Web Appendix: Projected effects of tobacco smoking on global tuberculosis control: mathematical modelling analysis

Sanjay Basu, MD
David Stuckler, PhD
Asaf Bitton, MD
Stanton A. Glantz, PhD

Table of Contents

1. Model Structure

To simulate the process of TB pathogenesis, our model (Figure 1) followed prior models in containing a compartment of susceptible persons (*S*), a latently-infected compartment (*L*) and an active TB disease compartment (*T*) [1 2]. Some people who become infected with TB do not manifest active disease but are infected in the dormant state of "latent TB", which is non-pathological (does not confer a higher risk of death when compared to uninfected persons) and is non-infectious; other infected people experience "primary progressive disease", which means they develop active (potentially deadly and infectious) TB disease soon after infection. Hence, the susceptible compartment has rates of flow to both the latently-infected compartment and the active TB compartment. Even latently-infected persons occasionally can, however, "reactivate" to active TB later in life, which is depicted in the model as flow from the latent compartment to the active TB compartment.

Some older models have divided the active TB compartment into two compartments, "non-infectious" and "infectious" active TB [1]. Others no longer separate non-infectious and infectious active TB, because active TB cases that were previously thought to be non-infectious are often infectious [3]. As a result, active TB cases are now thought to have a population distribution of infectiousness rather than a dichotomous separation into just two binary categories. Therefore, for people in the one active TB compartment, we multiply the contact rate *c* by the fraction of cases *f* that are sufficiently infectious to produce transmission, where the value of *f* is sampled from a probability distribution of possible values (see below) [4]. Because latent disease is a non-deadly state, the latent disease compartment is subject to the same mortality as the susceptible compartment. The active TB compartments have both background mortality and an additional mortality rate from active TB-induced death. The process of recovery from active TB is depicted as moving people back to the latent state compartment (that is, active TB may "regress" by chemotherapy or even by natural "self-cure" known as "natural regression") [5], rather than moving to a separate "recovered persons" compartment. These people can relapse from the latent compartment back to the active disease compartment because of reactivation of dormant bacteria or reinfection with TB during latent disease, both of which can lead to active TB. (Later in this Appendix, we compare this approach to modeling pathogenesis to other approaches, by applying formal model selection criteria to examine which among three alternative models provides the most parsimonious fit to population-level TB rates).

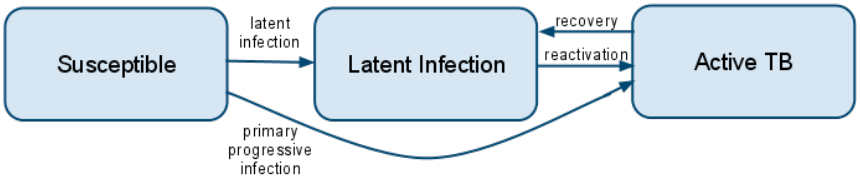


Figure 1: TB pathogenesis described by the model, using the same structure as prior models [6 7]. Smoking is integrated into the model by altering the rates of flow between compartments to reflect how smokers have different TB infection and mortality risks; HIV is similarly incorporated. These modifications are detailed below.

2. Model equations

The model can be described by a series of ordinary differential equations specifying the rates of change in the populations of each compartment over time, *t,* using the notation in Table 1.

TABLE 1: TB MODEL PARAMETERS.

PARAMETER	DEFINITION (UNITS)
c	contact rate (contacts per susceptible person per unit time that are sufficiently close as to potentially transmit infection if the contact is infectious)
f	fraction of infected contacts that are transmitting TB (fraction of active TB cases who are sufficiently infectious to result in transmission upon contact)
b	birth rate (1/year)
μ	background (non-TB) mortality rate (1/year)
μ_T	mortality rate due to TB (1/year)
p	proportion of newly-infected persons experiencing primary progressive disease (proportion of infected persons who undergo rapid progression from infection to active TB; hence, (1-p) is the proportion of newly-infected persons who progress to latent disease)
δ	proportion of active TB cases detected
ε	detection and treatment rate among active TB cases detected (how fast detection and therapy occurs after onset of active TB)
κ	proportion of treated cases who are successfully treated (regression to latency)
σ	rate of natural self-cure (rate of natural regression from active TB to latency, 1/year)
v	reactivation rate from latency to active TB (1/year)
x	proportion of latently-infected persons who are susceptible to primary progressive disease upon re-infection

The size of the *susceptible compartment* increases from births (b) among the total population $N(t)$, and decreases as people become infected or die (from both TB and non-TB causes). The infection rate depends on the number of people with active TB, $T(t)$, who can infect susceptible people (i.e., the number of new infections is dependent upon the number of currently-infectious people). The $T(t)$ infected people can make c contacts and potential transmissions of TB per unit time (i.e., the number of new people they come into close contact with and cough on, potentially transmitting *Mycobacterium tuberculosis* bacilli); fraction f of the infected people $T(t)$ are infectious and capable of transmitting the disease during these contacts. Therefore, the rate of infection is $cfT(t)$ per unit time. Among the newly-infected people, a fraction p undergo primary progression (rapid progression) to active TB and (1-p) develop latent disease (which can become active TB later). Therefore, the rate of infections leading to active TB is $pcfT(t)$ and the rate of infections leading to latent disease is $(1-p)cfT(t)$. Deaths from the susceptible compartment occur at rate μ per unit time.

Because smoking affects the risk of infection, we modify the rate of infection to reflect smokers' higher relative risk of infection. We define the proportion of the population who smokes $P_S(t)$ for any given time t, and the relative risk of infection among these smokers r_{ST}. If the rate of infection is $cfT(t)$ for nonsmoking susceptible people (proportion $(1-P_S(t))$ of the population), then the infection rate will be $r_{ST}cfT(t)$ among the proportion $P_S(t)$ of the population who smoke. Hence, the overall rate of infection will be $(1-P_S(t))cfT(t) + P_S(t)r_{ST}cfT(t) = (1-P_S(t)+P_S(t)r_{ST})cfT(t)$.

We apply the same logic to the higher background (non-TB) death rate that affects smokers (from cardiovascular disease, cancers, etc.). The nonsmoking susceptible proportion of the population $(1-P_S(t))$ dies at rate μ, while the smoking proportion of the population $P_S(t)$ die at a higher rate denoted μ_S. Hence, the overall mortality rate from the susceptible compartment will be $(1-P_S(t))\mu + P_S(t)\mu_S$.

Combining all these factors leads to equation 1, which describes the flow of people into and out of the susceptible compartment:

$$(1) \qquad \frac{dS(t)}{dt} = bN(t) - \{[1 - P_S(t) + r_{ST}P_S(t)]cfT(t) + [1 - P_S(t)]\mu + P_S(t)\mu_S\}S(t)$$

The number of *latently-infected people* increases as new infections become latent infections and as active TB cases regress back to latency. The latently-infected compartment is decreased by the rate at which latent cases reactivate to active TB disease, and the rate at which latent cases die.

The rate of infections entering the latent compartment is $(1\text{-}p)cfT(t)$, as described above. The rate at which active TB patients regress to latency is defined by two possibilities: regression can occur either by chemotherapy or by natural self-regression. Chemotherapeutic regression occurs when proportion δ of active TB cases are detected and treated at rate ε, of which proportion κ successfully regress; natural self-regression occurs at rate σ. Hence, the rate of flow from active TB to latent disease is $\delta\varepsilon\kappa+\sigma$.

The rate at which latently-infected people become active TB cases is composed of two pathogenic possibilities: natural "reactivation" of latent disease (which occurs at rate v)[1], and "exogenous reinfection," in which latently-infected persons are re-exposed and re-infected with TB, potentially causing them to experience primary progression to active TB. Exogenous reinfection transitions a person from latent to active TB when they are infected at rate $cfT(t)$ (as with susceptible people), but proportion x of the latently-infected people are not immune to re-infection (i.e., have not built up sufficient immunity from their prior infection). Of those people, proportion p undergo primary progression to active TB (the same p as the proportion of newly-infected susceptible people who undergo primary progression and become active cases). Hence, the rate at which people are exogenously re-infected is $xpcfT(t)$. The total rate at which people move from the latent TB to active TB compartment is therefore the sum of the rate of "reactivation" and the rate of "exogenous reinfection" for a total rate of $v + xpcfT(t)$.

Finally, those dying from latent TB only include those who are killed by the background non-TB death rate μ, because latent TB does not confer an additional mortality from TB itself.

Smoking changes the rate of new infections, as described above, such that the rate of infections that become latent will be $(1\text{-}p)cfT(t)$ for nonsmoking susceptible people (proportion $(1\text{-}P_S(t))$ of the susceptible population), and $r_{ST}(1\text{-}p)cfT(t)$ among the proportion $P_S(t)$ of the susceptible population who smoke. (Note that the variable p is the conventional TB variable for the proportion of infected persons who undergo primary progressive disease, while P_S is the proportion of the susceptible population who smoke.) Hence, the overall rate of new infections entering the latent compartment from the susceptible compartment will be $(1\text{-}P_S(t))(1\text{-}p)cfT(t) + P_S(t)(1\text{-}p)r_{ST}cfT(t) = (1\text{-}P_S(t)+P_S(t)r_{ST})(1\text{-}p)cfT(t)$.

Among latently-infected patients, we account for the proportion of the population who smoke $P_{SL}(t)$. The proportion of latently-infected persons who are exogenously re-infected ($xpcfT(t)$, as explained in the basic TB model description) will be modified by smoking, such that the rate of exogenous re-infection will be $xpcfT(t)$ for nonsmoking susceptible people (proportion $(1\text{-}P_{SL}(t))$ of the population) and will be $r_{ST}xpcfT(t)$ among the proportion $P_{SL}(t)$ of the population who smoke. Hence, the rate of exogenous reinfection will be $(1\text{-}P_{SL}(t))xpcfT(t) + P_{SL}(t)r_{ST}xpcfT(t) = (1\text{-}P_{SL}(t)+P_{SL}(t)r_{ST})xpcfT(t)$.

Finally, we must adjust the non-TB background mortality rate among latently-infected people to reflect the higher background mortality rate among smokers than nonsmokers. The proportion $(1\text{-}P_{SL}(t))$ of latently-infected persons die at rate μ, while smoking proportion $P_{SL}(t)$ die at higher rate μ_S. Hence, the overall mortality rate from the latently-infected compartment will be $(1\text{-}P_{SL}(t))\mu + P_{SL}(t)\mu_S$. There is no additional mortality from active TB in this group, since latency is a non-pathogenic state.

Hence, equation 2 describes the flow into and out of the latently-infected compartment: as:

$$(2) \qquad \frac{dL(t)}{dt} = [1 - P_S(t) + r_{ST}P_S(t)](1 - p)cfT(t)S(t) + (\delta\varepsilon\kappa + \sigma)T(t) - \{v + [1 - P_{SL}(t) + r_{ST}P_{SL}(t)]xpcfT(t) + [1 - P_{SL}(t)]\mu + P_{SL}(t)\mu_S\}L(t)$$

The number of *active TB cases* increases as people are infected and progress to active TB (from either the susceptible or latent compartments) and decreases when active TB cases regress to latency or die.

[1] As with all other natural history rates in this model, v is a constant. After infection with TB bacilli, a portion of people will rapidly advance to active TB disease, while another portion will progress later, slowly, from latency to active TB. Rather than using a complex function to account for this phenomenon, TB modelers (e.g., 1. Blower SM, McLean AR, Porco TC, Small PM, Hopewell PC, Sanchez MA, et al. The intrinsic transmission dynamics of tuberculosis epidemics. *Nat Med* 1995;1(8):815-21.) specified that one group of persons would undergo "primary progressive disease", or rapid transformation to active TB soon after infection (in long-term TB models, this is mathematically similar to skipping the latent state altogether), while other persons would progress to latency and experience the slow rate of reactivation, specified by v here.

As explained above, the population of infected people who undergo primary progressive disease is *pcfT(t)*. Latently infected people enter the active TB compartment by natural reactivation rate v and the exogenous reinfection rate $xpcfT(t)$. The rate of regression to latency is defined through chemotherapy (proportion δ detected, started on therapy at rate ε, and proportion κ successfully treated to produce regression) and natural regression (rate σ), such that the rate of movement from active to latent TB is $\delta\varepsilon\kappa+\sigma$. Finally, death from active TB can occur due to background (non-TB) deaths at rate μ as well as TB deaths at rate μ_T.

The rates of entry from the susceptible and latent compartments to the active TB compartment are modified by smoking as described earlier, as is the rate of active TB mortality to reflect the risk for active TB death among smokers. If the proportion of active TB cases that smoke is $P_{ST}(t)$, then the background (non-TB) death rate will be μ_S among proportion $P_{ST}(t)$ of actively-infected persons who smoke and μ among non-smokers (proportion $1-P_{ST}(t)$). The TB-related mortality rate will be increased by relative risk r_{SM} among smokers $P_{ST}(t)$, while being unchanged among portion $(1-P_{ST}(t))$ of TB-infected persons who do not smoke.

Hence, equation 3 describes the rate of flow into and out of the active TB compartment:

$$\begin{aligned}\frac{dT(t)}{dt} &= [1-P_S(t)+r_{ST}P_S(t)]pcfT(t)S(t) \\ &+\{v+[1-P_{SL}(t)+r_{ST}P_{SL}(t)].xpcfT(t)\}L(t) \\ &-\{\delta\varepsilon\kappa+\sigma+[1-P_{ST}(t)]\mu+P_{ST}(t)\mu_S+[1-P_{ST}(t)+r_{SM}P_{ST}(t)]\mu_T\}T(t)\end{aligned} \quad (3)$$

These equations were programmed into MATLAB version R2010b (MathWorks, Natick, MA) to perform the simulations.

3. Model parameterization

3.1 Natural history parameters

Table 1 of the primary manuscript shows the model parameter values, using log-normal distributions to describe the uncertainty in values[2 4]. Because the natural history parameters are derived from historical data prior to the smoking era, the rates of contact and natural progression do not incorporate the effect of smoking (i.e., our inclusion of the relative risk increase from smoking is not redundant with the other parameters). Following standard convention for TB models, parameters describe TB pathogenesis for persons above the age of 15, who constitute the majority of both TB transmission and TB case burden [2]. The natural history parameters are derived from a series of back-calculations from prevalence studies during prior TB epidemics as well as model-fitting exercises described in prior analyses [2].

To further justify the parameter values employed to simulate the impact of smoking on TB, we provide a more extensive review of prior studies examining the impact of smoking on TB in Table 2. Table 3 incorporates the results of prior studies that were included in the 2007 meta-analysis by Lin *et al*., and three additional studies published since 2007 quantifying the magnitude of smoking effects on TB infection or mortality risk. We replicated and updated Lin's meta-analysis using STATA's metan procedure to provide estimates and 95% confidence intervals around the effect sizes of smoking on TB infection and mortality, displayed at the end of appendix Table 2.

The risks of TB infection and progression in former smokers are the same as in nonsmokers according to a recent prospective cohort study among 486,341 adults that found smoking cessation reduced the risk of TB mortality to a level statistically indistinguishable from the population that had never smoked.[2] Furthermore, data from a multi-center TB case-control study on tobacco cessation indicates that the reduction in risk among active TB smokers occurs quickly enough that it is significantly manifest as a reduction of mortality to non-smoker levels during the course of TB therapy.[36] Hence, we simulated former smokers as returning to non-smokers' TB infection and mortality risks in this model.

HIV has a profound effect on TB pathogenesis and overall TB epidemic progression, so we used the common method of taking the weighted average (based on HIV prevalence each year) of the relevant HIV-positive and HIV-negative parameter values to incorporate HIV's impact on each parameter in the model (manuscript Table 1). For example, if 5% of the population is HIV-positive in a given year then the background nonsmokers' mortality rate, μ, would be $\mu = 0.05 * \mu_{HIV-} + 0.95 * \mu_{HIV+} = 0.05 * 0.10 + 0.95$

* 0.02 = 0.024 (see manuscript Table 1, row 5, columns 5 and 6). New weighted averages for each parameter in Table 1 are calculated for each year based on the HIV prevalence that year.

TABLE 2: SUMMARY OF RESEARCH CHARACTERIZING THE RELATIONSHIP BETWEEN TOBACCO SMOKING AND LATENT TB INFECTION, ACTIVE TB DISEASE OR TB MORTALITY. Adapted and updated from [8].

STUDY	EVALUATING RISK OF:	SAMPLE SIZE AND STUDY DESIGN	ODDS RATIO/RELATIVE RISK IN SMOKERS VS. NONSMOKERS (95% CONFIDENCE INTERVALS)
Anderson (1997) USA [9]	Latent infection	116 smokers vs. 177 nonsmokers (case-control)	1.78 (0.98-3.21)
Den Boon (2005) South Africa [10]	Latent infection	1,832 infections among 2,401 participants (case-control)	1.77 (1.41-2.21)
Hussain (2003) Pakistan [11]	Latent infection	225 infections among 425 participants (case-control)	3.20 (1.30-8.20)
Plant (2002) Vietnam [12]	Latent infection	898 infections among 1,395 participants (case-control)	2.31 (1.58-3.38)
Solsona (2001) Spain [13]	Latent infection	335 infections among 447 participants (case-control)	1.72 (1.02-2.86)
McCurdy (1997) USA [14]	Latent infection	49 infections among 296 participants (case-control)	1.87 (0.73-4.80)
Leung (2004) Hong Kong [15]	Active disease	252 cases among 42,655 participants (cohort)	2.87 (2.00-4.11)
Jick (2006) UK [16]	Active disease	497 cases vs. 1,956 controls (case-control)	0.80 (0.50-1.20)
Shetty (2006) India [17]	Active disease	189 cases vs. 189 controls (case-control)	2.37 (1.00-5.62)
Lienhardt (2005) West Africa [18]	Active disease	822 cases vs. 687 controls (case-control)	2.54 (1.77-3.66)
Wang (2005) China [19]	Active disease	158 cases vs. 316 controls (case-control)	1.54 (1.16-2.04)
Crampin (2004) Malawi [20]	Active disease	598 cases vs. 992 controls (case-control)	1.30 (0.70-2.40)
Ariyothai (2004) Thailand [21]	Active disease	100 cases vs. 100 controls (case-control)	2.70 (1.04-6.97)
Tekkel (2002) Estonia [22]	Active disease	248 cases vs. 248 controls (case-	4.62 (2.44-8.73)

		control)	
Kolappan (2002) India [23]	Active disease	112 cases vs. 553 controls (case-control)	2.24 (1.27-3.94)
Tocque (2001) UK [24]	Active disease	112 cases vs. 198 controls (case-control)	1.46 (0.87-2.47)
Dong (2001) China [25]	Active disease	174 cases vs. 174 controls (case-control)	1.65 (1.00-2.73)
Alcaide (1996) Spain [26]	Active disease	46 cases vs. 46 controls (case-control)	3.60 (1.50-7.20)
Buskin (1994) USA [27]	Active disease	151 cases vs. 545 controls (case-control)	1.30 (0.80-2.10)
Lewis (1963) UK [28]	Active disease	100 cases vs. 100 controls (case-control)	1.01 (0.55-1.85)
Brown (1961) Australia [29]	Active disease	100 cases vs. 100 controls (case-control)	0.95 (0.45-2.02)
Lowe (1956) UK [30]	Active disease	1,200 cases vs. 979 controls (case-control)	1.61 (1.27-2.02)
Gupta (1997) India [31]	Active disease	Unreported number of cases among 707 participants (case-control)	1.38 (0.80-2.39)
Yu (1988) China [32]	Active disease	202 cases among 30,289 participants (case-control)	2.17 (1.29-3.63)
Adelstein (1967) UK [33]	Active disease	96 cases among 76,589 participants (case-control)	3.90 (2.02-7.57)
Shah (1959) India [34]	Active disease	46 cases among 439 participants (case-control)	2.70 (1.37-5.29)
Gupta (2005) India [35]	Mortality	Male: 370 person-years lost among 210,129 Female: 174 lost among 323,316 (cohort)	2.30 (1.68-03.15) (male); 5.92 (2.31-15.17) (female)
Gajalakshmi (2003) India [36]	Mortality	1,840 cases vs. 16,488 controls (urban); 1,529 cases vs. 13,363 controls (rural) (case-control)	4.50 (4.00-5.00) (urban); 4.20 (3.70-4.80) (rural)
Sitas (2004) South Africa [37]	Mortality	414 cases vs. 1,124 controls (case-control)	1.61 (1.23-2.11)
Lam (2001) Hong	Mortality	47 cases vs. 1,480	2.54 (1.25-5.22) (male,

Kong [38]		controls (male, age 35-69); 9 cases vs. 4,930 controls (female, 35-69); 88 cases vs. 2,425 controls (male, >70); 53 cases vs. 4,183 controls (female, >70)	35-69); 1.49 (0.18-12.57) (female, 35-69); 1.63 (1.01-2.64) (male, >70); 1.03 (0.49-2.15) (female, >70)
Liu (1998) China [39]	Mortality	7,916 cases vs. 52,755 controls (male); 4,250 case vs. 34,560 controls (female)	1.20 (1.11-1.30) (male); 1.29 (1.10-1.51) (female)
Ariyothai (2004) Thailand [21]	Passive smoking and active disease	100 cases vs. 100 controls	2.37 (0.94-6.01)
Alcaide (1996) Spain [26]	Passive smoking and active disease	46 cases vs. 46 controls	2.50 (1.00-6.20)

3.2 Regional Parameters and Variables

Unlike the natural history parameters in manuscript Table 1, some components of the model vary with time and by WHO region: the proportion of active TB cases detected (δ), the proportion of detected active TB cases that are successfully treated (κ), the prevalence of smoking ($P_S(t)$, $P_{SL}(t)$, and $P_{ST}(t)$), and HIV prevalence. We estimated these parameters and variables as functions of time by fitting regression models to data from each WHO region.

3.2.1 TB detection and treatment rates

We estimated δ and κ over time by fitting saturating exponential functions to annual case detection and treatment success data for each WHO region from 1994 to 2008, [40 41]:,

(4) $$f(year) = s(1 - e^{-(year - t_0)/\tau})$$

In Equation 4, s is the asymptotic steady-state level (80% for case detection and 90% for treatment, based on WHO estimates of the trajectory of regional rates [42]), t_0 is a parameter that shifts the curve left or right, and τ is the time constant for approaching steady state. Table 3 provides estimates for t_0 and τ for each WHO region and Figure 2 shows the raw data and fits.

TABLE 3: MODELS FOR THE PROPORTION OF CASES DETECTED AND PROPORTION OF DETECTED CASES SUCCESSFULLY TREATED. Regions: AFR: Africa; AMR: Americas; EMR: Eastern Mediterranean, EUR: Europe, SEA: Southeast Asia; WPR: Western Pacific.

Region	*Case detection t_0 (95% CI)*	*Case detection τ (95% CI)*	R^2	*Treatment success t_0 (95% CI)*	*Treatment success τ (95% CI)*	R^2
AFR	1935 (1895-1974)	94 (37-151)	0.55	1982 (1977-1986)	12 (9-16)	0.88
AMR	1967 (1950-1985)	15 (6-23)	0.62	1990 (1987-1993)	6 (4-9)	0.81
EMR	1989 (1982-1995)	20 (9-32)	0.63	1963 (1970-1997)	8 (1-16)	0.44
EUR	1991 (1990-1993)	2 (2-3)	0.90	1930 (1794-2067)	47 (45-139)	0.93
SEA	1963 (1931-1996)	34 (4-64)	0.39	1994 (1992-1995)	6 (4-8)	0.83
WPR	1986 (1978-1995)	15 (5-24)	0.58	1992 (1988-1996)	2 (0-4)	0.55

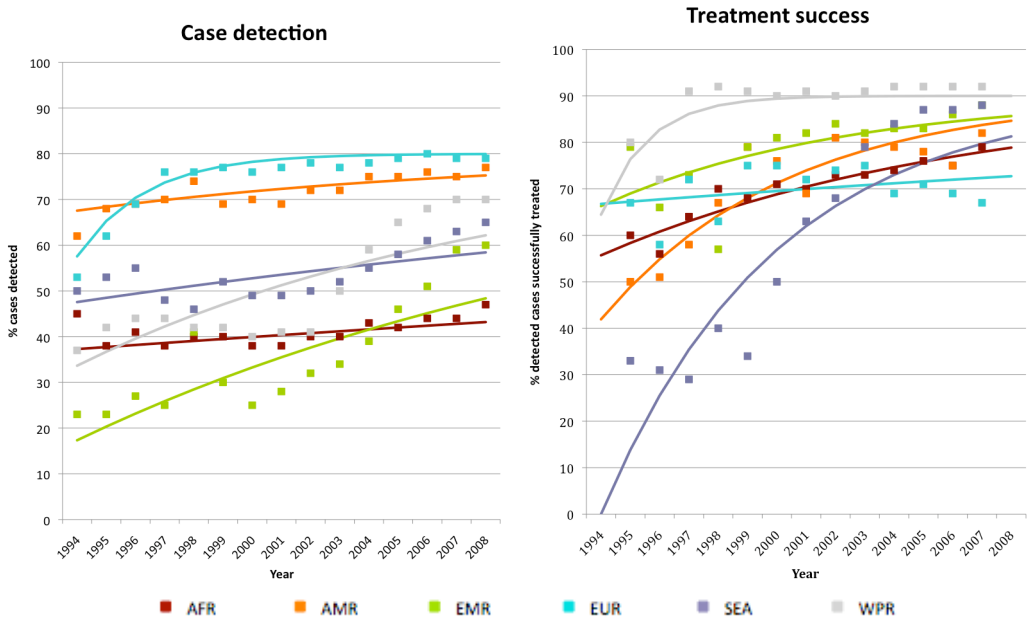


Figure 2: The proportion of cases detected in each WHO region (left) and the proportion of cases successfully treated in each WHO region (right) [43]. A saturating exponential function (solid lines) was fit to data points for each WHO region to estimate the percent of cases detected, δ, , and the percent of cases successfully treated, κ, , at each time in the model simulation for each WHO region (Table 4). These were projected into the future to perform model simulations.

3.2.2 HIV prevalence

HIV prevalence in each region was similarly estimated by regressing a double-logistic equation to United Nations data from 1990 to 2009 [44] (Table 4, Figure 3). This double-logistic equation simulates the initial epidemic rise of the disease, its peak, and then decline to an asymptotic steady-state, and is the model of choice in simulating HIV prevalence as described in the UNAIDS Estimation and Projections Package [45].

$$f(year) = \left[\frac{e^{\alpha(year-t_0)}}{1+e^{\alpha(year-t_0)}}\right]\left[\frac{ae^{-\beta(year-t_0)}}{1+e^{-\beta(year-t_0)}} + b\right] \quad (5)$$

In Equation 5, the parameters are a peak prevalence (a), rate of epidemic rise (α), an asymptotic (steady-state) prevalence (b). a rate of epidemic fall (β) to the steady-state, and a year at which the epidemic reaches half of its peak ($t_{1/2}$).

TABLE 4: STATISTICAL MODELS FOR THE PREVALENCE OF HIV IN EACH WHO REGION.

Region	*Parameter a (95% CI)*	*Parameter α (95% CI)*	*Parameter b (95% CI)*	*Parameter β (95% CI)*	*$t_{1/2}$ (95% CI)*	*R^2*
AFR	0.08 (0.03-0.11)	0.29 (0.21-0.37)	0.05 (0.04-0.05)	0.25 (0.12-0.38)	1995 (1991-1999)	0.99
AMR	0.01 (0.004-0.013)	0.19 (0.06-0.32)	0.006 (0.005-0.0007)	0.25 (0.19-0.31)	1994 (1987-2002)	0.98
EMR	0.0008 (00004-0.0012)	0.29 (0.26-0.32)	0.003 (0.002-0.003)	1.3 (0.3-2.4)	1997 (1997-1998)	0.99
EUR	0.008 (0.007-0.009)	0.287 (0.15-0.42)	0.02 (0.005-0.03)	0.02 (0.015-0.025)	2000 (1954-2047)	0.98
SEA	0.005 (0.003-0.007)	0.27 (0.13-0.41)	0.002 (0.0008-0.003)	0.49 (0.03-0.94)	2001 (1998-2005)	0.92
WPR	-0.003 (-0.1-0.1)	0.27 (0.014-0.03)	0.003 (0.002-0.003)	0.27 (0.21-0.3)	1995 (1971-2018)	0.96

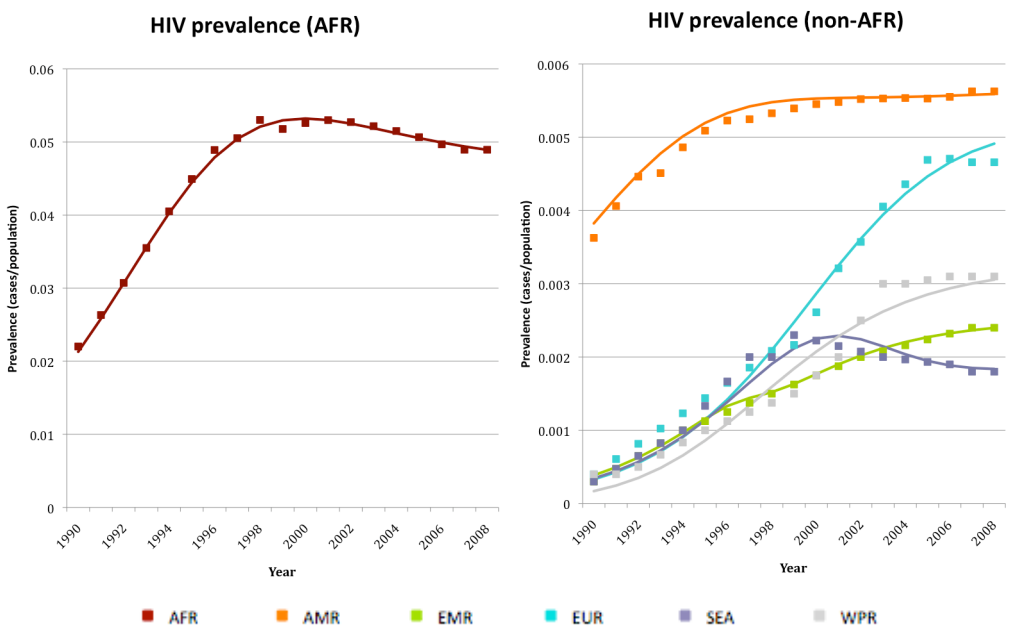


Figure 3: Regional HIV prevalence and fit for (left) the AFR region and all other regions (right; plotted on a different y-axis due to their lower rates). See Table 5 for details.

3.2.3 Smoking prevalence

Because regional smoking prevalence data are available only at scattered times for each WHO region [43], we estimated regional smoking prevalence over time using the UN Food and Agricultural Organization (FAO) data on tons of tobacco consumed annually in each region [46]. Both smoking prevalence (among each gender for persons aged >15 years) and total tobacco consumed are available for each WHO region in 1991 [47]. We calculated the ratio between smoking prevalence and total tobacco consumed per capita in 1991, corrected for gender distribution, then multiplied annual total tobacco consumption by this ratio to estimate annual smoking prevalence (Figure 4).[2]

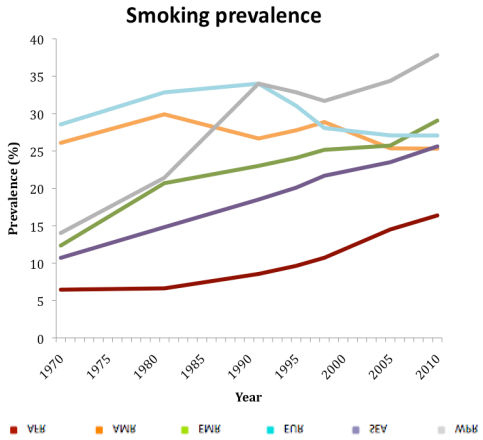


Figure 4: Smoking prevalence in each region, based on tobacco consumption data [48].

4. Smoking scenarios

We created three alternative scenarios for future smoking prevalence trends in each WHO region: (1) a baseline scenario in which smoking rates follow recent trends, (2) a pessimistic scenario in which smoking increases to high levels (an "upper bound"), and (3) an optimistic scenario in which smoking decreases dramatically in response to tobacco control efforts (a "lower bound"). In the *baseline* scenario, smoking rates continue changing at the same annual rate as from 2005 to 2010 (Figure 4) up to a maximum of 40% (the maximum in any overall WHO region to date) [47] (Figure 5a). In the *pessimistic* scenario, smoking prevalence rates rise at twice the global average rate observed from 2005 to 2010, 0.72%/year, up to a maximum of 50% (maximum rate in any country among persons >15 years of age) [47] (Figure 5b). In the *optimistic* scenario, smoking prevalence follows the baseline simulation until 2015, then falls by 1%/year (the rate of decline observed in some U.S. states that have strong tobacco control programs) [49] until reaching 0% (Figure 5c).

[2] The rate of smoking among the latent and active TB population is, given the relative risk of smoking on TB disease, higher than among the general population. We used the estimates from a population-representative study from India (36. Gajalakshmi V, Peto R, Kanaka TS, Jha P. Smoking and mortality from tuberculosis and other diseases in India: retrospective study of 43000 adult male deaths and 35000 controls. *Lancet* 2003;362(9383):507-15.) to estimate that the cross-sectional proportion of active TB cases who are active smokers ($P_{ST}(t)$) was consistently an average of ~2.8 times higher than the proportion of the general population who smoke ($P_S(t)$) (sampling from a triangular distribution with range 2.1-3.6), and in our simulations set $P_{SL}(t) = P_{ST}(t)$ in the absence of further data on latently-infected smokers.

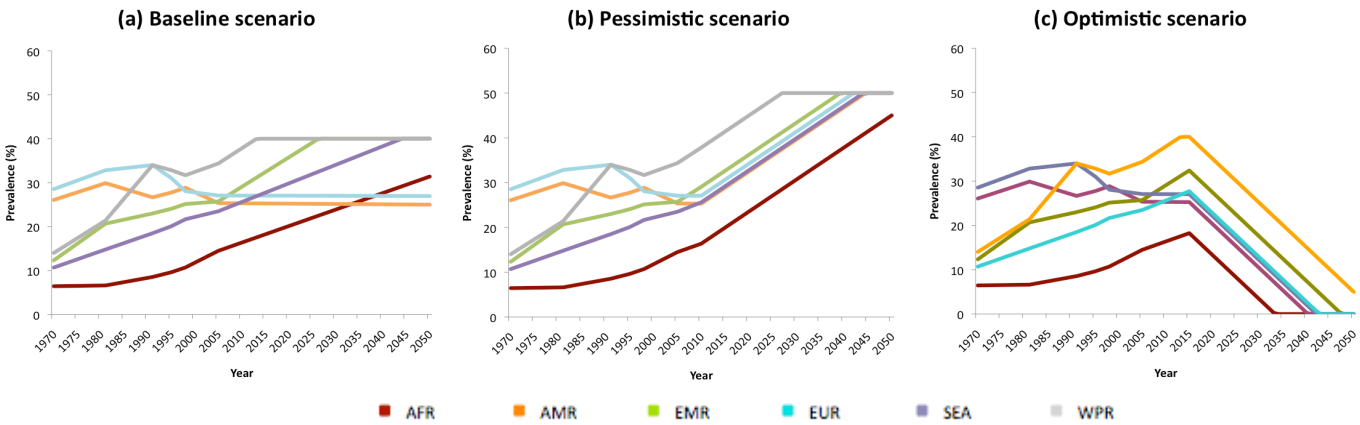


Figure 5: (a) Baseline, (b) pessimistic, and (c) optimistic scenarios for smoking prevalence.

5. Effect of Passive Smoking and HIV-smoking synergy

Studies available to date suggest that passive smoking may increase the relative risk of TB *infection* by a magnitude equal to that of active smokers, but passive smoking does not appear to increase the relative risk of *mortality* among active TB patients [21 26]. To incorporate the higher risk of TB infection among passive smokers, we increased the population size of persons at risk for the higher smoking-related relative risk of TB infection in equations (1)-(3) by replacing the smoking prevalence P_S in the infection rate expression $[1 - P_S(t) + r_{ST}P_S(t)]\,cfT(t)$ with prevalence P_S' that represents the combined active and passive smoking prevalence, yielding equations (1a)-(3a) (this is more parsimonious than creating a separate compartment, but is equivalent because when a separate compartment is added with the same infection risk between passive and active smokers, the two collapse into this more simplified set of equations):

(1a)

$$\frac{dS(t)}{dt} = bN(t) - \{[1 - P_S'(t) + r_{ST}P_S'(t)]cfT(t) + [1 - P_S(t)]\mu + P_S(t)\mu_S\}S(t)$$

(2a)

$$\begin{aligned}\frac{dL(t)}{dt} &= [1 - P_S'(t) + r_{ST}P_S'(t)](1 - p)cfT(t)S(t) \\ &+(\delta\kappa + \sigma)T(t) \\ &-\{v + [1 - P_{SL}'(t) + r_{ST}P_{SL}'(t)]xpcfT(t) + [1 - P_{SL}(t)]\mu + P_{SL}(t)\mu_S\}L(t)\end{aligned}$$

(3a)

$$\begin{aligned}\frac{dT(t)}{dt} &= [1 - P_S'(t) + r_{ST}P_S'(t)]pcfT(t)S(t) \\ &+\{v + [1 - P_{SL}'(t) + r_{ST}P_{SL}'(t)]xpcfT(t)\}L(t) \\ &-\{\delta\kappa + \sigma + [1 - P_{ST}(t)]\mu + P_{ST}(t)\mu_S + [1 - P_{ST}(t) + r_{SM}P_{ST}(t)]\mu_T\}T(t)\end{aligned}$$

(Note that the smoking prevalence in the mortality section of the equations was not changed.) We set P_S' to 1.4 times the value of P_S (sampling, in uncertainty analyses, across a normal distribution with the 95%CI set to 1.1-1.6) based on prior estimates of the typical relative proportion of the population exposed to passive smoke [47].

We also performed a simulation in which we accounted for the possibility that HIV and smoking synergize to increase mortality beyond each of their individual effects—an assertion based on limited cohort data (though these analyses do not find evidence for a smoking-HIV synergy for primary TB infection) [50 51]. We incorporated this potential synergy by increasing the mortality rate among HIV+ smokers to reflect a 7-month shorter lifespan among this group beyond the individual impact of smoking and HIV (varied from 5 to 9 months, [50 52]).

The results of the passive smoking and HIV-smoking synergy simulations are described in the main text, and depicted in Figure 6.

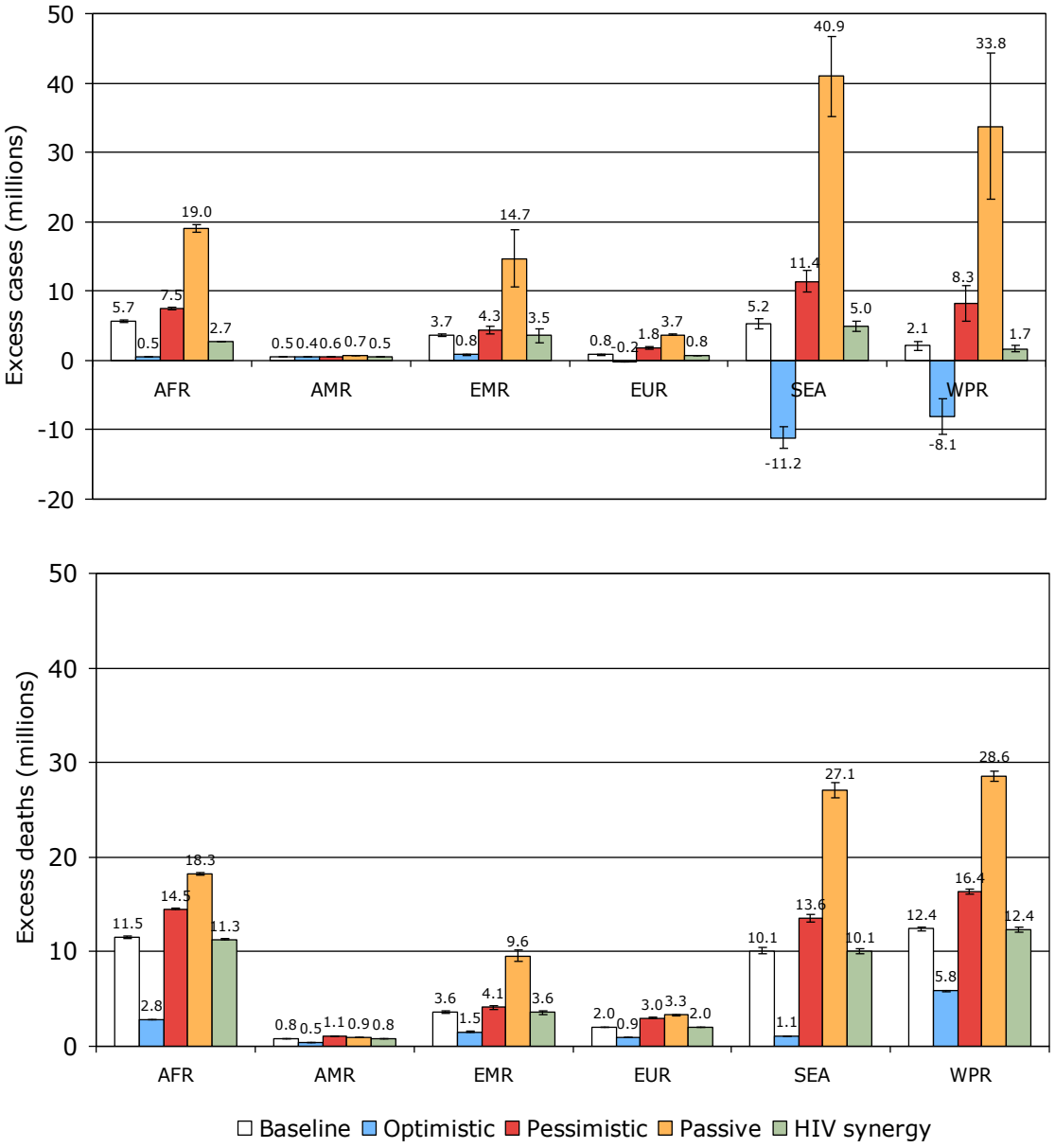


Figure 6: Alternative scenarios demonstrate the potentially large impact of passive smoking on TB cases and deaths. Excess cases and deaths are those that would not have occurred in the absence of smoking (tobacco-attributable cases and deaths from 2010-2050). We depict three alternative smoking trajectories, a scenario incorporating both passive and active smoking ("passive" bar), and a synergistic higher mortality rate among persons affected by both HIV and smoking ("HIV synergy"). Negative bars in the optimistic scenario occur because smoking first produces a lower prevalence due to its high mortality impact (resulting in fewer infectious cases alive to infect others), then tobacco control lowers TB risk for future generations to nearly non-smoking levels. Hence, fewer net cases are produced in the optimistic scenario than in the scenario without smoking, demonstrating the benefits of aggressive tobacco control.

6. Model selection

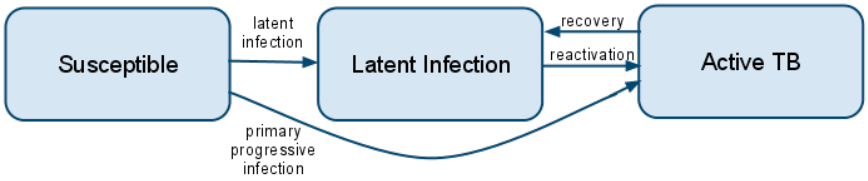
We compared the model to other common models of TB epidemics in terms of the trade-off between model complexity and parsimony. Models with a greater number of parameters may look more "realistic" and will always allow a model to fit data more easily due to more degrees of freedom (more adjustable parameters), but may have poor validity or highly uncertain values [53]. To balance the need for realism in a model against the desire for parsimony, we used "information criteria", which are objective scoring systems to compare a model's fit to given data against the number of parameters in the model used to fit that data. When a model predicts data better using less parameters, it gets a better score, and vice versa.

The most widely-accepted of these scoring systems is the Akaike Information Criterion (AIC) [54 55], defined as:

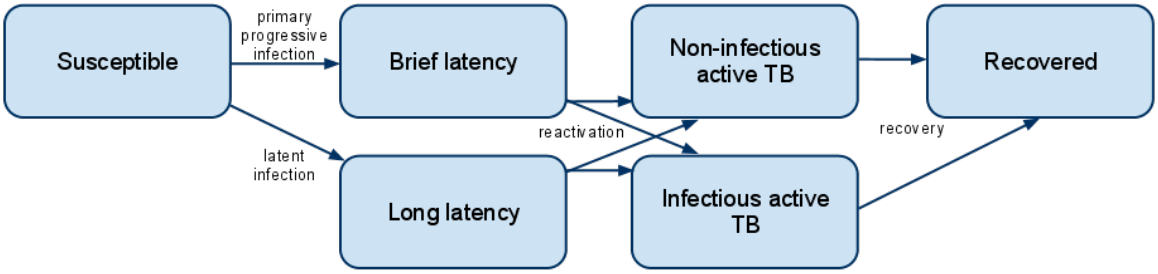
(9) $$AIC = 2k + n\ln(RSS),$$

in which k is the number of parameters being fit, n is the number of data points, and RSS is the residual sum of squared errors between the model fit and data. A lower AIC score is better, and a greater than 10-point difference in scores is considered a statistically significant difference between models [56]. We compared the model in Figure 1 with an earlier, more complex model structure that the first author published for another TB study [57] and a recent WHO model [58] (Figure 7).

Model 1: proposed TB-smoking model (most parsimonious; Figure 1)



Model 2: WHO model (intermediate complexity) [58]



Model 3: model previously published by lead author (most complex) [57]

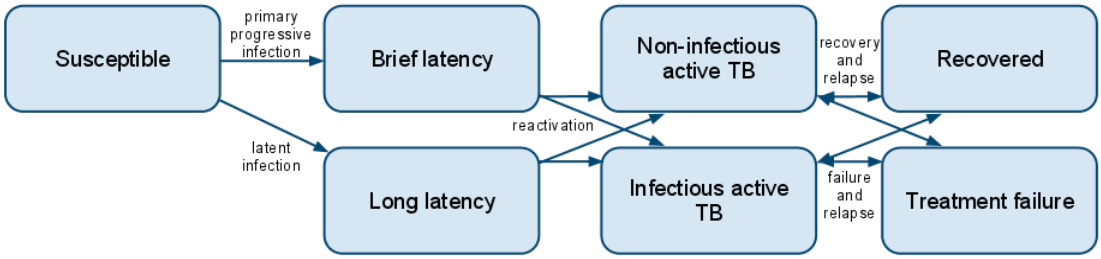


Figure 7: Three alternative models used for the model comparison exercise.

We fit the three alternative models to each WHO regions' annual TB prevalence, incidence and mortality data from 1990 to 2008 [59] using a simulated annealing algorithm in MATLAB version R2010b (Mathworks, Natick, MA) [60]. The algorithm starts with the median value of each natural history parameter distribution and varies the value across its distribution (manuscript Table 1) to find the parameter values that minimize the residual sum of squared errors between the model and data. (We used published equations and parameter distributions for models 2 and 3 [57 58].) Table 5 displays the SEA region result, which produced the smallest difference in *AIC* scores between the models (hence, is the most critical of the winning model). Model 1 has a significantly smaller *AIC* score than the other two models ($\Delta AIC>10$). Table 6 shows that Model 1 provides good fits to the data from each WHO region. We validated the data against annual case notification and annual risk of infection data from each WHO region from 1990 to the present[59] to ensure that all estimates were within 5% error prior to further simulation. Hence, we believe the presented model is improved compared to the two alternatives in that: (1) with regard to validation, the new model is more extensively validated against external data than earlier models; (2) with regard to TB pathogenesis, the new model incorporates recent discoveries about TB by incorporating factors like regression to latency and exogenous reinfection; and (3) with regard to predictive accuracy, the new model is capable of predicting longitudinal trends in every WHO region using fewer uncertain parameters than the older models, as quantified by the objective AIC statistic.

TABLE 5: MODEL COMPARISON FAVORS THE MODEL IN FIGURE 1	
Model # (corresponding to Appendix Figure 7)	*AIC* score (lower number is a better score)
1	302
2	421
3	413

TABLE 6: MODEL 1 FIT TO EACH REGION'S TB DATA	
Region	R^2
AFR	0.95
AMR	0.94
EMR	0.83
EUR	0.83
SEA	0.87
WPR	0.80

7. Uncertainty Analysis

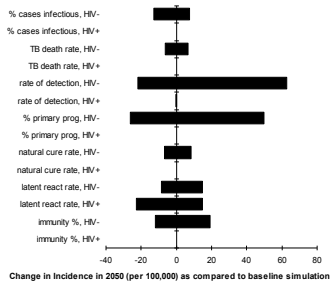
We computed the means and standard errors reported in the main text by performing Monte Carlo simulations with 10,000 replicates, sampling TB natural history parameters from their log-normal distributions listed in Table 2, and using normal distributions built around the mean and 95% confidence intervals for the case detection and treatment rates (Table 3) and HIV prevalence (Table 5). Each parameter value for each simulation run was selected independently of the other parameter values. Smoking prevalence trajectories were fixed for each scenario. All uncertainty and sensitivity analyses were performed in MATLAB version R2010b (Mathworks, Natick, MA).

8. Sensitivity Analysis

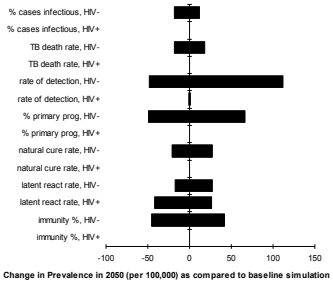
We also performed a univariate sensitivity analysis to examine how varying each parameter across the distributions shown in Table 2 affected the estimated TB incidence, prevalence and mortality predictions from the model for the year 2050 (Figure 8). We used the SEA region for the sensitivity analysis because this region is in the middle of the other regions for smoking prevalence, HIV prevalence, TB case detection, TB treatment success, and TB disease burden.

These analyses revealed that our model behaves similarly to prior models of TB. As with many TB models [2 4 57 61-63], the parameter with the greatest impact on future projections is the percent of newly-infected persons who undergo rapid primary progressive TB (this finding confirms that our model simulates the natural history of TB similarly to the cited prior models). Similarly, we find that the future rate of case detection will have the most significant impact on future TB burden among other exogenous (non-natural-history) parameters, consistent with the expectation that the number of active TB patients treated and rendered non-infectious will have an important impact on the number of future infections. The sensitivity analysis also finds that if our estimate of the relative risk of TB among smokers is low, the incidence of TB could be underestimated by as much as 50 per 100,000 persons, and the estimate of mortality by as much as 30 by 100,000 persons (Figure 8, plots d and f).

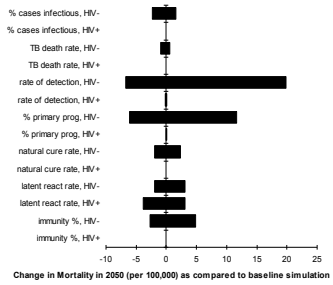
(a)



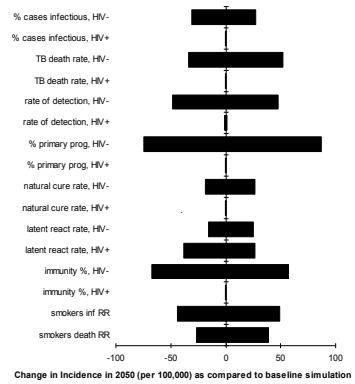
(b)



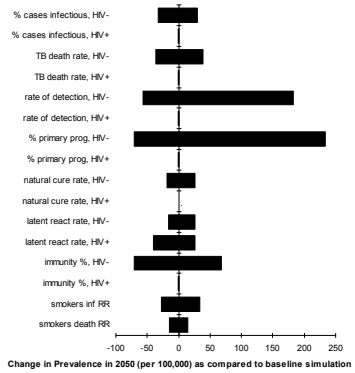
(c)



(d)



(e)



(f)

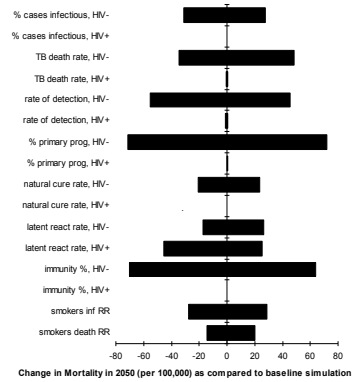


Figure 8: Tornado plots reflecting the sensitivity analysis. Plots a-c reflect a sensitivity analysis performed when smoking is excluded from the model (equivalent to older models of TB epidemics), while plots d-f reflect the full version of the model (employing the baseline smoking scenario).

9. References

1. Blower SM, McLean AR, Porco TC, Small PM, Hopewell PC, Sanchez MA, et al. The intrinsic transmission dynamics of tuberculosis epidemics. *Nat Med* 1995;1(8):815-21.
2. Dye C, Garnett GP, Sleeman K, Williams BG. Prospects for worldwide tuberculosis control under the WHO DOTS strategy. Directly observed short-course therapy. *Lancet* 1998;352(9144):1886-91.
3. Kawai V, Soto G, Gilman RH, Bautista CT, Caviedes L, Huaroto L, et al. Tuberculosis mortality, drug resistance, and infectiousness in patients with and without HIV infection in Peru. *Am J Trop Med Hyg* 2006;75(6):1027-33.
4. Dye C, Williams BG. Criteria for the control of drug-resistant tuberculosis. *Proc Natl Acad Sci U S A* 2000;97(14):8180-5.
5. Iseman MD. *A Clinician's Guide to Tuberculosis*. Hagerstown: Lippincott Williams & Wilkins, 2000.
6. Basu S, Galvani AP. The transmission and control of XDR TB in South Africa: an operations research and mathematical modelling approach. *Epidemiology and Infection* 2008;136(12):1585-98.
7. Basu S, Galvani AP. The evolution of tuberculosis virulence. *Bull Math Biol* 2009;71(5):1073-88.
8. Lin HH, Ezzati M, Murray M. Tobacco smoke, indoor air pollution and tuberculosis: a systematic review and meta-analysis. *PLoS Med* 2007;4(1):e20.
9. Anderson RH, Sy FS, Thompson S, Addy C. Cigarette smoking and tuberculin skin test conversion among incarcerated adults. *Am J Prev Med* 1997;13(3):175-81.
10. den Boon S, van Lill SW, Borgdorff MW, Verver S, Bateman ED, Lombard CJ, et al. Association between smoking and tuberculosis infection: a population survey in a high tuberculosis incidence area. *Thorax* 2005;60(7):555-7.
11. Hussain H, Akhtar S, Nanan D. Prevalence of and risk factors associated with Mycobacterium tuberculosis infection in prisoners, North West Frontier Province, Pakistan. *Int J Epidemiol* 2003;32(5):794-9.
12. Plant AJ, Watkins RE, Gushulak B, O'Rourke T, Jones W, Streeton J, et al. Predictors of tuberculin reactivity among prospective Vietnamese migrants: the effect of smoking. *Epidemiol Infect* 2002;128(1):37-45.
13. Solsona J, Cayla JA, Nadal J, Bedia M, Mata C, Brau J, et al. Screening for tuberculosis upon admission to shelters and free-meal services. *Eur J Epidemiol* 2001;17(2):123-8.
14. McCurdy SA, Arretz DS, Bates RO. Tuberculin reactivity among California Hispanic migrant farm workers. *Am J Ind Med* 1997;32(6):600-5.
15. Leung CC, Li T, Lam TH, Yew WW, Law WS, Tam CM, et al. Smoking and tuberculosis among the elderly in Hong Kong. *Am J Respir Crit Care Med* 2004;170(9):1027-33.
16. Jick SS, Lieberman ES, Rahman MU, Choi HK. Glucocorticoid use, other associated factors, and the risk of tuberculosis. *Arthritis Rheum* 2006;55(1):19-26.
17. Shetty N, Shemko M, Vaz M, D'Souza G. An epidemiological evaluation of risk factors for tuberculosis in South India: a matched case control study. *Int J Tuberc Lung Dis* 2006;10(1):80-6.
18. Lienhardt C, Fielding K, Sillah JS, Bah B, Gustafson P, Warndorff D, et al. Investigation of the risk factors for tuberculosis: a case-control study in three countries in West Africa. *Int J Epidemiol* 2005;34(4):914-23.
19. Wang GJ, Sleigh A, Zhou G, Jackson S, Liu XL. [Nonbiologic risk factors of pulmonary tuberculosis among adults in Henan: a case-control study]. *Zhonghua Liu Xing Bing Xue Za Zhi* 2005;26(2):92-6.
20. Crampin AC, Glynn JR, Floyd S, Malema SS, Mwinuka VK, Ngwira BM, et al. Tuberculosis and gender: exploring the patterns in a case control study in Malawi. *Int J Tuberc Lung Dis* 2004;8(2):194-203.
21. Ariyothai N, Podhipak A, Akarasewi P, Tornee S, Smithtikarn S, Thongprathum P. Cigarette smoking and its relation to pulmonary tuberculosis in adults. *Southeast Asian J Trop Med Public Health* 2004;35(1):219-27.
22. Tekkel M, Rahu M, Loit HM, Baburin A. Risk factors for pulmonary tuberculosis in Estonia. *Int J Tuberc Lung Dis* 2002;6(10):887-94.
23. Kolappan C, Gopi PG. Tobacco smoking and pulmonary tuberculosis. *Thorax* 2002;57(11):964-6.
24. Tocque K, Bellis MA, Beeching NJ, Syed Q, Remmington T, Davies PD. A case-control study of lifestyle risk factors associated with tuberculosis in Liverpool, North-West England. *Eur Respir J* 2001;18(6):959-64.
25. Dong B, Ge N, Zhou Y. Smoking and alchohol as risk factor of pulmonary tuberculosis in Chengdu: a matched case-control study. *Hua Xi Yi Ke Da Xue Xue Bao* 2001;32:104-6.
26. Alcaide J, Altet MN, Plans P, Parron I, Folguera L, Salto E, et al. Cigarette smoking as a risk factor for tuberculosis in young adults: a case-control study. *Tuber Lung Dis* 1996;77(2):112-6.
27. Buskin SE, Gale JL, Weiss NS, Nolan CM. Tuberculosis risk factors in adults in King County, Washington, 1988 through 1990. *Am J Public Health* 1994;84(11):1750-6.
28. Lewis JG, Chamberlain DA. Alcohol Consumption and Smoking Habits in Male Patients with Pulmonary Tuberculosis. *Br J Prev Soc Med* 1963;17:149-52.
29. Brown KE, Campbell AH. Tobacco, alcohol and tuberculosis. *Brit J Dis Chest* 1961;55:150-8.
30. Lowe CR. An association between smoking and respiratory tuberculosis. *Br Med J* 1956;2(5001):1081-6.
31. Gupta BN, Mathur N, Mahendra PN, Srivastava AK, Swaroop V. A study of household environmental risk factors pertaining to respiratory diseases. *Energy Environment Monitor* 1997;13:61-7.
32. Yu GP, Hsieh CC, Peng J. Risk factors associated with the prevalence of pulmonary tuberculosis among sanitary workers in Shanghai. *Tubercle* 1988;69(2):105-12.
33. Adelstein AM, Rimington J. Smoking and pulmonary tuberculosis: an analysis based on a study of volunteers for mass miniature radiography. *Tubercle* 1967;48:219-26.
34. Shah JR, Warawadekar MS, Deshmukh PA, Phutane PN. Institutional survey of pulmonary tuberculosis with special reference to smoking habits. *Indian J Med Sci* 1959;13(5):381-92.

35. Gupta PC, Pednekar MS, Parkin DM, Sankaranarayanan R. Tobacco associated mortality in Mumbai (Bombay) India. Results of the Bombay Cohort Study. *Int J Epidemiol* 2005;34(6):1395-402.
36. Gajalakshmi V, Peto R, Kanaka TS, Jha P. Smoking and mortality from tuberculosis and other diseases in India: retrospective study of 43000 adult male deaths and 35000 controls. *Lancet* 2003;362(9383):507-15.
37. Sitas F, Urban M, Bradshaw D, Kielkowski D, Bah S, Peto R. Tobacco attributable deaths in South Africa. *Tob Control* 2004;13(4):396-9.
38. Lam TH, Ho SY, Hedley AJ, Mak KH, Peto R. Mortality and smoking in Hong Kong: case-control study of all adult deaths in 1998. *BMJ* 2001;323(7309):361.
39. Liu BQ, Peto R, Chen ZM, Boreham J, Wu YP, Li JY, et al. Emerging tobacco hazards in China: 1. Retrospective proportional mortality study of one million deaths. *BMJ* 1998;317(7170):1411-22.
40. Williams BG, Granich R, Chauhan LS, Dharmshaktu NS, Dye C. The impact of HIV/AIDS on the control of tuberculosis in India. *Proc Natl Acad Sci U S A* 2005;102(27):9619-24.
41. World Health Organization. WHO Global Health Observatory (GHO): World Health Organization, 2010.
42. World Health Organization. Global Tuberculosis Control - epidemiology, strategy, financing. WHO/HTM/TB/2009.411. Geneva: WHO, 2009.
43. World Health Organization. WHO Report on the Global Tobacco Epidemic, 2009: Implementing smoke-free environments. Geneva: WHO, 2009.
44. UNAIDS. AIDS Epidemic Update 2009. Geneva: UNAIDS, 2010.
45. Ghys PD, Brown T, Grassly NC, Garnett G, Stanecki KA, Stover J, et al. The UNAIDS Estimation and Projection Package: a software package to estimate and project national HIV epidemics. *Sex Transm Infect* 2004;80 Suppl 1:i5-9.
46. Food and Agriculture Organization of the United Nations. Issues in the Global Tobacco Economy. Rome: United Nations, 2003.
47. American Cancer Society. The Tobacco Atlas. Atlanta: ACS, 2009.
48. Food and Agriculture Organization of the United Nations. Projections of Tobacco Production, Consumption and Trade, 2010.
49. Lightwood JM, Dinno A, Glantz SA. Effect of the California tobacco control program on personal health care expenditures. *PLoS Med* 2008;5(8):e178.
50. Crothers K, Goulet JL, Rodriguez-Barradas MC, Gibert CL, Oursler KA, Goetz MB, et al. Impact of cigarette smoking on mortality in HIV-positive and HIV-negative veterans. *AIDS Educ Prev* 2009;21(3 Suppl):40-53.
51. Marshall MM, McCormack MC, Kirk GD. Effect of cigarette smoking on HIV acquisition, progression, and mortality. *AIDS Educ Prev* 2009;21(3 Suppl):28-39.
52. Nieman RB, Fleming J, Coker RJ, Harris JR, Mitchell DM. The effect of cigarette smoking on the development of AIDS in HIV-1-seropositive individuals. *AIDS* 1993;7(5):705-10.
53. May RM. Uses and Abuses of Mathematics in Biology. *Science* 2004;303(5659):790-93.
54. Akaike H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 1974;19(6):716–23.
55. Bolker BM. *Ecological Models and Data in R*: Princeton University Press, 2008.
56. Gelman A, Carlin JB, Stern HS, Rubin DB. *Bayesian Data Analysis, Second Edition*: Chapman & Hall/CRC, 2003.
57. Basu S, Andrews JR, Poolman EM, Gandhi NR, Shah NS, Moll A, et al. Prevention of nosocomial transmission of extensively drug-resistant tuberculosis in rural South African district hospitals: an epidemiological modelling study. *Lancet* 2007;370(9597):1500-7.
58. Dye C, Williams BG. Eliminating human tuberculosis in the twenty-first century. *J R Soc Interface* 2008;5(23):653-62.
59. World Health Organization. Global TB database: WHO, 2010.
60. Kirkpatrick S, Gelatt CD, Vecchi MP. Optimization by Simulated Annealing. *Science* 1983;220(4598):671-80.
61. Blower SM, Dowlatabadi H. Sensitivity and uncertainty analysis of complex models of disease transmission. *International Statistical Review* 1994;2:229-43.
62. Basu S, Maru D, Poolman E, Galvani A. Primary and secondary tuberculosis preventive treatment in HIV clinics: simulating alternative strategies. *Int J Tuberc Lung Dis* 2009;13(5):652-8.
63. Basu S, Friedland GH, Medlock J, Andrews JR, Shah NS, Gandhi NR, et al. Averting epidemics of extensively drug-resistant tuberculosis. *Proc Natl Acad Sci U S A* 2009;106(18):7672-7.