Epidemiology and Infection

Modelling Tuberculosis Trends in the United States

A.N. Hill, J.E. Becerra, K.G. Castro

Supplementary Material

Model structure

A flowchart (Figure 1, main text) depicts the compartmental model. At any given time, an individual's TB status is categorized by membership of exactly one compartment. Individuals' categories change over time as they move between compartments according to their TB status. Temporal derivatives of compartment size are expressed in terms of rates of transfer of individuals between compartments. Subscripts 0, 1 refer to U.S.-born (USB) and foreign-born (FB) subpopulations, respectively. Upper case Roman letters refer to compartment populations. Lower case Roman letters denote proportions or probabilities (dimensionless). Greek letters denote rates, units are per time. As we are interested in annual incidence rates, time is measured in years and rates are expressed per year. The model compartments are as follows.

S: susceptible individuals.
F: infected individuals who progress to disease within two years of infection. This represents recent (or acute) infection, also referred to as primary (or fast) progression.
L: infected individuals who progress to disease over a longer time period. This route to disease represents endogenous reactivation, also referred to as slow progression, and individuals are said to have chronic latent TB infection (LTBI).
I: individuals with infectious TB.
J: individuals with non-infectious TB.
$N_i = S_i + F_i + L_i + I_i + J_i$: total population for USB ($i = 0$), FB ($i = 1$).

Individuals are assumed to contact each other randomly, with some preferred mixing within one's own group (a fraction e_0 of contacts is reserved for USB with USB, a fraction e_1 for FB with FB). When a susceptible person comes into contact with a person with infectious TB, the possibility of transmission is determined by the effective contact rate (β). If transmission occurs, the formerly susceptible individual becomes infected and either moves into the primary progression compartment, with probability p, or into the chronic LTBI compartment.

New individuals are recruited into the USB population either by birth (rate ρ) or into the FB population via arrival from other countries (rate α). We assume that the rates of change of the subpopulations due to births and arrivals are proportional to the size of the total population. Those born in the U.S., either of USB or FB parents, are assumed to be susceptible. A fraction f of those arriving from other countries is assumed to be already latently infected with TB, a fraction gp ($0 \leq g \leq 1$) of whom are assigned to the acute infection compartment.

Latently infected individuals can progress to TB disease either relatively quickly from state F (rate ν^F) or over a much longer period from state L (rate ν_i^L), if at all. A fraction q progresses to infectious TB, the rest to non-infectious TB. Progression will not occur if these individuals are first successfully treated for infection (rates σ^F for recent infection, σ^L for chronic LTBI), whereupon they move back into the susceptible compartment. Persons with chronic LTBI may be exogenously re-infected by infectious individuals and move into compartment F, but it is assumed they have partial immunity from their original infection, quantified by the fraction x. Individuals with disease may recover naturally or through treatment of disease (combined rate φ) and move back into the susceptible compartment. Individuals may die from natural causes at any stage (natural mortality rates μ_i), while those with disease are subject to an additional mortality rate due to disease (μ^d).

The prevalence among USB of chronic LTBI in 2000 is denoted by l_0 and among FB by l_1. The fraction of new cases in 2000 due to reactivation of chronic LTBI is denoted by r_0 for USB and r_1 for FB. The annual risk of infection for USB in 2000 is denoted by ARI_0.

We summarize the parameters used in the model with their symbols and best-fit values from the Latin hypercube sampling described in the main text.
Natural mortality rate: $\mu_0 = 1/78$ year^{-1} (USB); $\mu_1 = 1/53$ year^{-1} (FB).
USB birth rate: $\rho = 0.018$ year^{-1}.
FB arrival rate: $\alpha = 0.005$ year^{-1}.
Fraction of new infections which are acute (fast progressors): $p = 0.103$.
Progression rate of acute infection: $\nu^F = 1.5$ year^{-1}.
Prevalence of LTBI in the USB population in 2000: $l_0 = 0.015$.
Prevalence of LTBI in the FB population in 2000: $l_1 = 0.211$.
Fraction of cases due to reactivation in the USB population: $r_0 = 0.667$.
Fraction of cases due to reactivation in the FB population: $r_1 = 0.780$.
Progression rate for reactivation (chronic LTBI) in the USB population: $\nu_0^L = 0.0014$ year^{-1}.
Progression rate for reactivation (chronic LTBI) in the FB population: $\nu_1^L = 0.0010$ year^{-1}.
Fraction of infections progressing to infectious disease: $q = 0.708$.
Mortality rate due to TB: $\mu^d = 0.115$ year^{-1}.
Fraction of re-infected chronic LTBI moving to acute infection: $x = 0.111$.
Fraction of FB arrivals with LTBI: $f = 0.187$.
Annual risk of infection for USB in 2000: $\text{ARI}_0 = 0.030/100$.
Effective contact rate: $\beta = 10.39$ year^{-1}.
Fraction of preferred contacts with own population for USB: $e_0 = 0.965$.
Fraction of preferred contacts with own population for FB: $e_1 = 0.985$.
Fraction of FB arrivals with LTBI who are fast progressors: $g = 0.0047$.
Cumulative fraction self-cure and treatment of active disease for both populations and their corresponding rates: 0.897, $\varphi_0 = 1.114$ year^{-1} (USB), $\varphi_1 = 1.167$ year^{-1} (FB).
Cumulative fraction of treatment for acute infection for both populations and their corresponding rates: 0.461, $\sigma_0^F = 1.296$ year^{-1} (USB), $\sigma_1^F = 1.301$ year^{-1} (FB).
Treatment rate for chronic LTBI: $\sigma^L = 0.057$ year^{-1}.

Denoting the forces of infection (details below) by λ_i, the model is described by a series of ordinary differential equations, as follows. Dot superscripts denote derivatives with respect to

time.

$$
\begin{aligned}
\dot{S}_0 &= \rho(N_0+N_1) + \sigma_0^F F_0 + \sigma^L L_0 + \varphi_0(I_0+J_0) - \lambda_0 S_0 - \mu_0 S_0 \\
\dot{F}_0 &= p\lambda_0 S_0 + xp\lambda_0 L_0 - (\mu_0 + \nu^F + \sigma_0^F) F_0 \\
\dot{L}_0 &= (1-p)\lambda_0 S_0 - xp\lambda_0 L_0 - (\mu_0 + \nu_0^L + \sigma^L) L_0 \\
\dot{I}_0 &= q(\nu^F F_0 + \nu_0^L L_0) - (\mu_0 + \mu^d + \varphi_0) I_0 \\
\dot{J}_0 &= (1-q)(\nu^F F_0 + \nu_0^L L_0) - (\mu_0 + \mu^d + \varphi_0) J_0 \\
\dot{S}_1 &= (1-f)\alpha(N_0+N_1) + \sigma_1^F F_1 + \sigma^L L_1 + \varphi_1(I_1+J_1) - \lambda_1 S_1 - \mu_1 S_1 \\
\dot{F}_1 &= gpf\alpha(N_0+N_1) + p\lambda_1 S_1 + xp\lambda_1 L_1 - (\mu_1 + \nu^F + \sigma_1^F) F_1 \\
\dot{L}_1 &= (1-gp)f\alpha(N_0+N_1) + (1-p)\lambda_1 S_1 - xp\lambda_1 L_1 - (\mu_1 + \nu_1^L + \sigma^L) L_1 \\
\dot{I}_1 &= q(\nu^F F_1 + \nu_1^L L_1) - (\mu_1 + \mu^d + \varphi_1) I_1 \\
\dot{J}_1 &= (1-q)(\nu^F F_1 + \nu_1^L L_1) - (\mu_1 + \mu^d + \varphi_1) J_1
\end{aligned}
$$

Annual incidence per million in each subpopulation at time t is calculated as $10^6 \times (\nu^F F_i(t) + \nu_i^L L_i(t))/N_i(t)$ and the overall incidence is the weighted average of these, relative weights given by $N_i(t)$.

Forces of infection, λ_i, are determined by the assumptions of preferred mixing within one's own subpopulation. We follow the contact structure described in [1, 2] where the general expression for force of infection is

$$\lambda_i = \gamma_i p_i \left(c_{i0} \tfrac{I_0}{N_0} + c_{i1} \tfrac{I_1}{N_1} \right).$$

Here, γ_i is the average *per capita* contact rate for individuals in group i; p_i is the probability of infection of individuals in group i on coming into contact with any infectious individual; c_{ij} is the proportion of contacts made by members of group i with members of group j so that $c_{i0} + c_{i1} = 1$ for $i = 0, 1$. The construction assumes frequency dependent transmission [3].

The c_{ij} terms take the form

$$c_{ij} = e_i \delta_{ij} + (1-e_i) \frac{(1-e_j)\gamma_j N_j}{(1-e_0)\gamma_0 N_0 + (1-e_1)\gamma_1 N_1},$$

where e_i is the proportion of i-group contacts reserved for others in group i and δ_{ij} is 1 when $i = j$ and 0 otherwise. If $e_i = 0$ for all i, we have random mixing; if $e_i = 1$ for all i, we have completely assortative mixing. In between these two extremes, the e_i terms produce a convex combination of preferred mixing within one's own group and random mixing across the whole population including one's own group.

We further simplify by assuming that $\beta = \gamma p$ is the same for USB ($i = 0$) and FB ($i = 1$) so that the forces of infection are now given by

$$
\begin{aligned}
\lambda_i &= \beta \left(c_{i0} \tfrac{I_0}{N_0} + c_{i1} \tfrac{I_1}{N_1} \right), \\
c_{01} &= (1-e_0) \frac{(1-e_1)N_1}{(1-e_0)N_0 + (1-e_1)N_1}, \quad c_{00} = 1 - c_{01}, \\
c_{10} &= (1-e_1) \frac{(1-e_0)N_0}{(1-e_0)N_0 + (1-e_1)N_1}, \quad c_{11} = 1 - c_{10}.
\end{aligned}
$$

Relationship between cumulative fractions and rates

If a compartment has n outflows, denoted by rates $\omega_1, \ldots, \omega_n$, the cumulative fraction leaving by outflow corresponding to ω_i is

$$h_i = \frac{\omega_i}{\omega_1 + \cdots + \omega_n}.$$

It follows that

$$\omega_i = \frac{h_i}{1 - h_i} \times \sum_{k \neq i} \omega_k.$$

With constant outflow parameters, the proportion leaving the compartment by any means over a period of time t is $1 - e^{-(\omega_1 + \cdots + \omega_n)t}$ and the proportion leaving by outflow route corresponding to ω_i is

$$h_i \times \left(1 - e^{-(\omega_1 + \cdots + \omega_n)t}\right) = \frac{\omega_i}{\omega_1 + \cdots + \omega_n} \times \left(1 - e^{-(\omega_1 + \cdots + \omega_n)t}\right).$$

If ω_i is much greater than the sum of the other rates, the last expression may be approximated by $1 - e^{-\omega_i t}$. These relationships show how to calculate a particular rate from the cumulative fraction and the other rates for that compartment specified in Table 1 of the main text. As the cumulative fraction of treatment for active disease is assumed to be the same for both the USB and FB populations, the corresponding rate differs very slightly between the USB and FB populations on account of their different natural mortality rates. For the same reason, the treatment rates of acute infection differ slightly for USB and FB.

Calculation of demographic rates, effective contact rate, reactivation rates and initial conditions

Demographic rates were estimated from U.S. Census Bureau [4] and [5]. The overall life expectancy in 2000 was 76.9 years and 78.3 over the period 2005 – 2010. Accordingly, we set $\mu_0 = 1/78$ year^{-1}. The average age of entry of FB persons into the United States was around 25 years in the period 2000 – 2007 [5]. We set $\mu_1 = 1/(78 - 25) = 1/53$ year^{-1}.

Overall population sizes may be approximated in closed form when dealing with low prevalence disease. Ignoring deaths due to TB, governed by the μ^d term, the equations for the subpopulations N_0, N_1 are

$$\begin{aligned} \dot{N}_0 &\approx (\rho - \mu_0)N_0 + \rho N_1, \\ \dot{N}_1 &\approx \alpha N_0 + (\alpha - \mu_1)N_1. \end{aligned}$$

This first-order homogeneous linear system may be solved exactly. We used United States Census Bureau projections to 2050 [6] for total population, births, vital events and net international migration to estimate projections for the USB and FB populations by difference equations, using the fact that the FB population was approximately 10% of the total in 2000. This resulted in the long-term proportion of FB being approximately 18%. In conjunction with the exact solution for the system involving N_0 and N_1 above, we estimated ρ and α as in Table 1 of the main text by minimizing the least squares fit of the exact solution to overall population data 2000 – 2008 on the log scale.

The annual risk of infection (ARI) at time t is given in terms of the force of infection by

$$\mathrm{ARI}(t) = 1 - \exp\left(-\int_t^{t+1} \lambda(s)\, ds\right).$$

When the force of infection is small, as it is for TB in the United States, we may approximate the ARI as

$$\mathrm{ARI}(t) \simeq 1 - e^{-\lambda(t)\times 1} \simeq \lambda(t).$$

The factor of 1 in the second term refers to one year so that $\lambda(t)\times 1$ is dimensionless because the force of infection is a rate. The annual risk of infection for the USB population in 2000 (ARI_0) was estimated in [7] to be 0.02% – 0.03%. By using the crude approximation that prevalence of infectious disease is equal to the incidence of infectious disease multiplied by the duration of infectiousness, we can estimate the effective contact rate, β, from the following:

$$\begin{aligned}
\lambda_0(2000) &\approx \beta\left(c_{00}\tfrac{I_0(2000)}{N_0(2000)} + c_{01}\tfrac{I_1(2000)}{N_1(2000)}\right) \\
&\approx \beta\left(c_{00}q \times \text{USB incidence} + c_{01}q \times \text{FB incidence}\right) \times \text{duration} \\
\Rightarrow \mathrm{ARI}_0 &\approx \beta \times (c_{00} \times \text{USB incidence} + c_{01} \times \text{FB incidence}) \times q/(\mu + \mu^d + \varphi),
\end{aligned}$$

where incidences are those reported for USB and FB in 2000 (35 and 273 cases per million, respectively [8].

The progression rate for recent infection is determined by the assumption that 95% of those moving into the F_i compartment progress to disease within two years: $\nu^F = -\log(1-0.95)/2 \approx 1.5\ \mathrm{year}^{-1}$.

The National Health and Nutrition Examination Survey, 1999 – 2000 [9] estimated the prevalence of chronic LTBI in the USB, FB (L_i/N_i) and overall population in 2000. Using estimates of the proportion (r_i) of new cases due to reactivation of chronic LTBI, as discussed in the main text, and surveillance data for 2000, the endogenous reactivation rate (ν_i^L) and initial numbers of acute infections (F_i) are calculated thus:

$$\begin{aligned}
\nu_i^L &= r_i \times (\text{new cases in population } i) \div L_i = \frac{r_i \times \text{incidence in population } i}{\text{prevalence of LTBI in population } i} \\
F_i &= (1 - r_i) \times (\text{new cases in population } i) \div \nu^F
\end{aligned}$$

We estimate initial ($t = 2000$) active disease numbers I_i, J_i by again approximating prevalence as the product of initial incidence and duration of disease.

$$\begin{aligned}
I_i &= q \times (\text{new cases in population } i) \div (\mu + \mu^d + \varphi) \\
J_i &= (1 - q) \times (\text{new cases in population } i) \div (\mu + \mu^d + \varphi)
\end{aligned}$$

The initial number of susceptible persons was calculated as the difference between the total subpopulation N_i in 2000 and the sum of all the other compartments in 2000.

Rough estimate of the treatment rate for chronic LTBI

The number of treatment starts for LTBI in the United States in 2002 has been estimated between 291,000 and 433,000 [10]. Based on [9], our model estimates of LTBI in the United

States in 2000 were around 10 million. A quick estimate of the treatment rate for chronic LTBI (σ^L) is therefore 0.029 – 0.043 year^{-1}. The latter number corresponds to the median value of σ^L obtained from our Latin hypercube sample and the best-fit was 0.057 year^{-1}, indicating the right magnitude.

We also imputed the rate independently of published estimates and arrived at the same range. Assume there is no transmission of disease so that $\beta = 0$ and the forces of infection are zero. The differential equations for the two USB latent infection compartments, F_0 and L_0, simplify to:

$$\begin{aligned} \dot{F}_0 &= -(\mu_0 + \nu^F + \sigma_0^F)F_0 \\ \dot{L}_0 &= -(\mu_0 + \nu_0^L + \sigma^L)L_0 \end{aligned}$$

Each has an exponential function solution:

$$\begin{aligned} F_0(t) &= F(2000)\, e^{-(\mu+\nu^F+\sigma^F)(t-2000)} \\ L_0(t) &= L(2000)\, e^{-(\mu+\nu_0^L+\sigma^L)(t-2000)} \end{aligned}$$

The model expression for incident cases is

$$\nu^F \times F(2000)\, e^{-(\mu_0+\nu^F+\sigma_0^F)(t-2000)} + \nu_0^L \times L(2000)\, e^{-(\mu_0+\nu_0^L+\sigma^L)(t-2000)}$$

The fast progression rate σ_0^F is in the order of 100 times the magnitude of μ_0, which in turn is around 10 times the magnitude of ν_0^L. A rough approximation to the incident new cases is therefore given by dropping the first term, the cases arising from recent infection. Incident new cases are approximately

$$\nu_0^L \times L(2000)\, e^{-(\mu_0+\nu_0^L+\sigma^L)(t-2000)}.$$

From 2002 to 2008, the number of USB incident cases decreased at approximately -5.3% annually, translating to an exponential decay rate of $\log(1 - 5.3/100) = -0.0545\,\text{year}^{-1}$. At the year 2000, we therefore have

$$\mu_0 + \nu_0^L + \sigma^L \approx 0.0545.$$

There were 8,649 new cases among the USB in 2000 [8]. Assuming that 65% of new cases in the USB arose from reactivation in 2000 (Table 1), and that the USB chronic LTBI number in 2000 is 4.154 million [9],

$$\nu_0^L = \frac{0.65 \times 8649}{4.154 \times 10^6} = 0.0014\,\text{year}^{-1}.$$

The mortality parameter $\mu_0 = 1/78 = 0.0128$ year^{-1}. Hence, the treatment rate is

$$\sigma^L = 0.0545 - 0.0128 - 0.0014 = 0.0403\,\text{year}^{-1},$$

in accordance with the median rate for the best-fit samples reported in Table 1.

Sensitivity analysis

Calculation of partial rank correlation coefficients (PRCC) of model parameters with three different outcomes was performed: 1. projected year of elimination in the USB population; 2. projected annual incidence in 2100 in the FB population; 3. projected annual incidence in 2100 in the overall population. Table S 1 summarizes these. PRCCs of smaller magnitude indicate less influence on the outcome [11, 12].

All three outcomes are highly sensitive to the level of treatment of chronic LTBI (σ^L). Negative PRCCs indicate that increases in this rate, within the range of values sampled, bring forward the elimination year and reduce both FB and overall incidence in 2100.

Beyond this, there are differences between the sensitivity of USB elimination year to the model parameters when compared with the sensitivities of long-term incidence in the FB and overall populations, which behave similarly. Other influential parameters ($|\text{PRCC}| > 0.5$) for the USB elimination year are the fraction of contacts reserved for FB with FB (e_1) (negatively correlated) and the annual risk of infection (ARI_0) which determines the effective contact rate (β) (positively correlated). Parameter e_1 is negatively correlated with this outcome because, when the fraction of contacts increases, there is less infection of USB persons by FB persons and the time to elimination in the USB population goes down. For the FB and overall incidence in 2100, the FB progression rate of chronic LTBI (ν_1^L), the importation fraction (f) of LTBI in the FB population, and the proportion of infections which are acute (p) are influential, all positively correlated with outcome ($|\text{PRCC}| > 0.5$). For all three outcomes, the treatment rates of recent infection (σ^F) and active disease (φ) are not influential ($|\text{PRCC}| < 0.5$).

Table S 1: PRCCs of model parameters with the outcomes: elimination year for the USB subpopulation; incidence in 2100 for the FB subpopulation; incidence in 2100 for the overall population.

parameter	USB	parameter	FB	parameter	overall
σ^L	-0.9335	σ^L	-0.9374	σ^L	-0.9381
e_1	-0.7380	ν_1^L	0.8411	ν_1^L	0.8309
ARI_0	0.5230	f	0.8193	f	0.8072
p	0.4868	p	0.5978	p	0.6100
q	0.4619	ARI_0	0.4713	ARI_0	0.4939
σ^F	-0.2968	g	0.4655	q	0.4543
ν_0^L	0.1930	q	0.4329	g	0.4517
ν_1^L	0.1767	σ^F	-0.3675	σ^F	-0.3772
f	0.1475	e_1	0.1577	r_1	-0.1109
g	0.0902	r_1	-0.1133	r_0	0.0760
e_0	-0.0752	r_0	0.0740	μ^d	0.0513
φ	-0.0441	μ^d	0.0515	x	0.0345
r_0	0.0334	x	0.0382	ν_0^L	0.0266
r_1	-0.0229	φ	0.0184	φ	0.0177
x	-0.0174	ν_0^L	0.0094	e_0	-0.0072
μ^d	-0.0007	e_0	-0.0010	e_1	0.0046

Sensitivity of the model to multidrug-resistant (MDR) TB

The model presented in the main text can be modified to include MDR TB. We add two more compartments in each subpopulation, one for infectious MDR TB (I_i^{R}) and one for non-infectious MDR TB (J_i^{R}), $i = 0, 1$. Previous compartments I_i, J_i now represent drug sensitive TB. The MDR TB compartments are subject to reduced treatments rates, $h\varphi_i$, where $0 < h < 1$. To further test the sensitivity, we assume that MDR TB increases in the FB population from 1.2% of new cases in 2008 to 20% of new cases in 2100, while it remains constant at its 2008 level of 0.06% of new cases in the USB population. The 2008 levels are taken from [8]. This can be

modelled by the following system of differential equations.

$$
\begin{aligned}
\dot{S}_0 &= \rho(N_0+N_1) + \sigma_0^F F_0 + \sigma^L L_0 + \varphi_0(I_0+J_0+hI_0^{\mathrm{R}}+hJ_0^{\mathrm{R}}) - \lambda_0 S_0 - \mu_0 S_0 \\
\dot{F}_0 &= p\lambda_0 S_0 + xp\lambda_0 L_0 - (\mu_0+\nu^F+\sigma_0^F)F_0 \\
\dot{L}_0 &= (1-p)\lambda_0 S_0 - xp\lambda_0 L_0 - (\mu_0+\nu_0^L+\sigma^L)L_0 \\
\dot{I}_0 &= (1-u_0)q(\nu^F F_0+\nu_0^L L_0) - (\mu_0+\mu^d+\varphi_0)I_0 \\
\dot{J}_0 &= (1-u_0)(1-q)(\nu^F F_0+\nu_0^L L_0) - (\mu_0+\mu^d+\varphi_0)J_0 \\
\dot{I}_0^{\mathrm{R}} &= u_0 q(\nu^F F_0+\nu_0^L L_0) - (\mu_0+\mu^d+h\varphi_0)I_0^{\mathrm{R}} \\
\dot{J}_0^{\mathrm{R}} &= u_0(1-q)(\nu^F F_0+\nu_0^L L_0) - (\mu_0+\mu^d+h\varphi_0)J_0^{\mathrm{R}} \\
\dot{S}_1 &= (1-f)\alpha(N_0+N_1) + \sigma_1^F F_1 + \sigma^L L_1 + \varphi_1(I_1+J_1+hI_1^{\mathrm{R}}+hJ_1^{\mathrm{R}}) - \lambda_1 S_1 - \mu_1 S_1 \\
\dot{F}_1 &= gpf\alpha(N_0+N_1) + p\lambda_1 S_1 + xp\lambda_1 L_1 - (\mu_1+\nu^F+\sigma_1^F)F_1 \\
\dot{L}_1 &= (1-gp)f\alpha(N_0+N_1) + (1-p)\lambda_1 S_1 - xp\lambda_1 L_1 - (\mu_1+\nu_1^L+\sigma^L)L_1 \\
\dot{I}_1 &= (1-u_1(t))q(\nu^F F_1+\nu_1^L L_1) - (\mu_1+\mu^d+\varphi_1)I_1 \\
\dot{J}_1 &= (1-u_1(t))(1-q)(\nu^F F_1+\nu_1^L L_1) - (\mu_1+\mu^d+\varphi_1)J_1 \\
\dot{I}_1^{\mathrm{R}} &= u_1(t)q(\nu^F F_1+\nu_1^L L_1) - (\mu_1+\mu^d+h\varphi_1)I_1^{\mathrm{R}} \\
\dot{J}_1^{\mathrm{R}} &= u_1(t)(1-q)(\nu^F F_1+\nu_1^L L_1) - (\mu_1+\mu^d+h\varphi_1)J_1^{\mathrm{R}}
\end{aligned}
$$

The forces of infection are

$$\lambda_i = \beta\left(c_{i0}\tfrac{I_0+I_0^{\mathrm{R}}}{N_0} + c_{i1}\tfrac{I_1+I_1^{\mathrm{R}}}{N_1}\right).$$

The parameters u_i represent the fraction of new TB cases that are MDR. We ran simulations over the 5000 best-fit Latin hypercube samples (Table 1, main text) and assumed that MDR TB first occurred in 2008. Thus, $u_0 = 0.006$ and $u_1(t)$ increases from 0.012 in 2008 to 0.2 in 2100. We modelled a linear increase in $u_1(t)$. We took $h = 0.25$ which represented a reduction of approximately 50% of treatment of disease per year due to MDR TB.

Figure S 1 shows the resulting densities for elimination year for the USB population (c) and incidence in 2100 in the overall and FB populations (d). For comparison we include the corresponding densities for the non-MDR model (a and b). (These are Figures 2a and 2b in the main text.) As can be seen, the densities do not differ significantly, although a few simulations (82 of 5000) achieve long-term incidence above 200 cases per million annually in the FB population. More of the simulations for the USB population did not reach elimination by the year 2100 (2899 *vs.* 3860 of 5000), although most were relatively close. The median incidence of those not reaching elimination by 2100 was 1.5 per million and the maximum was 4.8 per million.

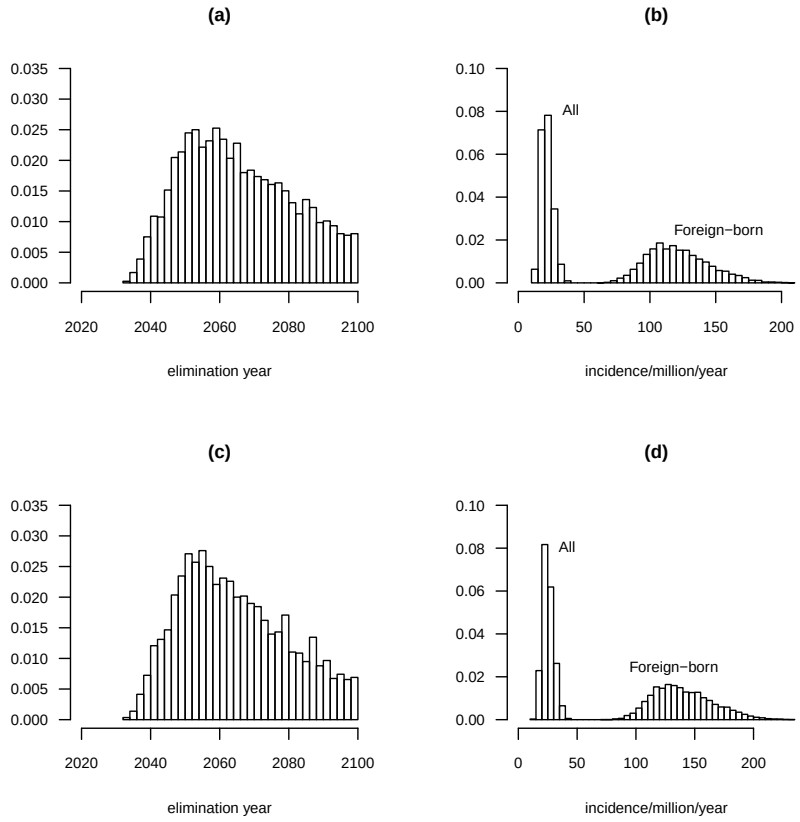

Figure S 1: The effect of including MDR TB in the model. Densities are shown for elimination years in the USB population, incidence in 2100 for the FB and overall population. These correspond to best-fit parameter sets Table 1 (main text). 1a: USB elimination year for non-MDR model. 1b: FB and overall incidence in 2100 for non-MDR model. 1c: USB elimination year for MDR model. 1d: FB and overall incidence in 2100 for MDR model.

References

[1] **Jacquez JA, *et al.*** Modeling and analyzing HIV transmission: the effect of contact patterns. *Mathematical Biosciences* 1988; **92**: 119–199.

[2] **Busenberg S, Castillo-Chavez C.** A general solution of the problem of mixing of subpopulations and its application to risk- and age-structured epidemic models for the spread of AIDS. *IMA Journal of Mathematics Applied in Medicine and Biology* 1991; **8**: 1–29.

[3] **Keeling MJ, Rohani P.** *Modeling Infectious Diseases in Humans and Animals.* Princeton, New Jersey: Princeton University Press, 2008, pp.17–18.

[4] **U.S. Census Bureau.** Current Population Survey. 2010 (`http://www.census.gov/cps/`). Accessed 7 July 2011.

[5] **Camarota SA.** Backgrounder. Immigrants in the United States, 2007: a profile of Americas foreign-born population. Center for Immigration Studies, November 2007. (`http://www.cis.org/immigrants_profile_2007`) Accessed 17 October 2011.

[6] **U.S. Census Bureau.** U.S. Population Projections. 2008 (`http://www.census.gov/population/www/projections/summarytables.html`). Accessed 7 July 2011.

[7] **Daniel TM, Debanne SM.** Estimation of the annual risk of tuberculous infection for white men in the United States. *Journal of Infectious Diseases* 1997; **175**: 1535–1537.

[8] **Centers for Disease Control and Prevention.** Reported Tuberculosis in the United States, 2008. Atlanta, GA: U.S. Department of Health and Human Services, CDC, September 2009 (`http://www.cdc.gov/tb/statistics/reports/2008/default.htm`). Accessed 7 July 2011.

[9] **Bennett DE, *et al.*** Prevalence of tuberculosis infection in the United States population. The National Health and Nutrition Examination Survey, 1999–2000. *American Journal of Respiratory and Critical Care Medicine* 2008; **177**: 348–355.

[10] **Sterling TR, *et al.*** The scope and impact of treatment of latent tuberculosis infection in the United States and Canada. *American Journal of Respiratory and Critical Care Medicine* 2006; **173**: 927–931.

[11] **Blower SM, Dowlatabadi H.** Sensitivity and uncertainty analysis of complex models of disease transmission: an HIV model, as an example. *International Statistical Review* 1994; **62**: 229 – 243.

[12] **Sanchez MA, Blower SM.** Uncertainty and sensitivity analysis of the basic reproductive rate. Tuberculosis as an example. *American Journal of Epidemiology* 1997; **145**: 1127 – 1137.