented and are associated with the natural mortality $\mu$.

distance between the survival curves generated by the data and by the model. The time-points considered when performing this optimisation were equally spaced by one day and ranged between 0 and the maximal time $t_i$ measured in the data.

The two methods introduced above are conceptually different. In effect, the survival likelihood approach searches for the set of model parameter values that maximises the probability of observing the data, given a particular model and a particular parameter set. In contrast, the least squares optimisation is based on a measure of a distance between two curves: the activation curve observed from the data and the one produced by a model. Although these two approaches are fundamentally distinct, they are both forms of optimisation applied to the model's parameter values. This optimisation is based on an iterative method which allows the parameters to vary in a multi-dimensional space, from which we retain only the parameter set that yields the optimal measure, i.e. the greatest likelihood (for the probabilistic approach) or the smallest distance (for the least squares method). This process was obtained using R v.3.3 and its incorporated function "constrOptim".

Model fitting was performed separately using three age categories ("<5 years old", "5 to 14 years old" and "≥15 years old") in addition to a pooled analysis including the whole population. In order to fully explore the parameter spaces and to report all acceptable parameter sets, we used a Metropolis-Hastings algorithm presented in detail in the Appendix in Supplementary file.

## 3. Results

### 3.1. Literature review

From our review of 88 publications related to TB modelling, we found that six different compartmental structures had been employed to model TB latency and these form the basis of the following analysis. The six models are represented in Fig. 1, along with the labels used for the associated parameters. The level of complexity ranges from a model incorporating no latency compartment (Model 1) to models based on two latency compartments positioned either in series (Models 4 and 5) or in parallel (Model 6). Some structures incorporate a bypass from the susceptible compartment to the active disease compartment, allowing for consideration of instantaneous activation of TB disease after exposure (Models 1, 3 and 5). Table S1 (Appendix in Supplementary file) reports our classification of the 88 reviewed studies according to the latency structure employed.

### 3.2. Model calibration

Fig. 2 presents the dynamics of TB activation obtained from each model when optimally calibrated to our data (see Methods paragraph for fitting approach). Only the models incorporating two latency compartments (Models 4, 5 and 6) suitably replicate the dynamics of TB activation. Model 1, which is not shown, does not contain any free parameters and assumes that all infected patients progress to active disease immediately after exposure. This is not compatible with our data nor our knowledge of the lifelong risk of TB disease in infected individuals (Trauer et al., 2016a; Sloot et al., 2014). Both Models 2 and 3 produce unreasonably poor fits to the data. In the Appendix in Supplementary file we show that the equations describing Models 2 and 3 only involve a single exponential function, which is not sufficient to replicate the two distinct patterns observed in the dynamics of activation—a high risk of disease activation over the first few months, followed by a dramatically lower risk in a second phase. In contrast, in Models 4, 5 and 6, which include two latency compartments, the activation dynamics are driven by two exponential components that are associated with two independent growth rates, leading to accurate replication of the patterns empirically observed for each age category. The fitting scores (see Fig. 2) for Models 4, 5 and 6 indicate that none provides a significantly better fit than the others; however, Model 5 presents a higher level of complexity than Models 4 and 6, as it involves estimating one additional parameter. Since the inclusion of this parameter did not improve model fitting, we consider that it is unnecessary to employ this structure. This proposition is strongly supported by a more detailed analysis of Model 5 where different approaches all support that the additional parameter does not improve fitting (see Appendix in Supplementary file). Accordingly, the remainder of our analysis relates only to Models 4 and 6.

Table 1 presents the parameter values (with units of days$^{-1}$) obtained from the calibration of Models 4 and 6 to the data. We observe that the rate of re-activation ($\nu$) is much lower in the age category "<5 years old" than in the other age categories for both Model 4 and Model 6. In contrast, the rate of rapid progression is higher in the "<5 years old" category than in the other age categories when considering Model 4; and when considering Model 6, the proportion of infected individuals experiencing a high risk of TB disease ($1-g$) is higher for children than for the other categories (35% for "<5 y.o.", 19% for "5 to 14 y.o.", 5% for "15 y.o. and more", and 9% for all ages together).

The lifelong risk of TB activation in infected individuals can be calculated analytically for the different models (see Appendix in Supplementary file, p.4). The parameter values reported in Table 1 for Models 4 and 6 correspond to total proportions of activation ranging between 10% ("15 y.o. and more") and 32% ("< 5 y.o.").

Given the very low values observed for the rate of reactivation ($\nu$) in the "<5 years old" category, we undertook an additional analysis in order to explore the possibility of the absence of late reactivation in this category. In this analysis, Models 4 and 6 were fitted to the data under the constraint that $\nu = 0$, with both models found to perform equally well in absence of reactivation. Fig. 3 presents the corresponding simplified models as well as the best fits obtained. A further exploration of the contribution of endogenous reactivation versus primary activation is presented in the Appendix in Supplementary file (Section 3.) for Model 4 and for the different age categories. This analysis demonstrates that endogenous reactivation contributes very little to the burden of active TB, especially in young individuals ("<5 years old" and "5 to 14 y.o."). Although this

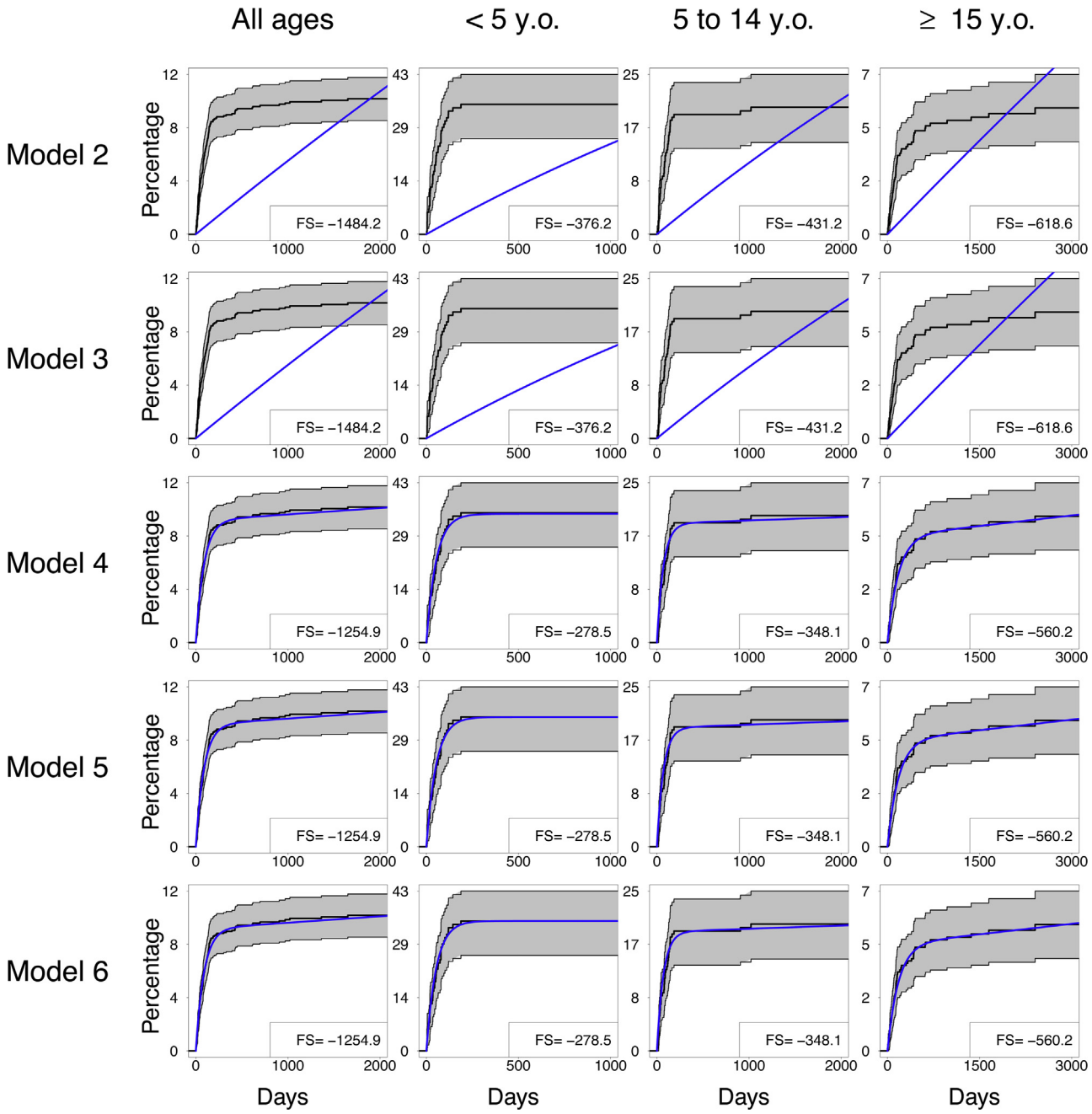


**Fig. 2.** Calibrations obtained with the different models for the percentage of active TB among infected individuals over time since infection. The black lines and grey shade represent the estimates (central and 95% CI) obtained from the Kaplan-Meier analysis of our data. The blue line represents the percentage of active TB over time obtained from the different models, when optimally calibrated with the survival likelihood method. The x-scales were chosen in order to allow for a decent visualisation of the early stages of infection and do not cover the entire time windows corresponding to the dataset. Fitting was realised by maximising the fitting score (FS). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

contribution is more substantial in the "≥15 years old" category, we estimate that only 1% of infected individuals of this category would have progressed to active disease transitioning from compartment $L_B$ after five or less years from infection (using Eq. (23) in Appendix in Supplementary file).

Our analysis of the analytical solutions associated with Model 4 and Model 6 demonstrated that the two expressions coincide if the parameter values of the two models satisfy the following relationships:

$$\varepsilon_6 = \varepsilon_4 + \kappa_4\nu_6 = \nu_4 g_6 = \frac{\kappa_4}{\kappa_4 + \varepsilon_4 - \nu_4}$$

where the subscripts indicate to which of Models 4 or 6 the parameters apply. This finding indicates that the dynamics of activation simulated by the two models are identical, differing only in the value of the parameter values that should be applied.

### 3.3. Probability distribution of parameters

We used a Bayesian framework to infer the posterior distributions for the parameters in the model. Uniform priors were used and tools of Bayesian inference applied including Markov Chain Monte-Carlo exploration using the Metropolis-Hastings acceptance algorithm. The resulting posterior distributions for the

**Table 1**
Parameter estimates. Calibration issued from the survival likelihood maximisation applied to the merged dataset (Victoria and Amsterdam data). Point estimates correspond to the parameters maximising the likelihood while values into brackets indicate the narrowest interval containing 95% of the accepted values during the Metropolis-Hastings simulation. Rates are presented as daily values.

| | | $\varepsilon$ | $\kappa$ | $\nu$ | $g$ |
|---|---|---|---|---|---|
| Model 4 | All ages | $1.1e^{-3}$ ($8.4e^{-4}$–$1.5e^{-3}$) | $1.0e^{-2}$ ($8.5e^{-3}$ $1.4e^{-2}$) | $5.5e^{-6}$ ($2.5e^{-6}$ $1.1e^{-5}$) | – |
| | <5 y.o. | $6.6e^{-3}$ ($4.4e^{-3}$ $9.5e^{-3}$) | $1.2e^{-2}$ ($8.5e^{-3}$ $1.8e^{-2}$) | $1.9e^{-11}$ ($5.0e^{-9}$ $1.6e^{-5}$) | – |
| | 5–14 y.o | $2.7e^{-3}$ ($1.7e^{-3}$ $3.9e^{-3}$) | $1.2e^{-2}$ ($7.9e^{-3}$ $1.6e^{-2}$) | $6.4e^{-6}$ ($6.7e^{-7}$ $1.9e^{-5}$) | – |
| | ≥15 y.o | $2.7e^{-4}$ ($1.6e^{-4}$ $5.1e^{-4}$) | $5.4e^{-3}$ ($3.5e^{-3}$ $1.1e^{-2}$) | $3.3e^{-6}$ ($1.9e^{-6}$ $1.0e^{-5}$) | – |
| Model 6 | All ages | $1.1e^{-2}$ ($9.2e^{-3}$ $1.5e^{-2}$) | – | $5.5e^{-6}$ ($3.4e^{-6}$ $1.0e^{-5}$) | 0.91 (0.89 0.93) |
| | <5 y.o. | $1.9e^{-2}$ ($1.2e^{-2}$ $2.5e^{-2}$) | – | $3.4e^{-11}$ ($2.7e^{-9}$ $2.0e^{-5}$) | 0.65 (0.55 0.73) |
| | 5–14 y.o | $1.4e^{-2}$ ($9.4e^{-3}$ $1.8e^{-2}$) | – | $6.4e^{-6}$ ($8.0e^{-7}$ $2.2e^{-5}$) | 0.81 (0.75 0.86) |
| | ≥15 y.o | $5.6e^{-3}$ ($3.8e^{-3}$ $9.6e^{-3}$) | – | $3.3e^{-6}$ ($7.3e^{-7}$ $9.3e^{-6}$) | 0.95 (0.94 0.97) |

Parameter estimates. Calibration issued from the survival likelihood maximisation applied to the merged dataset (Victoria and Amsterdam data). Point estimates correspond to the parameters maximising the likelihood while values into brackets indicate the narrowest interval containing 95% of the accepted values during the Metropolis-Hastings simulation. Rates are presented as daily values.

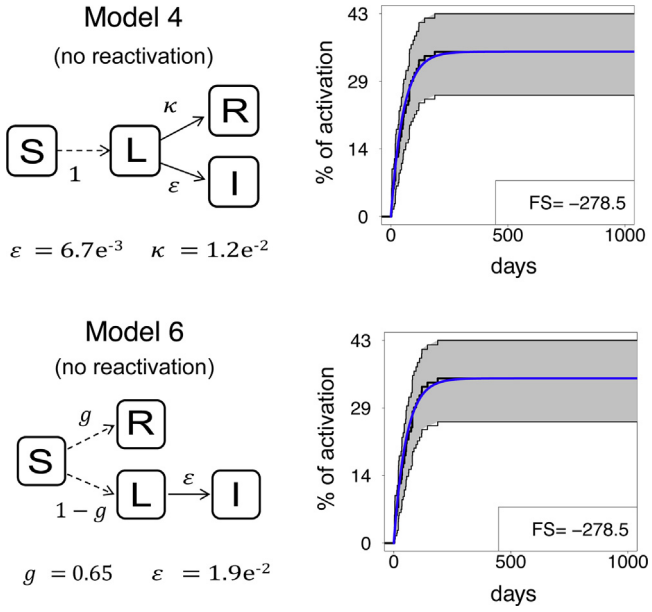


**Fig. 3.** Simplified model structures adapted to simulate TB latency in young children (<5 years old) Here, it is assumed that progression from the previous compartments $L_B$ of Model 4 and Model 6 to active disease $I$ cannot occur. The compartment $L_B$ therefore becomes a protected state which is labelled $R$ in this illustration. Fitting was realised by maximising the fitting score (FS).

different parameters of Models 4 and 6 are presented (Table 1). Proposed statistical distributions that could be used to generate similar sets of parameters are available in the Appendix in Supplementary file, along with the posterior distributions obtained for all parameters. The acceptable ranges of values for the parameter $\nu$ are wide for the category "<5 y.o.", due to the small number of reactivation cases in our dataset. However, our analysis showed that the rate of reactivation is limited by an upper bound of around $2.3e^{-5}$ in all categories, which represents a relatively small annual risk of re-activation of 0.8%.

By analysing the distributions of each of the retained parameter sets in pairs, we observed a collinearity between the parameters $\kappa$ and $\varepsilon$ for Model 4. This correlation (represented in Fig. 4) suggests that when individuals are assumed to stabilise infection more rapidly ($\kappa$ increases), the rate of rapid progression to TB disease ($\varepsilon$) tends to increase to compensate. This affine relationship between parameters $\kappa$ and $\varepsilon$ is also described through a formal mathematical analysis (Appendix in Supplementary file, Section 10.1.2.).

Another result provided by the Metropolis-Hastings simulation was the distribution of the average duration spent in the first latency compartment for Model 4. This quantity was obtained via the formula $\frac{1}{\kappa+\varepsilon+\mu}$ where $\mu$ represents the natural death rate. Fig. 5 presents the distribution of these durations for the different age categories. The average duration spent in the first latency compartment is estimated at 82 days for all age categories combined, this duration being shorter for children (52 days for "<5 y.o." category) than for older individuals (70 days for "5 to 14 y.o." and 146 days for "≥15 y.o.").

### 3.4. Validation and sensitivity analyses

Our findings were consistent when we employed an alternate fitting method (least squares minimisation), performed separate fits for each dataset (Victoria and Amsterdam) instead of the single merged dataset, or used alternate extraction methods for the data of Sloot et al. When fitting Model 3 using least squares minimisation, we obtained a model calibration that differed from that obtained with the survival likelihood method (See Appendix in Supplementary file, Fig. S14). However, while this second calibration allowed for a better simulation of the late stages of latency compared to the baseline fitting method (Fig. 2), it completely failed to capture the early dynamics of activation. More details about the different sensitivity analyses are available in the Appendix in Supplementary file.

Finally, by using imputed data as described in Trauer et al. accounting for the censorship for migration, death, and preventive treatment and therefore associated with higher estimates for the risk of TB disease (Trauer et al., 2016a), our conclusions regarding optimal model selection remained unchanged, while we obtained slightly different optimised parameter values (see Appendix in Supplementary file). In particular, in Model 4, imputation led to a slight increase in the rate of rapid progression ($\varepsilon$) combined with a slight reduction in the rate of transition towards the late latency compartment ($\kappa$). Concerning Model 6, the best calibration to imputed data was obtained with a higher proportion of infected individuals transitioning to the high risk compartment ($1-g$) while the rate of rapid progression ($\varepsilon$) was slightly reduced.

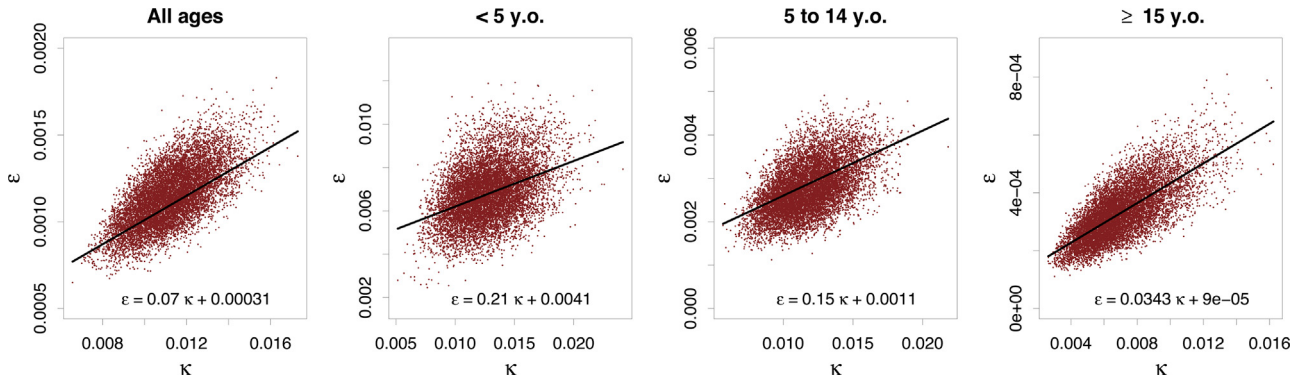


**Fig. 4.** Representation of the collinearity observed between the parameters $\kappa$ and $\varepsilon$ for Model 4. The red dots represent the 10,000 accepted parameter sets obtained from the Metropolis-Hastings simulation. The black line represents the affine model approximating the data with a least square minimisation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

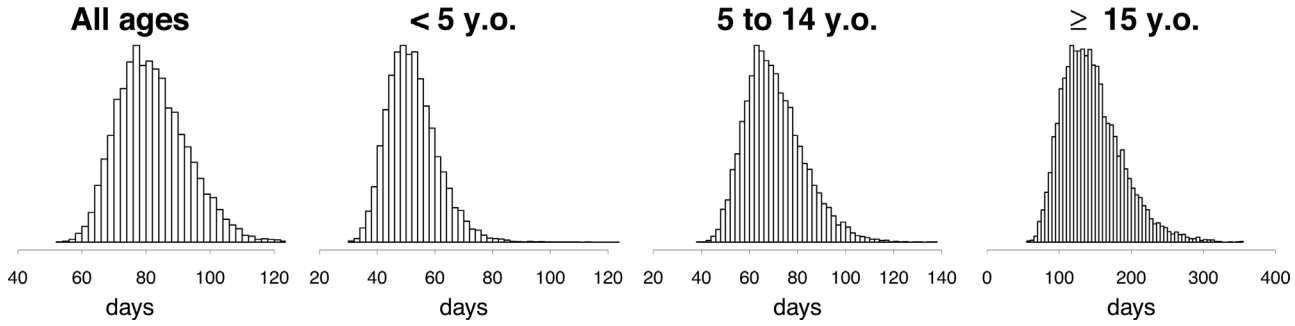


**Fig. 5.** Distribution of the times spent in the first latency compartment $L_A$ for model 4. Distribution associated with 10,000 accepted parameters sets from the Metropolis-Hastings simulation.

### 3.5. Comparison of our results with previous works

We reviewed the parameter values that had been employed in the previous studies incorporating the latency structures of Model 4 and Model 6 and compared these values to our estimates (see Appendix in Supplementary file, Section 9). Concerning Model 4, we found that the rate of progression from compartment $L_A$ to compartment $L_B$ ($\kappa$) was considerably lower than our estimate in all previous studies, which typically assume long periods spent in early latency (2–5 years). Similarly, the rate of fast progression to active TB ($\varepsilon$) used in the existing literature was much lower than our estimate. For Model 6, the proportion of slow progressors ($g$) found in previous works was close to our estimate, whereas the rate of fast progression ($\varepsilon$) had been markedly underestimated. Finally, while the rate of slow progression to active TB ($\nu$) was generally underestimated in studies incorporating either structure (Models 4 or 6), two studies used point estimates that fall inside of the 95% CI we report, and nine other studies used parameter ranges overlapping our 95% CI.

### 3.6. Rapid estimation of the parameters for future works

A theoretical analysis of the equations governing the dynamic models allowed us to develop a method to estimate rapidly the different parameters of Models 4 and 6 from any dataset. This approach involves simple graphical measurements performed on the curve representing the proportion of active cases among infected individuals over time from infection (denoted $\Gamma$), such as the curves represented in Fig. 2. The detailed description and demonstration of the method is available in the Appendix in Supplementary file and the main results are presented here. The profile of $\Gamma$ can be decomposed into two phases (see Fig. 2 and detailed description in Appendix in Supplementary file). The method consists of first measuring the initial slope of $\Gamma$ (denoted $s_0$) as well as the characteristics of a tangent to $\Gamma$ on a point situated at the beginning of the second rise phase (slope denoted $s$, and y-intercept denoted $y_0$). Then, estimates for the different parameters ($\hat{\varepsilon}$, $\hat{\kappa}$, $\hat{\nu}$ and $\hat{g}$) can be obtained by using the three measures ($s_0$, $s$ and $y_0$) as follows:

$$\textit{For Model 4}: \hat{\varepsilon} = s_0\hat{\kappa} = \frac{\hat{\varepsilon}}{y_0} - \hat{\varepsilon} - \mu\hat{\nu} = \frac{s\left(\hat{\kappa}+\hat{\varepsilon}+\mu\right)}{\hat{\kappa}}$$

$$\textit{For Model 6}: \hat{\varepsilon} = \frac{s_0}{y_0} - \mu\hat{\nu} = \frac{s_0 s}{s_0 - y_0\left(s_0+\mu\right)}\hat{g} = \frac{\hat{\varepsilon}-s_0}{\hat{\varepsilon}-\hat{\nu}}$$

where $\mu$ designates the natural mortality rate.

## 4. Discussion

### 4.1. Main contributions of the study

In this study, we determine the most appropriate model structures for accurately simulating TB latency. We demonstrate that of the structures employed in the past, only those that incorporate two compartments for latent infection are able to reproduce the specific dynamics of TB activation. Such approaches therefore involve two different levels for the rate of activation, allowing for more TB cases to occur after recent infection than through late reactivation. This work also provides a detailed and flexible catalogue of parameter values associated with the retained model structures, that was validated by the use of two independent fitting methods leading to similar estimates and that highlighted marked gaps with parameter values employed in previous works. This study may become a refer-

ence for calibration of future TB models and our approach involving estimations with and without age stratification will allow our findings to be directly applied whether models are age-structured or not.

### 4.2. Serial versus parallel structure

We demonstrate that the two compartments required to model TB latency could be placed either in series or in parallel, and would lead to identical activation dynamics. The difficulty in making this distinction between the two models is unfortunate because they reflect two alternate biological mechanisms that could explain the higher burden of disease observed after recent infection, each leading to different recommendations for TB prevention strategies. First, the serial structure models a decreasing risk of activation over time in every individual, indicating that preventive care should be targeted at the most recent infections as a priority through interventions such as contact tracing. On the other hand, the parallel structure suggests that individual predispositions may make some infected individuals more or less likely to develop disease, regardless of the time from infection. If this second scenario was verified with epidemiological data, these results would suggest that identifying and finding priority populations could dramatically enhance efficiency of interventions. Some factors such as HIV-infection, diabetes or smoking are already known to enhance the risk of TB activation (Bishwakarma et al., 2015; Houben et al., 2010; Horsburgh et al., 2010; Jeon and Murray, 2008), although our parameter estimates under the parallel structure scenario suggest that around 9% of infected individuals would belong to a high-risk group in which the risk of TB activation would be as high as 2000-fold that of the low-risk group. Alternatively, for the series structure of the latency compartments, rapid identification and early treatment of infected people would be the priority.

### 4.3. Age dependency

Our estimations highlight important discrepancies between age categories, with parameter estimates for young children (<5 years old) differing markedly from the other age categories. In particular, our study suggests that for young children, the rate of disease activation in the high-risk population (early latency for serial structure or high-risk group for parallel structure) is much greater than for 15+ year olds. In contrast, among the lower risk population (late latency for serial structure or low risk group for parallel structure), young children seem to be at lower risk of activation than the other individuals. These findings indicate that the youngest infected population presents a much higher risk of TB activation early after infection, while their risk of activation reduces dramatically in the later stages of infection. Previous works have already identified age as an important factor influencing the natural history of TB (Trauer et al., 2016a; Sloot et al., 2014; Vynnycky and Fine, 1997), and our analysis brings additional insight that reinforces the importance of providing young children with early preventive treatment for infection as late treatments would be useless. By revealing such age-specific characteristics, our study also implies that future TB modelling works should incorporate age-stratified structures in order to replicate TB activation dynamics accurately. Our analysis also demonstrates that simplified model structures incorporating no reactivation are adapted to simulate TB latency in the "<5 years old" category, suggesting that a significant proportion of infected children will never reactivate. Depending on the model structure employed, these young individuals could either become immune after some duration of infection (serial structure) or be protected immediately after infection (parallel structure).

### 4.4. Comparison with previous works

When two latency compartments are positioned in series, the rate of transition from early latency to late latency reflects the period for which individuals remain at high risk of disease. We estimate that the average early latency period is 52, 70 days and 146 days, for child(<5), adolescents (5–14) and adults, respectively. Most previous works employing the serial structure have used 5 years to define early latency, based on a previous convention to define primary disease versus endogenous disease (Vynnycky and Fine, 1997; Holm, 1969), and therefore may have greatly overestimated the actual duration of high risk. Our review of articles showed that the models with these long durations at high risk were also associated with lower rates of fast progression to active disease than the rates estimated from our study which compensates for the long early latency and leads to similar overall risks of activation over a lifetime. However, we have shown that our approach, which has dramatically shorter durations for the time spent in early latency and at high risk, is required in order to reproduce the profile of activation dynamics accurately.

Our estimate for the rate of late reactivation corresponds to a risk of 0.20 cases per 100 person-years for all ages, but is much lower among the '<5 years old' category. However, although our analysis clearly highlights different patterns of reactivation by age, accurate estimation of the reactivation parameter for the '<5 years old' category was limited by the small number of reactivation cases observed in this sub-group, which explains the wide variation in the associated confidence intervals. While previous works reported somewhat lower estimates for the late reactivation rate for all ages, ranging from 0.04 to 0.16 cases per 100 person-years (Vynnycky and Fine, 1997; Horsburgh et al., 2010; Comstock et al., 1967; Ferebee, 1970), our findings suggest that these results can only be interpreted in the context of age mix in the studies, and we can speculate that some of this variability may be accounted for by different proportions of "<5 year olds" in the study populations. Detailed data by age are not available from previous works so formal analysis by age has not been possible. Another possible explanation for the difference between our estimate and previous values for the reactivation rate is that the definition of infection in our datasets may be more specific than what was used in previous studies. In particular, looser definitions for infection may include false positives which would tend to increase the number of persons "at risk" and therefore reduce the inferred reactivation rate.

While our study shows that previous TB models have not always incorporated the optimal structure or parameterisation to account for the specific patterns observed in TB activation dynamics, the potential consequences of using suboptimal approaches remains undetermined. However, an analysis by Dowdy and colleagues of the most influential parameters to TB dynamics demonstrated that the parameters describing early latency are among those with the greatest impact on model predictions of steady-state TB incidence (Dowdy et al., 2013). This suggests that it is critical to reproduce early dynamics of TB infection closely in order to provide accurate insight into the epidemics trajectory. Therefore, future works investigating the consequences of employing inappropriate model structures or parameterisation are needed.

### 4.5. Potential for other future works

Our analysis was based on data from very low TB endemic settings where re-infection is expected not to play an important role (Wang et al., 2007). This allowed our estimation to focus on the potential progression to active disease following a single infection event and to avoid the confusion that would emerge from repeated infections when estimating the time from infection to activation. In the event that similar datasets became available in higher endemic

settings, a follow-up study could be conducted by integrating our "re-infection free" parameterisation into a model that would include an additional pathway for re-infection during latency. By fitting such a model to the new data, the re-infection parameter could then be estimated, thus providing valuable insights into the relative contribution of re-infection compared to the risk of first infection.

It is important to remember that the calibrations presented in this study are associated with the specific models that we selected. Accordingly, such estimates could not directly be used in models that incorporate a different structure or that do not belong to the category of deterministic compartmental transmission dynamic models. However, provided that both the structure and the nature of the model correspond to those used in this study, our work could also be used to re-estimate parameters associated with many different settings. To this end, we provide a step-by-step method which allows for rapid estimation of the different model parameters by performing simple measures on the reactivation failure curve. Consequently, our study could also be used to inform models in settings that present different characteristics, such as high HIV-endemic settings where the activation dynamics are known to be different (Houben et al., 2010; Horsburgh et al., 2010).

One natural limitation of this work is that it is linked to the epidemiological challenges of timing infection and reactivation accurately. However, the estimated durations that we use in this study are derived from two recently published works that employed rigorous definitions for both dates of infection and activation (Trauer et al., 2016a; Sloot et al., 2014).

## 5. Conclusions

Only models employing two latency compartments are able to reproduce TB latency dynamics accurately. We provide parameter values to optimally simulate epidemiological observations in such models along with an approach to obtaining such values from future epidemiological studies. Our analysis also reinforces the importance of age-stratification for capturing the dramatic differences between age groups in patterns of reactivation, which imply fundamental biological differences between age groups. However, we also demonstrate that data of the type this analysis is based upon cannot be used to determine the ideal configuration for the two latency compartments.

## Competing interests

We have no competing interests.

## Author's contributions

RR, JMT and ESM designed the study. RR performed the analysis. RR, JMT, NS, MTM, JTD and ESM interpreted the results. RR, JMT, NS, MTM, JTD and ESM wrote the manuscript.

## Funding

RR was supported by an Australian Government Research Training Program Scholarship.

## Acknowledgments

We thank the nurses and other staff of the Victorian TB Program for collecting the data In particular, we would like to thank Nompilo Moyo and Ee-Laine Tay who compiled the dataset.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.epidem.2017.06.002.

## References

Abu-Raddad, L.J., Sabatelli, L., Achterberg, J.T., Sugimoto, J.D., Longini Jr., I.M., Dye, C., Halloran, M.E., 2009. Epidemiological benefits of more-effective tuberculosis vaccines, drugs, and diagnostics. Proc. Natl. Acad. Sci. U. S. A. 106 (33), 13980–13985.

Aparicio, J.P., Castillo-Chavez, C., 2009. Mathematical modelling of tuberculosis epidemics. Math. Biosci. Eng. 6, 209–237.

Bishwakarma, R., Kinney, W.H., Honda, J.R., Mya, J., Strand, M.J., Gangavelli, A., Bai, X., Ordway, D.J., Iseman, M.D., Chan, E.D., 2015. Epidemiologic link between tuberculosis and cigarette/biomass smoke exposure: limitations despite the vast literature. Respirology 20 (4), 556–568.

Blower, S.M., Chou, T., 2004. Modeling the emergence of the 'hot zones': tuberculosis and the amplification dynamics of drug resistance. Nat. Med. 10, 1111–1116.

Castillo-Chavez, C., Feng, Z., 1997. To treat or not to treat: the case of tuberculosis. J. Math. Biol. 35 (6), 629–656.

Cohen, T., Lipsitch, M., Walensky, R.P., Murray, M., 2006. Beneficial and perverse effects of isoniazid preventive therapy for latent tuberculosis infection in HIV-tuberculosis coinfected populations. Proc. Natl. Acad. Sci. U. S. A. 103, 7042–7047.

Cohen, T., Colijn, C., Finklea, B., Wright, A., Zignol, M., Pym, A., Murray, M., 2008. Are survey-based estimates of the burden of drug resistant TB too low? Insight from a simulation study. PLoS One 3 (6), e2363.

Colijn, C., Cohen, T., Murray, M., 2008. Latent coinfection and the maintenance of strain diversity. Bull. Math. Biol. 71, 247–263.

Comstock, G.W., Ferebee, S.H., Hammes, L.M., 1967. A controlled trial of community-wide isoniazid prophylaxis in Alaska. Am. Rev. Respir. Dis. 95 (6), 935–943.

Dowdy, D.W., Dye, C., Cohen, T., 2013. Data needs for evidence-based decisions: a tuberculosis modeler's 'wish list'. Int. J. Tuberc. Lung Dis. 17 (7), 866–877.

Dye, C., 2012. The potential impact of new diagnostic tests on tuberculosis epidemics. Indian J. Med. Res. 135, 737–744.

Ferebee, S.H., 1970. Controlled chemoprophylaxis trials in tuberculosis: a general review. Bibl. Tuberc. 26, 28–106.

Gomes, M.G., Franco, A.O., Gomes, M.C., Medley, G.F., 2004. The reinfection threshold promotes variability in tuberculosis epidemiology and vaccine efficacy. Proc. Biol. Sci. 271, 617–623.

Hill, A.N., Becerra, J., Castro, K.G., 2012. Modelling tuberculosis trends in the USA. Epidemiol. Infect. 140, 1862–1872.

Holm, J., 1969. Development from Tuberculous Infection to Tuberculous Disease. KNVC, The Hague, The Netherlands.

Horsburgh Jr., C.R., O'Donnell, M., Chamblee, S., Moreland, J.L., Johnson, J., Marsh, B.J., Narita, M., Johnson, L.S., von Reyn, C.F., 2010. Revisiting rates of reactivation tuberculosis: a population-based approach. Am. J. Respir. Crit. Care Med. 182 (3), 420–425.

Houben, R.M., Dodd, P.J., 2016. The global burden of latent tuberculosis infection: a Re-estimation using mathematical modelling. PLoS Med. 13 (10), e1002152.

Houben, R.M., Glynn, J.R., Mallard, K., Sichali, L., Malema, S., Fine, P.E., French, N., Crampin, A.C., 2010. Human immunodeficiency virus increases the risk of tuberculosis due to recent re-infection in individuals with latent infection. Int. J. Tuberc. Lung Dis. 14 (7), 909–915.

Jeon, C.Y., Murray, M.B., 2008. Diabetes mellitus increases the risk of active tuberculosis: a systematic review of 13 observational studies. PLoS Med. 5 (7), e152.

Lin, H.H., Langley, I., Mwenda, R., Doulla, B., Egwaga, S., Millington, K.A., Mann, G.H., Murray, M., Squire, S.B., Cohen, T., 2011. A modelling framework to support the selection and implementation of new tuberculosis diagnostic tools. Int. J. Tuberc. Lung Dis. 15, 996–1004.

Menzies, N.A., Cohen, T., Lin, H.H., Murray, M., Salomon, J.A., 2012. Population health impact and cost-effectiveness of tuberculosis diagnosis with Xpert MTB/RIF: a dynamic simulation and economic evaluation. PLoS Med. 9, e1001347.

Sloot, R., Schim van der Loeff, M.F., Kouw, P.M., Borgdorff, M.W., 2014. Risk of tuberculosis after recent exposure: a 10-year follow-up study of contacts in Amsterdam. Am. J. Respir. Crit. Care Med. 190 (9), 1044–1052.

Trauer, J.M., Denholm, J.T., McBryde, E.S., 2014. Construction of a mathematical model for tuberculosis transmission in highly endemic regions of the Asia-Pacific. J. Theor. Biol. 358, 74–84.

Trauer, J.M., Moyo, N., Tay, E.L., Dale, K., Ragonnet, R., McBryde, E.S., Denholm, J.T., 2016a. Risk of active tuberculosis in the five years following infection. 15%? Chest 149 (2), 516–525.

Trauer, J.M., Denholm, J.T., Waseem, S., Ragonnet, R., McBryde, E.S., 2016b. Scenario analysis for programmatic tuberculosis control in western province, Papua New Guinea. Am. J. Epidemiol. 183 (12), 1138–1148.

Vynnycky, E., Fine, P.E., 1997. The natural history of tuberculosis: the implications of age-dependent risks of disease and the role of reinfection. Epidemiol. Infect. 119 (2), 183–201.

WHO, 2016. Global Tuberculosis Report. World Health Organization.
Wallinga, J., Lipsitch, M., 2007. How generation intervals shape the relationship between growth rates and reproductive numbers. Proc. Biol. Sci. 274 (1609), 599–604.
Wang, J.Y., Lee, L.N., Lai, H.C., Hsu, H.L., Liaw, Y.S., Hsueh, P.R., Yang, P.C., 2007. Prediction of the tuberculosis reinfection proportion from the local incidence. J. Infect. Dis. 196 (2), 281–288.
Wearing, H.J., Rohani, P., Keeling, M.J., 2005. Appropriate models for the management of infectious diseases. PLoS Med. 2 (7), e174.