Published in final edited form as:

J Theor Biol. 2011 November 21; 289: 197–205. doi:10.1016/j.jtbi.2011.08.032.

Modeling socio-demography to capture tuberculosis transmission dynamics in a low burden setting

Giorgio Guzzetta^{a,b}, Marco Ajelli^a, Zhenhua Yang^c, Stefano Merler^a, Cesare Furlanello^a, and Denise Kirschner^d

Giorgio Guzzetta: guzzetta@fbk.eu; Marco Ajelli: ajelli@fbk.eu; Zhenhua Yang: zhenhua@umich.edu; Stefano Merler: merler@fbk.eu; Cesare Furlanello: furlan@fbk.eu; Denise Kirschner: kirschne@umich.edu

^aFondazione Bruno Kessler, Trento, Italy

^bDepartment of Statistics and Mathematics Applied to Economics, Univ. of Pisa

^cSchool of Public Health, University of Michigan, USA

^dDepartment of Microbiology and Immunology, University of Michigan Medical School, USA

Abstract

Evidence of preferential mixing through selected social routes has been suggested for the transmission of tuberculosis (TB) infection in low burden settings. A realistic modelization of these contact routes is needed to appropriately assess the impact of individually targeted control strategies, such as contact network investigation of index cases and treatment of latent TB infection (LTBI).

We propose an age-structured, socio-demographic individual based model (IBM) with a realistic, time-evolving structure of preferential contacts in a population. In particular, transmission within households, schools and work-places, together with a component of casual, distance-dependent contacts are considered. We also compared the model against two other formulations having no social structure of contacts (homogeneous mixing transmission): a baseline deterministic model without age structure and an age-structured IBM.

The socio-demographic IBM better fitted recent longitudinal data on TB epidemiology in Arkansas, USA, which serves as an example of a low burden setting. Inclusion of age structure in the model proved fundamental to capturing actual proportions of reactivated TB cases (as opposed to recently transmitted) as well as profiling age-group specific incidence. The socio-demographic structure additionally provides a prediction of TB transmission rates (the rate of infection in household contacts and the rate of secondary cases in household and workplace contacts).

These results suggest that the socio-demographic IBM is an optimal choice for evaluating current control strategies, including contact network investigation of index cases, and the simulation of alternative scenarios, particularly for TB eradication targets.

© 2011 Elsevier Ltd. All rights reserved.

Correspondence to: Giorgio Guzzetta, guzzetta@fbk.eu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

Individual Based Models; Agent-based Models; epidemiological modeling; endemic diseases; latent TB infection; contact network investigation

1. Introduction

Tuberculosis (TB) is a major global health concern, with an estimated 9 million new cases and 1.7 million deaths worldwide each year [1]. Most TB disease occurs in resource-limited countries, and TB is among the most deadly infectious diseases in the world. The strong interaction of TB with HIV dynamics, and the world-wide emergence of drug-resistant *Mycobacterium tuberculosis* (*Mtb*) strains all make TB an important potential threat for high resource countries as well [1].

TB is transmitted through airborne droplet nuclei, composed of *Mtb* organisms contained in saliva droplets, and emitted by actively infected hosts through sneezing, coughing, and talking. Uninfected individuals sharing the same environment can inhale droplets that leads to the initiation of an immune response that can have one of three different clinical outcomes: complete clearance of the pathogen [2, 3], latent TB infection (LTBI) or progression to primary active disease (about 10% of transmission cases [4]). Primary disease entails the uncontrolled growth of bacteria due to an inadequate immune response by the host. LTBI is characterized by containment of bacteria by the host immune response. Both outcomes, however, lead to the development of granulomas in lungs and other organs. The pathogen is physically contained and immunologically constrained by granulomas, often throughout the lifetime of the individual; however, containment can break down even as long as decades after infection (known as endogenous reactivation), especially in immunocompromised individuals due to malnutrition, aging, or HIV infection [5]. Latently infected individuals may also be re-infected and develop active disease (known as exogenous reinfection). Two billion people worldwide are estimated to be latently infected with *Mtb* and are thus at risk of reactivation [1]. TB can attack different physiological compartments, but the main manifestation in adults, and the only infectious type, is pulmonary TB; therefore, it is this type of TB we focus on in this study.

There is an extensive literature for the past century of compartmental ordinary differential equation (ODE) models describing transmission dynamics of TB, mostly structured as variations of the standard SEIR type (S, susceptible, E exposed, I, infected and R recovered, see [6] for a review of SEIR models). Upon contact with an infectious host, most susceptible individuals move to an exposed compartment (the latent state) before progressing to symptomatic, infectious active disease. However, direct progression from the susceptible to the infectious compartment is often allowed to mimic primary disease; recovery is possible due to antibiotic treatment or natural cure. Recovered individuals can move back to the latent compartment, since treatment is thought to not completely eliminate the bacteria nor to give protection against future exposure. Mathematical models for TB have been used for describing the natural history of the disease [7, 8], control strategy decisions at the global and local levels [9, 10] as well as for evaluating the potential effect of new clinical tools [11, 12]. Recently, the use of individual based models (IBMs, also termed agent based models, ABMs) is emerging in the field of epidemiological modeling [13, 14, 15]. IBMs are particularly useful for investigating the role of spatial behaviors and stochastic processes occurring in transmission dynamics, and for capturing effects of individual variability in disease epidemiology. Few IBMs have to date been proposed for the description of TB epidemiology [16, 17, 18]. While preferential transmission has been empirically demonstrated [19, 20, 21] at least in settings with a low TB prevalence, few models have

attempted to describe the pattern of social contacts in a population for this disease. This feature is particularly important to assess individually targeted control strategies such as those based on tracing of close contacts of index cases. To this aim, we propose an age-structured, socio-demographic IBM with a realistic, time-evolving structure of preferential contacts in the population, and compare it with a simple compartmental ODE model and an age-structured IBM, both with homogeneous mixing. These three models have the same basic epidemiological structure and are parametrized and validated against data derived from a low burden setting. In particular, we focused on the state of Arkansas, representing approximately 1% of the total USA population and characterized by average socio-demographical and epidemiological characteristics and low immigration. This choice was dictated by the availability of published data from a molecular epidemiology study on the proportion of endogenously reactivated over the total TB cases in Arkansas in the years 1997–2003 [22].

2. Materials and Methods

Among the several variants of the SEIR compartmental structure proposed in literature, we chose one, described in Figure 1, largely inspired by [8] and in agreement with previous published IBMs [16, 17, 18].

This structure has the advantage of distinguishing the three forms of active TB (primary, endogenously reactivated and exogenously re-infected). Moreover, it includes two *exposed* compartments, E_0 and E_s . Here, the term *exposed* refers to the subset of close contacts of an index case who have inhaled the droplet nuclei containing tubercle bacilli, known as *adequate contacts* in epidemiological modeling terminology [23]. This terminology is not to be confused with that adopted in epidemiological studies, where all close contacts of an index case are considered exposed, since in practice it is impossible to detect the inhalation of droplets. In this study, exposed compartments track the first several months after acquisition of the pathogen, where the immune status of the individual is not yet clinically defined. The introduction of this compartment is based on the observation that the risk of primary TB, while highest in the first year after infection, is non-negligible over the first 5 years post-exposure [24]. More recently within-host models of the human immune response to TB infection [25, 26, 27] show that the outcome of infection (latency or primary TB) in a naive host is undetermined until a time of at least a hundred days after inhalation of the pathogen. From the exposed or re-exposed state, individuals move to either the latent or the corresponding infectious compartment; however, hosts who inhale the pathogen for the first time may also completely clear the pathogen and revert to a state of susceptibility. Unsuccessful infections are a possible outcome of *Mtb* inhalation [2, 3], but they are generally neglected in epidemiological models, since their frequency is unknown. Therefore, most models estimate a transmissibility parameter that includes both the effect of social encounters and that of individual immunological mechanisms. In this paper we acknowledge the distinction of the two effects in the perspective of integrating immunological and epidemiological dynamics in future models. Since our goal is to capture transmission in a low burden setting (in the USA) where the case detection rate is close to 100%, we also need to account for treatment of individuals. It is not well understood whether treatment kills all bacteria in a host, or simply helps regain containment of bacteria [28, 29]. Treated individuals have been observed to spontaneously reactivate TB (relapse) [30], giving some support to the hypothesis that bacteria may not be completely cleared in treated individuals; on this basis, we assume that treated individuals will revert to a latent state. We also assume that case detection and effective treatment occurs randomly in a population: at each time step, all infectious individuals have the same fixed probability of being cured.

2.1. Modeling TB dynamics

We present three different models based on the described epidemiological structure:

- A. An ODE model with homogeneous mixing, no age structure and constant population size. The population is kept at steady state for simplicity, by using a number of susceptible newborns per year equal to the total number of yearly deaths. Model details are available in Section 1.1 of Supplementary material.
- B. An age-structured stochastic version of the ODE model, implemented as an IBM where transitions between epidemiological states for each individual are dictated by a probability. Homogeneous mixing is still assumed in this model. Age-structure evolves over time as individuals grow older or die according to age-specific mortality rates; a given number of newborns is introduced each year into the susceptible compartment according to estimated fertility rates. Thus, the population size is no longer assumed to be constant in time. Age-dependent risks of developing TB disease have been assumed. Moreover, an age-dependent probability of developing smear positive TB has been introduced to account for differential infectiousness related to smear status. The infectiousness of smear negative TB cases is assumed to be one fourth that of smear positive cases, as estimated from literature [31, 32] and adopted in other models [12].
- C. A socio-demographic IBM, which builds on model B by adding preferential transmission terms through social contacts of individuals, using as reference the one proposed for influenza in [15]. The population is allocated to a spatial grid in such a way to account for observed spatial population density. Each individual belongs to a household and can be assigned, with probability depending on age, to a school, a workplace, or neither. We assume that there are 6 types of schools representing different levels of education: day care, kindergarten, primary, secondary, high school, universities; and 6 types of workplaces, depending on their size (from [15]). Households, schools and workplaces have a location on the spatial grid; assignment to schools or workplaces is done to capture the commuting pattern in the region being modeled (see Section 1.3.3 of Supplementary material). Infectious individuals can thus transmit the pathogen to other susceptibles either at their home or at their school or workplace. A component of casual transmission to the general population is also maintained, with a probability of contact that decreases with distance between the residences of the infectious and susceptible individuals. In fact, it has been found that isolates from unrelated individuals have a higher probability to be genotypically identical if the cases live at close distance, at least in urban settings (J. H. Bates, Arkansas Department of Health, unpublished observations). This, as well as other published studies [33], supports the hypothesis that significant transmission occurs in public, closed environments such as theaters, public transportation vehicles, bars, stores and so on. Since TB is endemic and has very long dynamics [7], we account for the evolution in time of the population's socio-demographic structure. We followed the approach taken in [34], where an IBM for an endemic disease (Hepatitis A) accounting for socio-demography has been proposed for the first time. According to its rules, all individuals can die of natural death with an age-specific probability, and they can be born into households where a couple of individuals of suitable age for parenting are present; new households can be formed if an individual moves out of its original household to a new residence or to form a couple with another individual from another household (marriage), or when a couple splits (divorce). Individuals in school are re-assigned to a different school only when they change their educational level, according to their age. Individuals in workplaces can become unemployed with a given probability, and unemployed individuals can be assigned to a new workplace.

Assignments to a new school or workplace are done in such a way to maintain known commuting patterns. An update of the population is calculated at the end of each year. We refer the readers to Section 1.3 and 1.5 of Supplementary material for further details on Model C implementation and Section 1.4 for a validation of the socio-demographic evolution.

2.2. Transmission dynamics in models B and C

In models B and C, any susceptible or latently infected individual i at any time t has a probability $p_i = 1 - e^{-\Delta t \lambda_i(t)}$, (with $\Delta t = 1$ week the time step of the simulations) of inhaling pathogen and thus moving to the corresponding exposed compartment. The probability of this event depends on the instantaneous risk of infection $\lambda_i(t)$, computed at any time of the simulation (as in [35, 34, 15]).

- For model B, the risk of infection for each individual is defined as

$$\lambda_i = \sum_{\{k=1, \dots, N\}} \beta_r^{(B)} \frac{I_k s_k}{N}$$

where:

- $\beta_r^{(B)}$ (in years⁻¹) is the transmission rate.
- N is the total population at a given time.
- $I_k = 1$ if individual k is infectious (has active TB), 0 otherwise.
- s_k is the relative infectiousness of individual k ; in particular, $s_k = 1$ if the infectious individual is a smear-positive case, $s_k = 0.25$ if it is smear-negative [31, 32, 12].
- For model C, the risk of infection for each individual is defined as the sum of the risk factors coming from the different sources of infection considered, namely:
 1. contacts with infectious members of the household (first term in Eq. 1),
 2. contacts with infectious individuals attending the same school or workplace, if any (second term in Eq. 1),
 3. random contacts in the population (third term in Eq. 1).

$$\lambda_i = \sum_{\{k=1, \dots, N | H_k = H_i\}} \beta_H \frac{I_k s_k}{n_i} + \sum_{\{k=1, \dots, N | P_k = P_i\}} \beta_P \frac{I_k s_k}{m_i} + \sum_{\{k=1, \dots, N\}} \beta_r^{(C)} \frac{I_k s_k f(d_{ik})}{\sum_{\{k=1, \dots, N\}} f(d_{ik})} \quad (1)$$

The terms in Eq. 1 are defined as follows:

- β_H (expressed in years⁻¹) is the within-household transmission rate.
- H_i is the index of the household where individual i lives in and n_i is the household size.
- β_P (in years⁻¹) is the within-school/workplace transmission rate.
- P_i is the index of the school/workplace where individual i studies/works (depending on the employment of i) and m_i is the school/workplace size (if any).
- $\beta_r^{(C)}$ (in years⁻¹) is the transmission rate through casual contacts.

- $f(d_{ik})$ is the function defined in Eq. 1 in Supplementary material. It makes the casual transmission of TB in the general community explicitly dependent on distance through patterns of commuting ([14, 35, 34, 15]).

In both models B and C, individuals may become infectious by progressing to primary disease or exogenous re-infection from the exposed compartments, or to endogenous reactivation TB from the latent compartment, according to age-specific risks. Such risks were all described as functions of age determined by two parameters:

- the risk of primary disease following infection by age, $p(a)$, was assigned a functional form inspired by the piecewise linear used in [8] and originally estimated on epidemiological data from Norway. It is assumed to have a constant, low value, p_c , in children (0–10 years), a constant, typically higher value, p_a , in adults (≥ 20 years), and a linear growth in between, in such a way to be continuous (Figure 9 in Supplementary material);
- the protection from re-infection disease by age, $\sigma(a)$ was assumed to have the same functional form, with analogous parameters σ_c and σ_a . The age-specific risk of progression to TB disease by exogenous re-infection for re-exposed individuals is thus given by $p(a)(1 - \sigma(a))$;
- the functional form for the risk of endogenous reactivation $r(a)$ was assumed to be a class C^1 function, growing linearly for the first 50 years and quadratically for the last 50 years (Figure 10 in Supplementary material). Its two parameters modulate respectively the average reactivation risk over ages (r), and the slope of the linear growth in younger ages (r_m). Given r , r_m forces the value of concavity of the parabola in older ages, thus acting as a shape factor for the functional form. This ad-hoc choice for the functional form of the reactivation risk is more plausible than the piecewise linear used in [8]: in fact, the latter form assumes a constant risk of reactivation after the age of 20, in contrast with the biological hypothesis of an increased risk of reactivation with the decaying immunological response in aged individuals [5]. Moreover, this functional form allowed us to better capture the age-specific TB incidence in the modeled setting (see Section 3).
- The age-dependent probability of developing smear positive TB, $f_+(a)$, was approximated with another piecewise linear function (Figure 12 in Supplementary material) following [8] and using estimates from epidemiological data [22].

2.3. Parametrization

Parametrization of the models was obtained by searching for the minimum of a score function F , calculated as the combined error on the following variables:

- Longitudinal estimates on the prevalence (p), incidence (i) and mortality (m) per 100,000 population of active TB in the USA in years 2001–2007, available from the World Health Organization (WHO) [36];
- Average proportion of reactivated TB cases over the total (rf) in years 2000–2003, as measured in a published molecular epidemiology study [22]
- (for models B and C only) Age-group specific incidences (ai) in years 2001–2007 from U.S. born cases in Arkansas, provided by the Online TB Information System (OTIS) of the U.S. Center for Disease Control [37].

The relative errors between model output (in capital letters) and data (lowercase letters) were combined according to Eq. 2, to produce the score function F_A for Model A.

$$F_A = \frac{\|P - p\|_2}{\|P\|_2} + \frac{\|I - i\|_2}{\|I\|_2} + \frac{\|M - m\|_2}{\|M\|_2} + \left| \frac{RF - rf}{RF} \right| \quad (2)$$

Here, $\|\cdot\|_2$ represents the Euclidean norm of the corresponding array.

For Models B and C, we added to Eq. 2 a term accounting for the error between the simulated and observed age-specific incidence profiles over 5 age groups, as shown in Eq. 3 ($F_{B,C}$ is thus the score function for both model B and C).

$$F_{B,C} = F_A + \frac{\|AI - ai\|_2}{\|AI\|_2} \quad (3)$$

2.4. Parameter space exploration

Some of the model parameters could be precisely estimated from epidemiological data, and were kept fixed throughout all simulations (Table 1 - see Section 2.1 and 2.2 of Supplementary material for details on the estimation of these parameters). An exception is parameter χ (proportion of cleared infections) for models B and C, affecting the number of latently infected individuals and therefore the number of reactivations; while no information are available from literature on this parameter, we expect it to have a negligible effect in the time-span of 7 years considered in this study, since very few recently infected individuals will have the chance to reactivate. Therefore, we fix its value in models B and C to the best fitting value suggested by the calibration of Model A (see Table 2).

All the other model parameters (in number of $Z_A = 7$, $Z_B = 7$ and $Z_C = 9$ for respectively model A, B and C) were left free to vary over a range during the parametrization procedure. For models B and C, we introduce the following notation: we term P_1 the subset of free parameters whose ranges could be estimated from epidemiological observations, and P_2 the set of remaining free parameters, which were explored over a broad range. In particular, P_2 includes the transmissibilities ($\beta_r^{(B)}$ for model B and $\beta_H, \beta_P, \beta_r^{(C)}$ for model C) and the shape factor of the reactivation risk by age (r_m for both models). Therefore, the cardinality of P_2 is $|P_2^{(B)}| = 2$ for model B and $|P_2^{(C)}| = 4$ for model C. Sections 2.1 and 2.2 in Supplementary material report a detailed description of all model parameters, and the estimation of their ranges with the corresponding reference studies.

The parameter space was explored using the Latin Hypercube Sampling (LHS) method [38]. LHS allows an efficient sampling of the parameter space, as it requires a smaller sample size Q than uniform sampling to achieve the same accuracy [38].

For model A, simulations were run with $Q_A = 100,000$ combinations of the $Z_A = 7$ free parameters. The parameter set obtaining the minimum score F_A was selected (see Table 2).

IBMs are computationally more intensive than ODEs, and intrinsically stochastic. Several realizations of the model with the same parameter set are required to obtain stable results with respect to random variability: in fact, each result presented in the paper for IBMs is based on 100 realizations of the same experiment. The large sample size used for model A is thus unfeasible for models B and C, and the search of best fitting parameters was performed according to the following scheme:

1. Run a global LHS sampling with $Q=10,000$ on all free parameters of model B ($Z_B=7$).
2. Run a local LHS search around the best fitting parameter set, i.e., fix model parameters from the subset P_1 and search only on parameters from P_2 . Since model B is a special case of model C (with $\beta_H, \beta_P = 0$), a local search starting from the same minimum can be done for model C as well. The resulting parameter spaces have a dimension of respectively $Z_B^{local}=2$ and $Z_C^{local}=4$ and are explored with a sample size of $Q^{local} = 500$. The ranges of free parameters in the local search are reduced, based on indications from best fitting simulations in the global search (see Section 2.4 in Supplementary material for details).

The best fitting parameter set for both models B and C after the local searches are reported in Table 2 and discussed in Section 3.

2.5. Model initialization

The three models were initialized with prevalence data from the USA in 2000, the same epidemiological year as the census data used for the socio-demographic initialization. The initial number of prevalent cases was set to 5.61 per 100,000, as reported for the USA by the World Health Organization (WHO) [36]. The initial number of individuals with LTBI was estimated to an average 4.2% of the general USA population in 1999–2000 [39]; age-specific latent prevalences for five age groups (0–14, 15–24, 25–44, 45–64, 65–100) are also provided in the same study and were used as initial values in the age-structured IBMs. The initial number of latent cases was distributed randomly in the population, as no data is available about their clustering with respect to households, schools or workplaces in low-burden settings. Finally, the initial number of exposed individuals was initialized in such a way to give a correct estimate of incidence at the end of year 2000 (see Section 2.3 in Supplementary material), and the corresponding times since *Mtb* infection were chosen from a uniform distribution with range 0 – 5 years.

3. Results

Table 2 reports the estimated best fitting parameter sets for the three models. A discussion of variability of best fitting parameters from P_1 and P_2 in models B and C is given in Section 2.4 of Supplementary material.

The estimated risk of primary disease upon infection p (average of $p(a)$ over age for the IBMs) was about 15% for all models, slightly above the 5–10% risk of primary disease widely suggested in the literature [4, 8]. Best fitting parameters of the reactivation rate for models B and C suggest an increased risk (respectively 7.4 and 8.3 times) for individuals older than 50 years old, as compared to the general population. A 3.8-fold risk was also found for the same age group in a longitudinal study within a small USA community [40].

The transmissibility parameter, β_R , of model A was significantly smaller than that of models B and C. Calculating the Annual Risk of TB Infection (ARI) in year 2000 by the formula:

$$ARI(2000)=\beta_R^{(A)}(1-\chi)I(2000)$$

model A predicts a value of 0.003%, which is remarkably low when compared to the estimate of 0.03% in 1995 by [41, 42]; although TB incidence and prevalence have constantly declined in the USA between 1995 and 2000 [37], a 10-fold reduction of the ARI in this short time window seems unrealistic. The underestimation of transmission in model

A can be attributed to the absence of an age structure. In fact, not accounting for the age-dependent heterogeneity in the reactivation risk produces an overestimation of reactivated incident cases; therefore, in order to capture incidence values in the time period considered, model A compensates by reducing the effect of recent transmission.

For calculating ARI using model B, we need to consider the difference in infectiousness related to smear status. The proportion of smear-negative TB cases in Arkansas is approximately $f_- = 70\%$ [22]; considering the relative infectiousness of smear negative and smear-positive individuals ($s_- = 0.25$, $s_+ = 1$), the predicted ARI becomes:

$$ARI(2000) = \beta_R^{(B)} (1 - \chi) \sum_{z=+,-} f_z s_z I(2000)$$

(with $f_+ = 1 - f_-$), resulting in an estimated ARI of 0.014% in year 2000, in the same order of magnitude of the cited value of 0.03% [41, 42]. A calculation of the ARI in model C cannot be done due to the heterogeneous mixing introduced by the social structure. However, the best fitting transmissibility values in the three routes, taken together, have the same order of magnitude of model B: therefore, we expect the ARI in this model to be close to actual values.

Data on prevalence, incidence and mortality per 100,000 individuals in the years 2001–2007 for the USA show a steady, linear decrease in the years considered. Prevalence drops from 3.96 to 3.15 cases per 100,000 individuals, incidence from 5.29 to 4.22 new cases per 100,000 individuals per year and mortality from 0.53 to 0.42 cases per 100,000 individuals, with an the average yearly percentage drop of about 3.7% for all indicators [37]. All models capture this pattern (Figure 2), with relative errors staying below $\pm 15\%$ at all time points for all variables. The improvement in fit between models C and B was found to be significant by means of an F-test for nested models [43] ($F(2, 17) = 8.424$, $p = 2.87 \cdot 10^{-3}$). For the two IBMs, the average values over 100 runs with the 95% bootstrap confidence intervals are plotted.

If we examine the proportions of endogenously reactivated cases (Table 3), model A overestimates the reported proportion of 68.5% reactivated cases [22], confirming the previous observation that the lack of age structure overpredicts numbers of reactivations. Models B and C are equally able to account for the measured reactivation fraction, with confidence intervals that overlap with data.

Figure 3 compares data on age-specific incidences of U.S. born cases in Arkansas for years 2001–2007 [37] with corresponding predictions by models B and C (model A can not account for this variable, since it lacks an age structure). Data show a growth in the specific incidence for older age groups, rising from 1.13 cases per 100,000 in children under 14 to 9.32 per 100,000 in elderly people over 65. This trend can be explained with the observation that almost 70% of active TB cases in Arkansas are due reactivation (Table 3), and that the reactivation risk increases with age. Both IBMs produce a good estimate of age-specific incidence.

In addition to the other two models, model C is able to track down patterns of transmission within households and workplaces. We use data from three different large-scale studies where contacts of active TB cases were screened for TB infection by means of Tuberculin Skin Test (TST) administered at least 10 weeks after the last contact with the index case. In [19], 6,225 contacts of 1,080 cases diagnosed between July 1996 and June 1997 in 11 U.S. based TB programs were investigated. 2,664 were household contacts and 747 were workplace contacts. Another study [20] investigated workplace transmission for 724

contacts of 42 cases occurring in 1996 in 5 state TB control programs in the U.S. The largest study available [21] involved 26,542 contacts of 3,485 cases between 1990 and 2000 in British Columbia, Canada. The fraction of TST positive individuals was taken as a proxy of the proportion of adequate contacts for transmission of the pathogen; data from [19, 21] were used for household contacts and from [20, 21] for workplace contacts. The fraction of secondary active TB cases was also reported in [21]. These data were not used to calibrate best fitting parameters: instead, they are compared with model predictions as a validation of the plausibility of the model. Table 4 reports the estimated infection rates (proportions of contacts who developed LTBI or primary disease) and the proportion of secondary active TB cases for household and workplace contacts in data and in the model. According to [21], the rate of secondary cases in households is over 4 times the rate in workplaces, whereas the ratio of infection rates is approximately 2 times (see Table 4). Only contacts who did not receive isoniazid treatment for LTBI are considered; the secondary TB rate is expected to be lower in the presence of LTBI treatment. One explanation for the increased risk of secondary cases in households could be common susceptibility to TB disease within households, due to shared genetic and environmental risk factors (such as poor hygiene, malnutrition, passive smoking, etc.). Table 4 shows good agreement for all variables except for the infection rate in work-places, which is underestimated. Our model does not account explicitly for a differential risk of infection in households and in workplaces. This explains partly its inability to capture all the variables in Table 4. For instance, if the model predicted correctly the infection rate in workplaces, it would also overestimate the rate of secondary cases in the same setting due to the setting-independent risk of infection.

A significant difference ($p < 10^{-2}$) in infection rates among contacts of smear positive versus smear negative cases was found in contact investigations at workplaces [20]. Model C predicts respectively an average 14% and 7% infections among contacts of smear positive and smear negative cases in workplaces (significantly different, $p \ll 10^{-2}$).

Lower infection rates have also been suggested for workplaces of larger size: the majority of workplaces with infection rates beyond 30% had less than 6 employees, and only a 9% TST positivity was found in the largest workplace (124 contacts) investigated in [20]. Model C reproduces the proposed pattern, as shown in Figure 4. Infection rates around 16% are predicted in small workplaces (size ≤ 10), decreasing steeply to less than half (6%) for large workplaces (size ≥ 100).

4. Discussion

Understanding patterns of TB transmission in different social settings is key to assessing common TB prevention strategies, such as contact network investigation of index cases and LTBI treatment. To this purpose, we proposed a computational model for describing TB infection dynamics in an epidemic setting. The model we developed is an IBM and includes age-structure, socio-demographic dynamics and preferential transmission within households, schools, workplaces as well as distance-dependent casual contacts. The model better describes recent surveillance and molecular epidemiological data when compared to a simple ODE and an age-structured IBM with homogeneous mixing, based on the same epidemiological structure proposed by Vynnycky and Fine [8]. Furthermore, it was able to predict contact tracing investigation data from the USA and Canada.

Comparison of the three models allows us to isolate the contribution of incremental epidemiological features in improving model predictability. While the simple ODE model (Model A) is generally able to capture the decline in time of prevalence, incidence and mortality, it overestimates the fraction of reactivated cases, assigning a small role to recent transmission, in contrast with data [22]. This is important from the public health standpoint

because TB cases occurring from recently acquired infections can be prevented through contact tracing and timely LTBI treatment in people who have recently been infected [44]; the failure to recognize the contribution of recent transmission to TB burden in a population can facilitate the spread of TB in a population.

In contrast, the homogeneous, age-structured IBM (Model B) was able to correctly describe the time course of prevalence, incidence and mortality, and also gave correct estimates for fraction of reactivated cases. Moreover, it fit well the age-specific incidence of active TB. Thus, the inclusion of age-structure in the model plays a fundamental role in capturing relevant characteristics of TB epidemiology and can not be neglected in models that are intended to be used for evaluating control strategies.

The most sophisticated model, the socio-demographical IBM (Model C), could fit the same data better than the homogeneous IBM. An interesting added value was its ability to predict realistic values for infection rates within households and the proportion of secondary cases in households and work-places. The evidence for preferential mixing shows that an accurate model for socio-demographic transmission processes in the study setting is key to the quantitative evaluation of individually targeted control strategies, such as contact tracing and treatment of LTBI for exposed individuals [44]. Our results confirm the importance of using IBMs to model the effects of contact network investigation and preventive treatment, highlighted in a recent review of TB models [45].

Previous studies have used IBMs to understand the impact of preferential mixing on specific aspects of TB epidemiology. Cohen and colleagues [17] use a contact network IBM to investigate the effect of social interactions on the proportion of exogenous re-infection in low and middle TB prevalence settings. They found that re-infection can be locally important if the force of infection is not homogeneously distributed in the population. In our simulation, the proportion of cases due to exogenous re-infection is very low, because of the loose connectedness of the social network (e.g. small household sizes) and the low level of transmission in very low burden settings. Also, in our model initial cases and individuals with LTBI were uniformly distributed over the Arkansas territory. However, it is possible that under different assumptions re-infection could play a bigger role; for instance, modeling immigration could produce local areas of higher prevalence for latent and active TB where re-infection could actually play a role. Another socio-demographic IBM has theoretically investigated the determinants of TB cluster sizes testing the hypothesis regarding multiple circulating strains with identical transmissibility [16].

In this study, we parametrized our IBM to a specific low burden setting, reproducing quantitative age-structured and longitudinal surveillance data and predicting, to some extent, data on transmission for two relevant routes of infection, households and workplaces. We focused on transmission in low burden settings, where we can neglect the effect of HIV co-infection, due to its low prevalence (only 3.3% of TB cases in Arkansas affected HIV positive individuals between 2001 and 2007 [37]).

The parametrization was done in such a way to allow comparison with simpler models, incrementally adding relevant features of TB transmission dynamics. In this way, we could evaluate the contributions of the incremental features to the explanation of different aspects of TB epidemiology. Fitting IBMs is a difficult task, since a thorough exploration of high-dimensional parameter spaces is unfeasible, due to the computational burden and intrinsic stochasticity of such models. In this study, we do not claim that values for the best fitting parameters are biologically meaningful. However, we evaluate their plausibility, where possible, against available estimates from data.

For the purpose of control strategy assessment, a few issues will need to be addressed. First, a fit of transmission data in households and workplaces needs to be done, in order to maximize the plausibility of estimated parameters. In this work, no fit was done for such data. Second, we need to explore determinants of increased risk of secondary disease for household contacts, in order to evaluate the real impact of household contact tracing in reducing the TB burden. Familial risk factors such as genetic susceptibility, increased initial inoculum of bacteria due to smaller, less ventilated environments, and shared environmental conditions (e.g. malnutrition, low hygiene and passive smoking) may explain the increased risk of secondary TB in households. In particular, epidemiological studies found evidence that African-American populations have an increased susceptibility to TB disease [46]; about 16% of the Arkansas population belongs to this ethnicity, and the relative TB incidence is indeed higher than in the general population [37]. The effect of a genetic component of susceptibility was previously studied for TB under the assumption of homogeneous mixing [47, 48].

A limitation of this study concerns the exclusion of socio-demographical processes related to the immigration of individuals from high burden countries. These effects are negligible on the demographic component of our model. However, in terms of TB dynamics, incident cases from foreigners constitute more than half of the total in the United States (58% in 2008) [37]. In Arkansas, the context of our analysis, the proportion of cases from foreign born is still a minority (15.7% on average in the period 1998–2007), but rising (from 11.5% in 1998–2002 to 21.5% in 2003–2007) [37]. Immigrants from countries with high TB prevalence may live in more crowded households, often clustered in neighborhoods, and have increased risk factors (e.g., lower income, higher LTBI prevalence) and therefore feature a specific epidemiology. Identifying mixing patterns of the immigrant and autochthonous population needs careful analysis, which we leave to a separate study. Transmission of *Mtb* in other important social settings, such as within hospital wards, will also be addressed in future versions of the IBM.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to thank Dr. Joseph H. Bates and Dr. Leonard N. Mukasa from the Arkansas Department of Health for their assistance with data retrieval and interpretation, useful comments on the model, and careful reading of the manuscript. Additionally, we would like to thank Dr. Piero Manfredi from the University of Pisa for his remarks and suggestions.

References

1. World Health Organization. Global Tuberculosis Control, Technical Report. 2010.
2. Roach DR, Bean AGD, Demangel C, France MP, Briscoe H, Britton WJ. TNF Regulates Chemokine Induction Essential for Cell Recruitment, Granuloma Formation, and Clearance of Mycobacterial Infection. *Journal of Immunology*. 2002; 168:4620–4627.
3. Bhatt K, Salgame P. Host Innate Immune Response to Mycobacterium tuberculosis. *Journal of Clinical Immunology*. 2007; 27:347–362. [PubMed: 17364232]
4. Stead WW, Lofgren JP, Warren E, Thomas C. Tuberculosis as an endemic and nosocomial infection among the elderly in nursing homes. *New England Journal of Medicine*. 1985; 312:1487.
5. Kaufman SHE. Protection against tuberculosis: cytokines, T cells, and macrophages. *Annals of the Rheumatic Diseases*. 2002; 61:ii54–ii58. [PubMed: 12379623]
6. Anderson, RM. Infectious diseases of humans: dynamics and control. Oxford University Press; 1992.

7. Blower SM, McLean AR, Porco TC, Small PM, Hopewell PC, Sanchez MA, Moss AR. The intrinsic transmission dynamics of tuberculosis epidemics. *Nature Medicine*. 1995; 1:815–821.
8. Vynnycky E, Fine PEM. The natural history of tuberculosis: the implications of age-dependent risks of disease and the role of reinfection. *Epidemiology and Infection*. 1997; 119:183–201. [PubMed: 9363017]
9. Dye C, Garnett GP, Sleeman K, Williams BG. Prospects for worldwide tuberculosis control under the WHO DOTS strategy. *Lancet*. 1998; 352:1886–1891. [PubMed: 9863786]
10. Dye C, Williams BG. Eliminating human tuberculosis in the twenty-first century. *Journal of the Royal Society Interface*. 2008; 5:233–243.
11. Salomon JA, Lloyd-Smith JO, Getz WM, Resch S, Sanchez M, Porco T, Borgdorff MW. Prospects for advancing tuberculosis control efforts through novel therapies. *PLoS Medicine*. 2006; 3:e273. [PubMed: 16866578]
12. Abu-Raddad LJ, Sabatelli L, Achterberg J, Sugimoto J, Longini I, Dye C, Halloran M. Epidemiological benefits of more-effective tuberculosis vaccines, drugs, and diagnostics. *PNAS*. 2009; 106:13980–13985. [PubMed: 19666590]
13. Longini IM, Halloran ME, Nizam A, Yang Y. Containing Pandemic Influenza with Antiviral Agents. *American Journal of Epidemiology*. 2004; 159:623–633. [PubMed: 15033640]
14. Ferguson NM, Cummings DAT, Cauchemez S, Fraser C, Riley S, Meeyai A, Iamsrithaworn S, Burke DS. Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature*. 2005; 437:209–214. [PubMed: 16079797]
15. Merler S, Ajelli M. The role of population heterogeneity and human mobility in the spread of pandemic influenza. *Proceedings of the Royal Society B*. 2010; 277:557–565. [PubMed: 19864279]
16. Murray M. Determinants of cluster distribution in the molecular epidemiology of tuberculosis. *PNAS*. 2002; 99:1538–1543. [PubMed: 11818527]
17. Cohen T, Colijn C, Finklea B, Murray M. Exogenous re-infection and the dynamics of tuberculosis epidemics: local effects in a network model of transmission. *Journal of the Royal Society Interface*. 2007; 4:523–531.
18. France, AM. PhD thesis. University of Michigan; 2008. Integrating molecular typing into routine tuberculosis surveillance: an assessment of the strengths and limitations of current approaches.
19. Marks SM, Taylor Z, Qualls NL, Shrestha-Kuwahara RJ, Wilce MA, Nguyen CH. Outcomes of Contact Investigation of Infectious Tuberculosis Patients. *American Journal of Respiratory and Critical Care Medicine*. 2000; 162:2033–2038. [PubMed: 11112109]
20. Davidow AL, Mangura BT, Wolman MS, Bur S, Reves R, Thompson V, Ford J, Reichler MR. Workplace contact investigation in the United States. *International Journal of Tuberculosis and Lung Disease*. 2003; 7:S446–S452. [PubMed: 14677836]
21. Moran-Mendoza O, Marion SA, Elwood K, Patrick DM, FitzGerald JM. Tuberculin skin test size and risk of tuberculosis development: a large population-based study in contacts. *International Journal of Tuberculosis and Lung Disease*. 2007; 11:1014–1020. [PubMed: 17705981]
22. France AM, Cave MD, Bates JH, Foxman B, Chu T, Yang Z. What's Driving the Decline in Tuberculosis in Arkansas? A Molecular Epidemiologic Analysis of Tuberculosis Trends in a Rural, Low-Incidence Population, 1997–2003. *American Journal of Epidemiology*. 2007; 166:662–671. [PubMed: 17625223]
23. Hethcote HW. The Mathematics of Infectious Diseases. *SIAM Review*. 2000; 42:599–653.
24. Styblo, K. Epidemiology of tuberculosis: selected papers. Royal Netherlands Tuberculosis Association; 1991.
25. Wigginton JE, Kirschner DE. A Model to Predict Cell-Mediated Immune Regulatory Mechanisms During Human Infection with Mycobacterium tuberculosis. *The Journal of Immunology*. 2001; 166:1951–1967. [PubMed: 11160244]
26. Marino S, Pawar S, Fuller CL, Reinhart TA, Flynn JL, Kirschner DE. Dendritic Cell Trafficking and Antigen Presentation in the Human Immune Response to Mycobacterium tuberculosis. *The Journal of Immunology*. 2004; 173:494–506. [PubMed: 15210810]

27. Sud D, Bigbee C, Flynn JAL, Kirschner DE. Contribution of CD8 T Cells to control of *Mycobacterium tuberculosis* infection. *The Journal of Immunology*. 2006; 176:4296–4314. [PubMed: 16547267]
28. Gomez JE, MJD. M. tuberculosis persistence, latency, and drug tolerance. *Tuberculosis*. 2004; 84:29–44. [PubMed: 14670344]
29. Van Rie A, Warren R, Richardson M, Victor TC, Gie RP, Enarson DE, Beyers N, Van Helden P. Exogenous reinfection as a cause of recurrent tuberculosis after curative treatment. *New England Journal of Medicine*. 1999; 341:1174–1179. [PubMed: 10519895]
30. Weis SE, Slocum PC, Blais FX, King B, Nunn M, Matney GB, Gomez E, FBH. The Effect of Directly Observed Therapy on the Rates of Drug Resistance and Relapse in Tuberculosis. *New England Journal of Medicine*. 1994; 330:1179–1184. [PubMed: 8139628]
31. Behr MD, Warren SA, Salamon H, Hopewell PC, Ponce de Leon A, Daley CL, Small PM. Transmission of *Mycobacterium tuberculosis* from patients smear-negative for acid-fast bacilli. *Lancet*. 1999; 353:444–449. [PubMed: 9989714]
32. Tostmann A, Kik SV, Kalisvaart NA, Sebek MM, Verver S, Boeree MJ, van Soolingen D. Tuberculosis transmission by patients with smear-negative pulmonary tuberculosis in a large cohort in the Netherlands. *Clinical Infectious Diseases*. 2008; 47:1135–1142. [PubMed: 18823268]
33. Verver S, Warren RM, Munch Z, Richardson M, van der Spuy GD, Borgdorff MW, Behr MA, Beyers N, van Helden PD. Proportion of tuberculosis transmission that takes place in households in a high-incidence area. *Lancet*. 2004; 363:212–214. [PubMed: 14738796]
34. Ajelli M, Merler S. An individual-based model of hepatitis A transmission. *Journal of Theoretical Biology*. 2009; 259:478–488. [PubMed: 19361529]
35. Ciofi degli Atti M, Merler S, Rizzo C, Ajelli M, Massari M, Manfredi P, Furlanello C, Scalia Tomba G, Iannelli M. Mitigation measures for pandemic influenza in Italy: an individual based model considering different scenarios. *PLoS ONE*. 2008; 3:e1790. [PubMed: 18335060]
36. World Health Organization. Global Tuberculosis Control: Epidemiology, Strategy, Financing. Report 2009, Technical Report. 2009.
37. U.S. Center for Disease Control and Prevention. [accessed November 30, 2010.] Online Tuberculosis Information System. <http://wonder.cdc.gov/tb.html>
38. Marino S, Hogue IB, Ray CJ, Kirschner DE. A Methodology For Performing Global Uncertainty and Sensitivity Analysis in Systems Biology. *Journal of Theoretical Biology*. 2008; 254:178–196. [PubMed: 18572196]
39. Bennett DE, Courval JM, Onorato I, Agerton T, Gibson JD, Lambert L, McQuillan GM, Lewis B, Navin TR, Castro KG. Prevalence of Tuberculosis Infection in the United States Population. *American Journal of Respiratory and Critical Care Medicine*. 2008; 177:348–355. [PubMed: 17989346]
40. Horsburgh CR, O'Donnell M, Chamblee S, Moreland JL, Johnson J, Marsh BJ, Narita M, Scoles Johnson L, Fordham von Reyn C. Revisiting rates of reactivation tuberculosis. *American Journal of Respiratory and Critical Care Medicine*. 2010; 182:425.
41. 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. [PubMed: 9180202]
42. Salpeter EE, Salpeter SR. Mathematical Model for the Epidemiology of Tuberculosis, with Estimates of the Reproductive Number and Infection-Delay Function. *American Journal of Epidemiology*. 1998; 142:398–406. [PubMed: 9508108]
43. Lomax, RG. Statistical Concepts. Lawrence Erlbaum Assoc Inc; 2007.
44. Cohn DL. Treatment of Latent Tuberculosis Infection: Renewed Opportunity for Tuberculosis Control. *Clinical Infectious Diseases*. 2000; 31:120–124.
45. Aparicio JP, Castillo-Chavez C. Preventive treatment of tuberculosis through contact tracing. *Mathematical Biosciences and Engineering*. 2009; 6:209–237. [PubMed: 19364150]
46. Velez DR, Hulme WF, Myers JL, Weinberg JB, Levesque MC, Stryjewski ME, Abbate E, Estevan R, Patillo SG, Gilbert JR, Hamilton W, Scott CD. NOS2A, TLR4, and IFNGR1 interactions influence pulmonary tuberculosis susceptibility in African-Americans. *Human Genetics*. 2009; 126:643–653. [PubMed: 19575238]

47. Murphy BM, Singer BH, Anderson S, Kirschner DE. Comparing epidemic tuberculosis in demographically distinct heterogeneous populations. *Mathematical Biosciences*. 2002; 180:161–185. [PubMed: 12387922]
48. Murphy BM, Singer BH, Kirschner DE. On the treatment of tuberculosis in heterogeneous populations. *Journal of Theoretical Biology*. 2003; 223:391–404. [PubMed: 12875819]

Highlights

- Considering age structure in TB models is key to reproducing realistic proportions of reactivation vs. recently transmitted TB;
- The use of Individual Based Models allows the tracking of transmission in specific settings;
- Modeling socio-demography contributes to a deeper understanding of TB transmission dynamics.

Figure 1. Epidemiological structure of the proposed models

S : Susceptible, E_0 : Exposed for the first time, I_p : Primary TB, L : Latent infection, I_n : Endogenously reactivated TB, E_s : Re-exposed, I_x : Exogenous re-infection resulted in TB. *Exposed* here refers to individuals who have inhaled the pathogen but whose immunological outcome is still undefined. Only individuals in compartments I_p , I_n , I_x can transmit the disease. Gray arrows indicate either TB related mortality for the infectious compartments, or the influx of newborns in the susceptible compartment. All compartments also have an equal (age-specific) rate of natural mortality, which is not shown for simplicity.

Figure 2. Comparison of prevalence, incidence and mortality

Longitudinal prevalence, incidence and mortality per 100,000 individuals in data (2001–2007) and in the three model fits. Vertical lines in panels on the left represent 95% bootstrap confidence intervals around average values for 100 simulations at each time point.

Figure 3. Comparison of age specific incidences

Incidences per 100,000 individuals for five age groups in data (average 2001–2007) and in the two IBMs. Vertical lines represent 95% bootstrap confidence intervals around average values for 100 simulations.

Figure 4. Infection rate by workplace size as predicted by Model C

The red background indicates 95% bootstrap confidence intervals around average values for 100 simulations.

Values adopted for fixed parameters.

Table 1

Par	Model	Unit	Description	Value	Ref
μ	A	yrs^{-1}	natural death rate	$1.25 \cdot 10^{-2}$	*
b	A	yrs^{-1}	number of newborns per year	33,750	**
μ_T	A, B, C	yrs^{-1}	TB related death rate	0.133	[36]
σ_c	B, C	%	protection from re-infection (≤ 10 yrs)	0	[8]
σ_a	B, C	%	protection from re-infection (≥ 20 yrs)	40	[8]
f_c^+	B, C	%	proportion of smear positive cases (≤ 5 yrs)	0	[22]
f_a^+	B, C	%	proportion of smear positive cases (≥ 25 yrs)	29	[22]
A	B, C	-	time-from-exposure-dependent relative risk of primary TB, scaling factor	1.54	[8]
γ	B, C	yrs^{-1}	time-from-exposure-dependent relative risk of primary TB, time constant	0.92	[8]
χ	B, C	%	proportion of cleared infections	68	***

* corresponds to an average lifetime of $\mu^{-1} = 80$ years;
** chosen so to obtain a constant population of $b/\mu = 2,700,000$ individuals, close to Arkansas values in the considered period;
*** fixed from best estimate of model A.

Table 2

Best fitting parameter sets for the three models.

Par	Unit	Description	Model A	Model B	Model C
σ	%	protection from re-infection	28.1	-	-
p	%	proportion of primary TB	15.5	-	-
χ	%	proportion of unsuccessful infections	68	68*	68*
Subset P₁					
k	yr ⁻¹	rate of progression to outcome	2.04	0.947	0.947
d	yr ⁻¹	treatment rate	1.25	1.28	1.28
p_c	%	proportion of primary TB (≤ 10 yrs)	-	9.99	9.99
p_a	%	proportion of primary TB (≥ 20 yrs)	-	16.71	16.71
r	yr ⁻¹	average reactivation rate	1.02 · 10 ⁻³	1.43 · 10 ⁻³	1.43 · 10 ⁻³
Subset P₂					
r_m	yr ⁻²	slope of reactivation rate for ≤ 50 yrs	-	1.34 · 10 ⁻⁵	1.21 · 10 ⁻⁵
β_H	yr ⁻¹	transmissibility (households)	-	-	4.27
β_P	yr ⁻¹	transmissibility (schools/workplaces)	-	-	3.51
β_R	yr ⁻¹	transmissibility (casual contacts)	1.54	17.29	6.26

* the values of χ for Models B and C were not obtained from the fit, but fixed to the best estimate from Model A (See Section 2.4).

Table 3
Comparison of reactivation fractions

Percentage of reactivated cases over total in data (2000–2003) and as predicted by the three models.

	Reactivation fraction	
	Average (%)	95% CI
Data [22]	68.5	67.1–69.8
Model A	76.8	76.5–77.2
Model B	67.4	62.5–71.9
Model C	68.2	61.9–72.5

Table 4
Comparison of infected contacts and secondary active TB cases in households and workplaces

Proportions of infected contacts and secondary active TB cases in households and workplaces as estimated from large scale studies and predicted by Model C.

	reference study			Model C (95% CI)
	[19]	[20]	[21]	
Infection rates (households) (%)	44		36.4	59.7 (59.4–60.0)
Infection rates (workplaces) (%)		22	18.7	9.8 (9.6–9.9)
Secondary TB (households) (%)			2.8	2.87 (2.78–2.97)
Secondary TB (workplaces) (%)			0.63	0.46 (0.44–0.48)