Università di Pisa
Dipartimento di Statistica e Matematica
Applicata all'Economia

Report n. 354

Modelling socio-demography to capture tubercolosis transmission dynamics in a low burder setting.

Giorgio Guzzetta, Marco Ajelli, Zenhua Yang, Stefano Merler, Cesare Furlanello, Denise Kirschner

Pisa, 08 marzo 2012
- Stampato in Proprio –

Via Cosimo Ridolfi, 10 – 56124 PISA – Tel. Segr. Amm. 050 2216231 Segr. Stud. 050 2216317 Fax 050 2216375
Cod. Fisc. 80003670504 - P. IVA 00286820501 - Web http://statmat.ec.unipi.it/

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], Denise Kirschner[d]

[a] *Fondazione Bruno Kessler, Trento, Italy*
[b] *Department of Statistics and Mathematics Applied to Economics, Univ. of Pisa*
[c] *School of Public Health, University of Michigan, USA*
[d] *Department 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 workplaces, 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):

*Corresponding author
Email addresses: guzzetta@fbk.eu (Giorgio Guzzetta), ajelli@fbk.eu (Marco Ajelli), zhenhua@umich.edu (Zhenhua Yang), merler@fbk.eu (Stefano Merler), furlan@fbk.eu (Cesare Furlanello), kirschne@umich.edu (Denise Kirschner)

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.

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 immuno-compromised 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, sus-

ceptible, 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. Unsuccesful 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 $p_i = 1 - e^{-\Delta t \cdot \lambda_i(t)}$, (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

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

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).

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

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 ($\geq$ 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 epi-

demiological 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 (lowcase letters) were combined according to Eq. 2, to produce the score function F_A for Model A.

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

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).

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

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

Table 1: **Values adopted for fixed parameters.**

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_0	B,C	%	protection from re-infection ($\leq$ 10 yrs)	0	[8]
σ_d	B,C	%	protection from re-infection ($\geq$ 20 yrs)	10	[8]
f_c^+	B,C	%	proportion of smear positive cases ($\leq$ 5 yrs)	0	[22]
f_o^+	B,C	%	proportion of smear positive cases ($\geq$ 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]
λ	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.

C) and the shape factor of the reactivation risk by age (r_m for both models). Therefore, the cardinality of P_2 is $\left|P_2^{(B)}\right| = 2$ for model B and $\left|P_2^{(C)}\right| = 4$ 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.

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*	
				Subset P_1	
k	yrs^{-1}	rate of progression to outcome	2.04	0.947	
d	yrs^{-1}	treatment rate	1.25	1.28	
p_c	%	proportion of primary TB ($\leq$ 10 yrs)	-	9.99	
p_a	%	proportion of primary TB ($\geq$ 20 yrs)	-	16.71	
r	yrs^{-1}	average reactivation rate	$1.02 \cdot 10^{-3}$	$1.43 \cdot 10^{-3}$	
				Subset P_2	
r_m	yrs^{-2}	slope of reactivation rate for $\leq$ 50 yrs	-	$1.34 \cdot 10^{-5}$	$1.21 \cdot 10^{-5}$
β_H	yrs^{-1}	transmissibility (households)	-	-	1.27
β_P	yrs^{-1}	transmissibility (schools/workplaces)	-	-	3.51
β_R	yrs^{-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).

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.

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

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 workplaces, 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.

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			
	[19]	[20]	[21]	Model C (95% CI)
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)

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 $\leq$ 10), decreasing steeply to less than half (6%) for large workplaces (size $\geq$ 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 workplaces. 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 autochtonous 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.

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] D. R. Roach, A. G. D. Bean, C. Demangel, M. P. France, H. Briscoe, W. J. Britton, TNF Regulates Chemokine Induction Essential for Cell Recruitment, Granuloma Formation, and Clearance of Mycobacterial Infection., Journal of Immunology 168 (2002) 4620–4627.

[3] K. Bhatt, P. Salgame, Host Innate Immune Response to Mycobacterium tuberculosis, Journal of Clinical Immunology 27 (2007) 347–362.

[4] W. W. Stead, J. P. Lofgren, E. Warren, C. Thomas, Tuberculosis as an endemic and nosocomial infection among the elderly in nursing homes, New England Journal of Medicine 312 (1985) 1487.

[5] S. H. E. Kaufman, Protection against tuberculosis: cytokines, T cells, and macrophages, Annals of the Rheumatic Diseases 61 (2002) ii54–ii58.

[6] R. M. Anderson, M. R. M., Infectious diseases of humans : dynamics and control, Oxford University Press, 1992.

[7] S. M. Blower, A. R. McLean, T. C. Porco, P. M. Small, P. C. Hopewell, M. A. Sanchez, A. R. Moss, The intrinsic transmission dynamics of tuberculosis epidemics, Nature Medicine 1 (1995) 815–821.

[8] E. Vynnycky, P. E. M. Fine, The natural history of tuberculosis: the implications of age-dependent risks of disease and the role of reinfection, Epidemiology and Infection 119 (1997) 183–201.

[9] C. Dye, G. P. Garnett, K. Sleeman, B. G. Williams, Prospects for worldwide tuberculosis control under the WHO DOTS strategy, Lancet 352 (1998) 1886–1891.

[10] C. Dye, B. G. Williams, Eliminating human tuberculosis in the twenty-first century, Journal of the Royal Society Interface 5 (2008) 233–243.

[11] J. A. Salomon, J. O. Lloyd-Smith, W. M. Getz, S. Resch, M. Sanchez, T. Porco, M. W. Borgdorff, Prospects for advancing tuberculosis control efforts through novel therapies, PLoS Medicine 3 (2006) e273.

[12] L. J. Abu-Raddad, L. Sabatelli, J. Achterberg, J. Sugimoto, I. Longini, C. Dye, M. Halloran, Epidemiological benefits of more-effective tuberculosis vaccines, drugs, and diagnostics, PNAS 106 (2009) 13980–13985.

[13] I. M. Longini, M. E. Halloran, A. Nizam, Y. Yang, Containing Pandemic Influenza with Antiviral Agents, American Journal of Epidemiology 159 (2004) 623–633.

[14] N. M. Ferguson, D. A. T. Cummings, S. Cauchemez, C. Fraser, S. Riley, A. Meeyai, S. Iamsirithaworn, D. S. Burke, Strategies for containing an emerging influenza pandemic in Southeast Asia, Nature 437 (2005) 209–214.

[15] S. Merler, M. Ajelli, The role of population heterogeneity and human mobility in the spread of pandemic influenza, Proceedings of the Royal Society B 277 (2010) 557–565.

[16] M. Murray, Determinants of cluster distribution in the molecular epidemiology of tuberculosis, PNAS 99 (2002) 1538–1543.

[17] T. Cohen, C. Colijn, B. Finklea, M. Murray, Exogenous re-infection and the dynamics of tuberculosis epidemics: local effects in a network

model of transmission, Journal of the Royal Society Interface 4 (2007) 523–531.

[18] A. M. France, Integrating molecular typing into routine tuberculosis surveillance: an assessment of the strengths and limitations of current approaches, Ph.D. thesis, University of Michigan, 2008.

[19] S. M. Marks, Z. Taylor, N. L. Qualls, R. J. Shrestha-Kuwahara, M. A. Wilce, C. H. Nguyen, Outcomes of Contact Investigation of Infectious Tuberculosis Patients, American Journal of Respiratory and Critical Care Medicine 162 (2000) 2033–2038.

[20] A. L. Davidow, B. T. Mangura, M. S. Wolman, S. Bur, R. Reves, V. Thompson, J. Ford, M. R. Reichler, Workplace contact investigation in the United States, International Journal of Tuberculosis and Lung Disease 7 (2003) S446–S452.

[21] O. Moran-Mendoza, S. A. Marion, K. Elwood, D. M. Patrick, J. M. FitzGerald, Tuberculin skin test size and risk of tuberculosis development: a large population-based study in contacts, International Journal of Tuberculosis and Lung Disease 11 (2007) 1014–1020.

[22] A. M. France, M. D. Cave, J. H. Bates, B. Foxman, T. Chu, Z. Yang, 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 166 (2007) 662–671.

[23] H. W. Hethcote, The Mathematics of Infectious Diseases, SIAM Review 42 (2000) 599–653.

[24] K. Styblo, Epidemiology of tuberculosis: selected papers, Royal Netherlands Tuberculosis Association, 1991.

[25] J. E. Wigginton, D. E. Kirschner, A Model to Predict Cell-Mediated Immune Regulatory Mechanisms During Human Infection with Mycobacterium tuberculosis, The Journal of Immunology 166 (2001) 1951–1967.

[26] S. Marino, S. Pawar, C. L. Fuller, T. A. Reinhart, J. L. Flynn, D. E. Kirschner, Dendritic Cell Trafficking and Antigen Presentation in the Human Immune Response to Mycobacterium tuberculosis, The Journal of Immunology 173 (2004) 494–506.

[27] D. Sud, C. Bigbee, J. A. L. Flynn, D. E. Kirschner, Contribution of CD8 T Cells to control of Mycobacterium tuberculosis infection, The Journal of Immunology 176 (2006) 4296–4314.

[28] J. E. Gomez, M. J. D., M. tuberculosis persistence, latency, and drug tolerance, Tuberculosis 84 (2004) 29–44.

[29] A. Van Rie, R. Warren, M. Richardson, T. C. Victor, R. P. Gie, D. E. Enarson, N. Beyers, P. Van Helden, Exogenous reinfection as a cause of recurrent tuberculosis after curative treatment, New England Journal of Medicine 341 (1999) 1174–1179.

[30] S. E. Weis, P. C. Slocum, F. X. Blais, B. King, M. Nunn, G. B. Matney, E. Gomez, F. B. H., The Effect of Directly Observed Therapy on the

Rates of Drug Resistance and Relapse in Tuberculosis, New England Journal of Medicine 330 (1994) 1179–1184.

[31] M. D. Behr, S. A. Warren, H. Salamon, P. C. Hopewell, A. Ponce de Leon, C. L. Daley, P. M. Small, Transmission of Mycobacterium tuberculosis from patients smear-negative for acid-fast bacilli, Lancet 353 (1999) 444–449.

[32] A. Tostmann, S. V. Kik, N. A. Kalisvaart, M. M. Sebek, S. Verver, M. J. Boeree, D. van Soolingen, Tuberculosis transmission by patients with smear-negative pulmonary tuberculosis in a large cohort in the Netherlands., Clinical Infectious Diseases 47 (2008) 1135–1142.

[33] S. Verver, R. M. Warren, Z. Munch, M. Richardson, G. D. van der Spuy, M. W. Borgdorff, M. A. Behr, N. Beyers, P. D. van Helden, Proportion of tuberculosis transmission that takes place in households in a high-incidence area, Lancet 363 (2004) 212–214.

[34] M. Ajelli, S. Merler, An individual-based model of hepatitis A transmission, Journal of Theoretical Biology 259 (2009) 478–488.

[35] M. Ciofi degli Atti, S. Merler, C. Rizzo, M. Ajelli, M. Massari, P. Manfredi, C. Furlanello, G. Scalia Tomba, M. Iannelli, Mitigation measures for pandemic influenza in Italy: an individual based model considering different scenarios, PLoS ONE 3 (2008) e1790.

[36] World Health Organization, Global Tuberculosis Control: Epidemiology, Strategy, Financing. Report 2009, Technical Report, 2009.

[37] U.S. Center for Disease Control and Prevention, Online Tuberculosis Information System, http://wonder.cdc.gov/tb.html, accessed November 30, 2010.

[38] S. Marino, I. B. Hogue, C. J. Ray, D. E. Kirschner, A Methodology For Performing Global Uncertainty and Sensitivity Analysis in Systems Biology, Journal of Theoretical Biology 254 (2008) 178–196.

[39] D. E. Bennett, J. M. Courval, I. Onorato, T. Agerton, J. D. Gibson, L. Lambert, G. M. McQuillan, B. Lewis, T. R. Navin, K. G. Castro, Prevalence of Tuberculosis Infection in the United States Population, American Journal of Respiratory and Critical Care Medicine 177 (2008) 348–355.

[40] C. R. Horsburgh, M. O'Donnell, S. Chamblee, J. L. Moreland, J. Johnson, B. J. Marsh, M. Narita, L. Scoles Johnson, C. Fordham von Reyn, Revisiting rates of reactivation tuberculosis, American Journal of Respiratory and Critical Care Medicine 182 (2010) 425.

[41] T. M. Daniel, S. M. Debanne, Estimation of the Annual Risk of Tuberculous Infection for White Men in the United States, Journal of Infectious Diseases 175 (1997) 1535–1537.

[42] E. E. Salpeter, S. R. Salpeter, Mathematical Model for the Epidemiology of Tuberculosis, with Estimates of the Reproductive Number and Infection-Delay Function, American Journal of Epidemiology 142 (1998) 398–406.

[43] R. G. Lomax, Statistical Concepts, Lawrence Erlbaum Assoc Inc., 2007.

[44] D. L. Cohn, Treatment of Latent Tuberculosis Infection: Renewed Opportunity for Tuberculosis Control, Clincal Infectious Diseases 31 (2000) 120–124.

[45] J. P. Aparicio, C. Castillo-Chavez, Preventive treatment of tuberculosis through contact tracing, Mathematical Biosciences and Engineering 6 (2009) 209–237.

[46] D. R. Velez, W. F. Hulme, J. L. Myers, J. B. Weinberg, M. C. Levesque, M. E. Stryjewski, E. Abbate, R. Estevan, S. G. Patillo, J. R. Gilbert, W. Hamilton, C. D. Scott, NOS2A, TLR4, and IFNGR1 interactions influence pulmonary tuberculosis susceptibility in African-Americans, Human Genetics 126 (2009) 643–653.

[47] B. M. Murphy, B. H. Singer, S. Anderson, D. E. Kirschner, Comparing epidemic tuberculosis in demographically distinct heterogeneous populations, Mathematical Biosciences 180 (2002) 161–185.

[48] B. M. Murphy, B. H. Singer, D. E. Kirschner, On the treatment of tuberculosis in heterogeneous populations, Journal of Theoretical Biology 223 (2003) 391–404.

Figure Captions

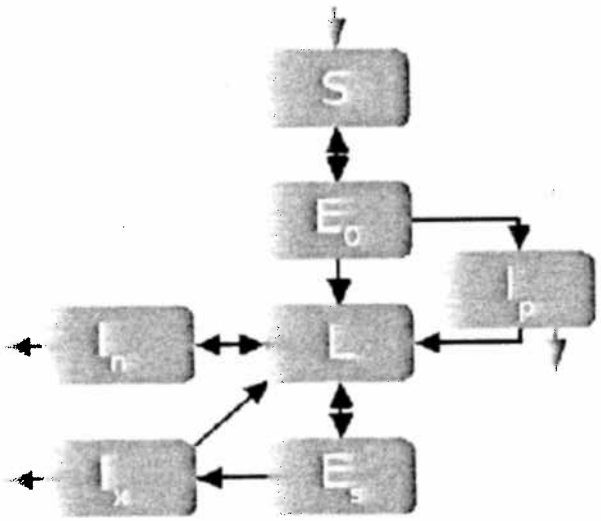

Figure 1: **Epidemiological structure of the proposed models.** S: Susceptible, E_0: Exposed for the first time, I_p: Primary TB, L: Latent infection, I_n: Endogenously reactivated TB, E_s: Re-exposed, I_x: Exogenous re-infection resulted in TB. *Exposed* here refers to individuals who have inhaled the 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.

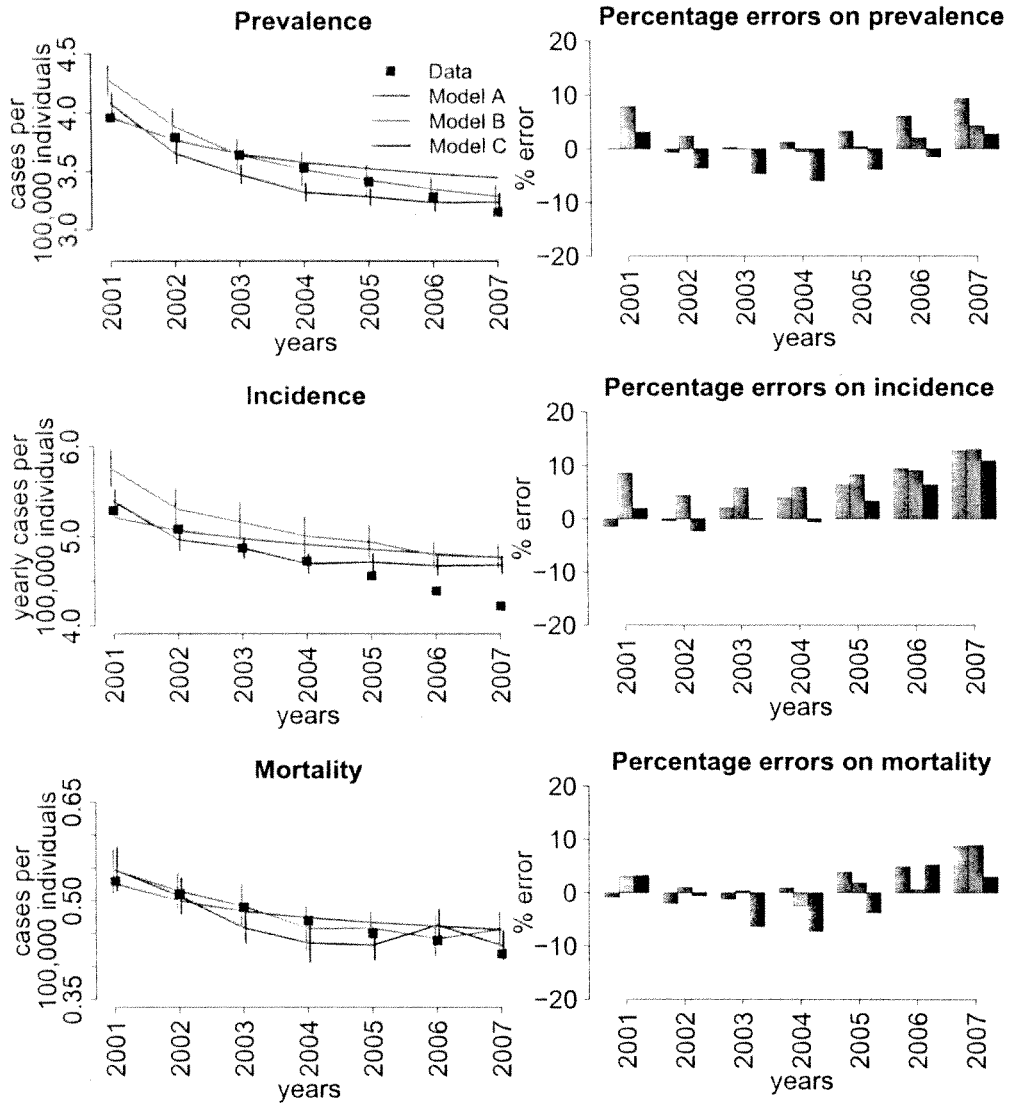

Figure 2: **Comparison of prevalence, incidence and mortality.** Longitudinal prevalence, incidence and mortality per 100,000 individuals in data (2001-2007) and in the three model fits. Vertical lines in panels on the left represent 95% bootstrap confidence intervals around average values for 100 simulations at each time point.

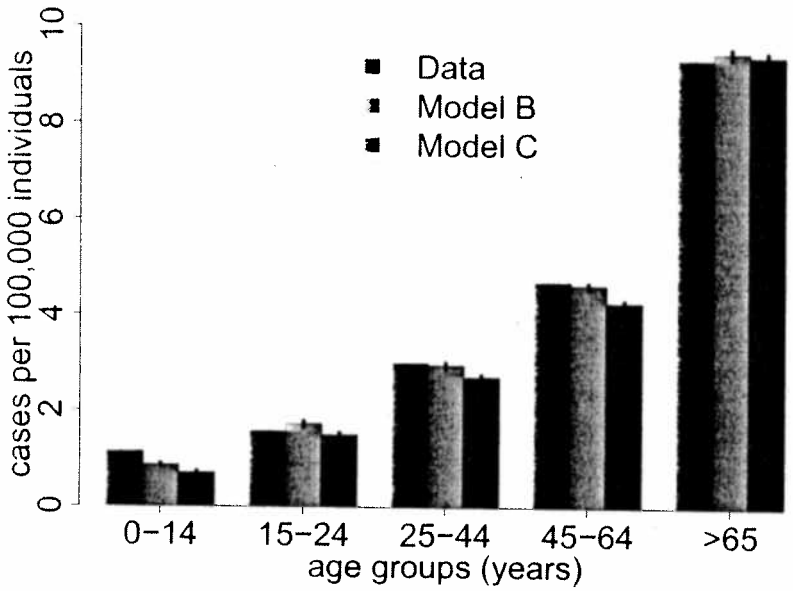

Figure 3: **Comparison of age specific incidences.** Incidences per 100,000 individuals for five age groups in data (average 2001-2007) and in the two IBMs. Vertical lines represent 95% bootstrap confidence intervals around average values for 100 simulations.

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.

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

Giorgio Guzzetta[1,2], Marco Ajelli[1], Zhenhua Yang[3], Stefano Merler[1], Cesare Furlanello[1] and Denise Kirschner[4].

[1] Fondazione Bruno Kessler, Trento, Italy
[2] Department of Mathematics and Statistics Applied to Economics, University of Pisa
[3] School of Public Health, University of Michigan
[4] Department of Microbiology and Immunology, University of Michigan Medical School

Supplementary material

Contents

1 Model implementations

1.1 Model A

The Ordinary Differential Equations for Model A, reported in this section, describe the epidemiological structure schematized in Figure 1.

The following notation is adopted:

- S: susceptible;
- E_0: exposed for the first time;
- I_p: primary TB;
- L: latently infected;
- I_n: endogenously reactivated TB;
- E_s: re-exposed;
- I_x: exogenously re-infected TB.

Both *exposed* compartments E_0 and E_s represent hosts who have recently inhaled the pathogen and still have an undecided outcome of disease (see Section Materials and Methods in the main text for further details). E_0 are individuals exposed for the first time while E_s have been previously exposed.

$$
\begin{aligned}
\frac{dS}{dt} &= b + \mu_T(I_p + I_n + I_x) + k\chi(1-p)E_0 - \beta_R \frac{I_p + I_n + I_x}{N} S - \mu S \\
\frac{dE_0}{dt} &= \beta_R \frac{I_p + I_n + I_x}{N} S - kE_0 - \mu E_0 \\
\frac{dI_p}{dt} &= pkE_0 - dI_p - (\mu + \mu_T) I_p \\
\frac{dL}{dt} &= k(1-p)(1-\chi)E_0 + k(1-(1-\sigma)p)E_s + d(I_p + I_n + I_x) \\
&\quad -\beta_R \frac{I_p + I_n + I_x}{N} L - rL - \mu L \\
\frac{dI_n}{dt} &= rL - dI_n - (\mu + \mu_T) I_n \\
\frac{dE_s}{dt} &= \beta_R \frac{I_p + I_n + I_x}{N} L - kE_s - \mu E_s \\
\frac{dI_x}{dt} &= (1-\sigma)pkE_s - dI_x - (\mu + \mu_T) I_x
\end{aligned}
$$

A description of the terms contributing to each equation is reported hereafter (all terms are intended per unit time):

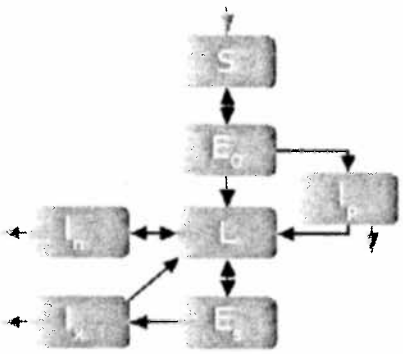

Figure 1: **Epidemiological structure of the proposed models.** S: Susceptible, E_0: Exposed for the first time, I_p: Primary TB, L: Latent infection, I_n: Endogenously reactivated TB, E_s: Re-exposed, I_x: Exogenous re-infection resulted in TB. *Exposed* here refers to individuals who have inhaled the 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.

- μX (with X any model compartment): natural death of individuals;
- $\mu_T I_z$ (with $z \in \{p, n, x\}$): TB-induced mortality;
- dI_z (with $z \in \{p, n, x\}$): individuals diagnosed and effectively treated;
- $b + \mu_T(I_p + I_n + I_x)$: births; this term replaces people who have died of natural death or because of TB disease, in order to keep the population size constant;
- $\beta_R \frac{I_p+I_n+I_x}{N} S$: first episode of exposure of susceptible individuals; homogeneous mixing is assumed;
- kE_0: individuals exposed for the first time who have resolved their outcome of infection, and move to either the susceptible, latent or primary TB compartments;
- pkE_0: individuals exposed for the first time who develop primary active TB;
- $k\chi(1-p)E_0$: individuals exposed for the first time who have cleared the pathogen;
- $k(1-p)(1-\chi)E_0$: individuals exposed for the first time who enter the latent state;
- $\beta_R \frac{I_p+I_n+I_x}{N} L$: re-exposure of latently infected individuals;
- kE_s: re-exposed individuals who have resolved their outcome of infection, and move to either the exogenously re-infected or to the latent compartment;
- $(1-\sigma)pkE_s$: re-exposed individuals developing exogenous re-infection;

Table 1: **Parameters description for Model A.**

Parameter	Description	Unit
b	newborns	yrs^{-1}
μ	natural mortality	yrs^{-1}
μ_T	TB-related mortality	yrs^{-1}
β_R	transmission rate	yrs^{-1}
k	rate of flow out from exposure	yrs^{-1}
p	proportion of active disease upon infection	%
γ	clearance of first infections	%
σ	protection from exogenous reinfection	%
r	rate of endogenous reactivation	yrs^{-1}
d	successful treatment rate	yrs^{-1}

- $k(1-(1-\sigma)p)E_s$: re-exposed individuals who do not develop disease and revert to the latent state;
- rL: latently infected individuals who reactivate;
- $d(I_p+I_n+I_x)$: individuals who have successfully completed treatment.

An interpretation of parameters involved in each term, together with their unit of measure is given in Table 1. See Section 2 for choices of parameter values and their ranges for model calibration, and Table 2 in the main text for their best-fitting parameter values after calibration.

1.2 Model B

An age-structured Individual Based Model with homogeneous mixing and closed population, i.e. without immigration and emigration, is described here. Each individual in the model population is initially assigned an age, according to the Arkansas population age structure in 2000 [1].

The vital dynamics of the population has to be considered for modeling endemic diseases. The modeled population is therefore updated once a year (i.e., each January 1 of the simulation time). At each population update, individuals may die on the basis of the mortality rates by age. Newborns are generated according to observed fertility rates [2], so that, unlike Model A, the population is no longer constant. The maximum age is 100 years: individuals of this age that are still alive will be removed from the population at the next population update.

The epidemiological rules, common to Model C, are described in Section 1.5.

1.3 Model C

An age-structured Individual Based Model with preferential mixing through selected socio-demographic routes and closed population is discussed in this section.

1.3.1 Population density

The Gridded Population of the World version 3 (GPW v3) [3] produced by the Center for International Earth Science Information Network (CIESIN) of the Earth Institute at Columbia University was used as the source of population density for the study area. The study area covers the state of Arkansas, for a total population of about 2,673,400 individuals in year 2000 [4]. A total surface of 137,000 square km is covered, divided in 219 cells of approximately 625 square kilometers. The average population in the cells is 12,207, ranging from 1,629 to 140,674. The grid allows spatial location of households, schools and workplaces, used for modeling commuting patterns (Section 1.3.3) and distance-dependent casual contacts (Section 1.5).

1.3.2 Households

Considering realistic household groups in the transmission of TB is crucial in the perspective of analyzing the effect of specific control strategies, such as treatment of latent TB infection (LTBI) to contacts of index cases. LTBI treatment is considered one of the key measures for reducing TB incidence in low-burden settings [5].

We used an heuristic model based on previously published models (e.g. [6, 7, 8]) that matches marginal distributions of household size and population age structure, and maintains realistic generational age gaps within households (by avoiding randomly assigned ages to the households members), respecting as best as possible the actual mix of age groups. A sketch of the heuristic model employed is shown below:

1. determine the household type by choosing among six household types from the proper frequency distribution (see Table 2);
2. determine the household size by sampling from the proper frequency distribution (see Table 3);
3. assign an age to the household head, a_h, by sampling from the distribution of the age classes (see Table 4) and according to type-specific constraints, detailed below;
4. assign an age to the additional members by sampling from the population age structure and taking into account general constraints, detailed hereafter, as well as type specific ones.

General constraints These constraints are applied throughout the initialization of households and used for the demographic evolution as well (Section 1.4):

- All families are composed by a household head, at most one spouse, at most one aggregated elderly person (assumed to be a parent of the household head or of the spouse), and at most six children;
- The age of the spouse, a_s satisfies $a_h - 12 \leq a_s \leq a_h$ (a_h is the age of the household head) and $a_s \geq 15$; thus, by spouse we indicate the younger member of the couple;

Table 2: **Distribution of household types in Arkansas [1].** The asterisk (*) in types 2-5 indicates the possible presence of an aggregated elderly member.

	Household type	Proportion
1	Single person households	25.6%
2	Married couples without children under 18 years old (*)	31.6%
3	Married couples with children under 18 years old (*)	22.7%
4	Singles without children under 18 years old (*)	6.6%
5	Singles with children under 18 years old (*)	9.3%
6	Other households	4.2%

Table 3: **Distribution of household sizes in Arkansas** [1].

Household size	1	2	3	4	5	6	≥ 7
Proportion	25.6%	35.1%	17.3%	13.5%	5.6%	1.9%	1.0%

- The age of children, a_c, satisfies $a_h - 35 \leq a_c \leq a_s - 15$ if there is a couple in the house, and $a_h - 35 \leq a_c \leq a_h - 15$ otherwise;
- the age of the aggregated elderly, a_e satisfies $a_s + 15 \leq a_e \leq a_h + 35$ if there is a couple in the house, and $a_h + 15 \leq a_e \leq a_h + 35$ otherwise.

Spouses are present only in households of type 2 and 3 (see Table 2). Aggregated elderly are present only in households of type 2,3,4,5. The probability of having an aggregated elderly was estimated from census data [1] as about 1% for types 2 and 3, and 1.8% for types 4 and 5.

Type-specific constraints

- Households of type 2 represent one of the following three subcases:
  a. married couples without children (size 2);
  b. married couples without children and an aggregated elderly person (size 3);
  c. married couples with one or more children at least 18 years old (size ≥ 2) and possibly an aggregated elderly.

  Therefore:
  - if size is > 3 (a subset of case (c)), at least a child is present; therefore, the age of the household head and spouse are forced to be ≥ 33 since otherwise they would be too young to have a child of age ≥ 18;

Table 4: **Distribution of ages of household heads in Arkansas** [1].

Age of household head	15-25	25-34	35-44	45-54	55-64	65-74	75-84	85-100
Proportion	6.3%	16.4%	20.5%	18.9%	14.5%	12.3%	8.5%	2.6%

- if the size is 3 and the spouse is < 33 years old, the couple is too young to have children of age ≥ 18; therefore, the presence of an elderly aggregate is forced;
- the remaining individuals are forced to be children, with $18 \leq a_c \leq a_s - 15$.

- Households of type 3:
  - the age of the household head is limited to be < 54 since otherwise the couple would be be too old to have a child of age < 18;
  - the first other individual besides the couple is a child of age < 18;
  - other children are assigned with age ranges compatible with parents' ages according to the general constraints.

- Households of type 4 represent one of the following two subtypes:

  a. singles without children and with an aggregated elderly (size 2);

  b. singles with one or more children at least 18 years old (size ≥ 2) and possibly an aggregated elderly.

  Therefore:
  - if size is > 2 (a subset of case (b)), at least a child is present; therefore, the age of the household head is forced to be ≥ 33 since otherwise he/she would be too young to have a child of age ≥ 18;
  - if size is 2 and the household head is < 33 years old, the household head is too young to have children of age ≥ 18; therefore, the presence of an elderly aggregate is forced;
  - the remaining individuals are forced to be children, with $18 \leq a_c \leq a_s - 15$.

- Households of type 5:
  - the age of the household head is limited to be < 54 since otherwise he/she would be too old to have a child of age < 18;
  - the first other individual besides the household head is a child of age < 18;
  - other children are assigned with age ranges compatible with parents' ages according to the general constraints.

- Households of type 6: all members of households of type 6 other than the household head are not labeled as spouses, aggregated elderly, or children; thus, they are not subject to age constraints. Members of these households can form a new household according to rules described in Section 1.4.

1.3.3 Schools and workplaces

Data for the allocation of individuals in schools and workplaces were drawn from routine public statistics. Schools and workplaces were spatially-distributed proportionally to the population. The distribution of school sizes by type for Arkansas were available from [9]; school attendance rates by school type were provided by [1]. The proportion of workplace establishments by 6 sizes was obtained from [10], and the probabilities of employment by age group were available from [11].

Students and workers were randomly assigned to a school or workplace, in such a way that the probability density function of travel distances complies with a gravity law, as proposed in [12]. The probability c_{ij} of commuting from the place of residence, situated in county i (donor), to the school or workplace, located in county j (receiver), is described by the formula:

$$c_{ij} = \theta \frac{p_i^{\tau_d} p_j^{\tau_r}}{d_{ij}^{\rho}} \tag{1}$$

where:

- p_i is the population of the donor county;
- p_j is the population of the receiver county;
- d_{ij} is the distance between the centroids of counties i and j;
- τ_d, τ_r and ρ are the model's parameter; τ_d and τ_r tune the dependence of dispersal on donor and recipient sizes and ρ tunes the dependence on the distance.
- θ is a scaling factor so that c_{ij} is adimensional and has values of a probability density function (i.e., $\sum_{i,j} c_{ij} = 1$) [6].

County to county worker flow data for all residents of all 75 Arkansas counties are publicly available from [13]. According to these data, only 3% of the people worked outside Arkansas, and only 0.03% outside the USA. Therefore, limiting our commuting model to just within-state movements represents a good approximation of the real figure.

For each couple donor-receiver, the fraction C_{ij} of corresponding commuters over the total is calculated from data. Optimal parameters of model 1 are obtained by minimizing the square error between C_{ij} and c_{ij}, averaged over all i and j. We used a Nelder-Mead gradient descent algorithm [14] for finding the best parameter sets, obtaining $\tau_d = 0.21$, $\tau_r = 1.00$ and $\rho = 1.21$.

Outgoing commuters from a given county i for the model and observed data are defined respectively as q_O and Q_O:

$$q_O(i) = \sum_j c_{ij}$$

$$Q_O(i) = \sum_j C_{ij}$$

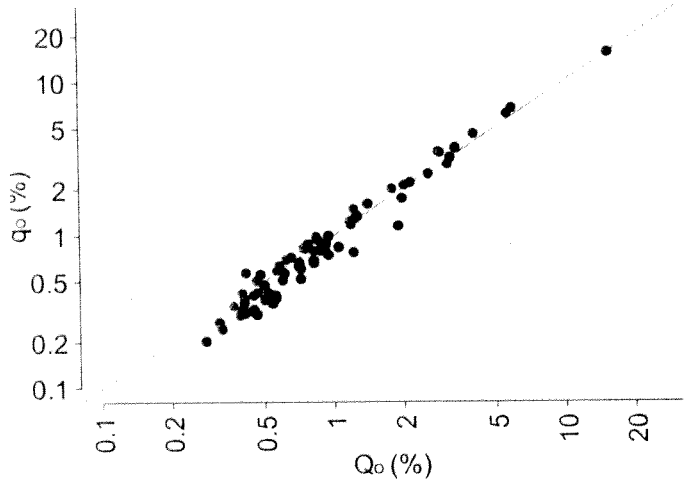

Figure 2: **Fraction of outgoing commuters over the total by county.** The scatter plot represents in a log-log scale the proportion of outgoing commuters for each of the 75 Arkansas counties. Observed values [13] are on the x axis, while the y axis represents values predicted by the gravity model described in Eq. 1 with $\tau_d = 0.21$, $\tau_r = 1.00$ and $\rho = 1.21$.

Similarly, incoming commuters to a given county j are defined as q_I and Q_I:

$$q_I(j) \sum_i c_{ij}$$

$$Q_I(j) \sum_i C_{ij}$$

The scatter plot of predicted (q_O) versus observed (Q_O) outgoing commuters is displayed in Figure 2 for all 75 counties. Figure 3 shows the same scatter plot for the incoming commuters, Q_I. In both cases, a very good agreement is found, with an R^2 between observation and predictions of respectively 0.989 and 0.987.

Finally, Figure 4 describes the probability of commuting across a given distance d_{ij}, as suggested from data analysis and by the model; distances are grouped into 20 logarithmically scaled classes for clarity. This figure shows that the proposed gravity model describes well the average distance covered by commuters in Arkansas.

1.4 Evolution of socio-demography

Since we are modeling an endemic disease, the demography and the network of contacts among individuals have to be kept updated, as in [7]. In model C individuals come to life, grow, can generate new households, can procreate and die; moreover, they can attend school (following the educational career), can be employed and retire.

At each population update, individuals may die on the basis of the mortality rates by age. Individuals of age > 100 are also removed from the population. Newborns are located according to observed fertility rates [2] in an existing household, chosen among those of

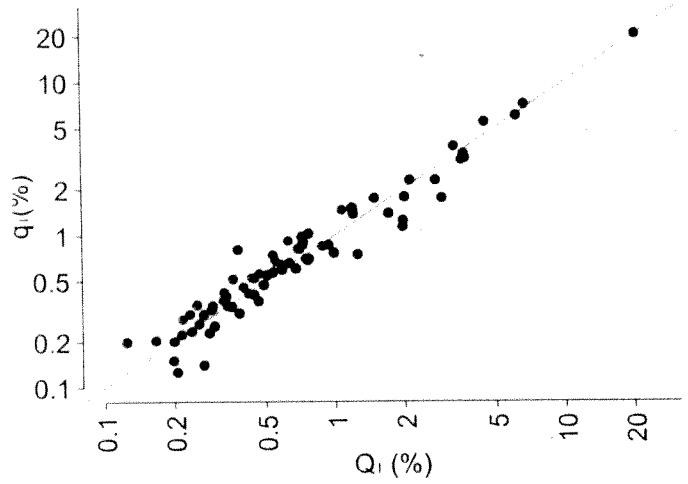

Figure 3: **Fraction of incoming commuters over the total by county.** The scatter plot represents in a log-log scale the proportion of incoming commuters for each of the 75 Arkansas counties. Observed values [13] are on the x axis, while the y axis represents values predicted by the gravity model described in Eq. 1 with $\tau_d = 0.21$, $\tau_r = 1.00$ and $\rho = 1.21$.

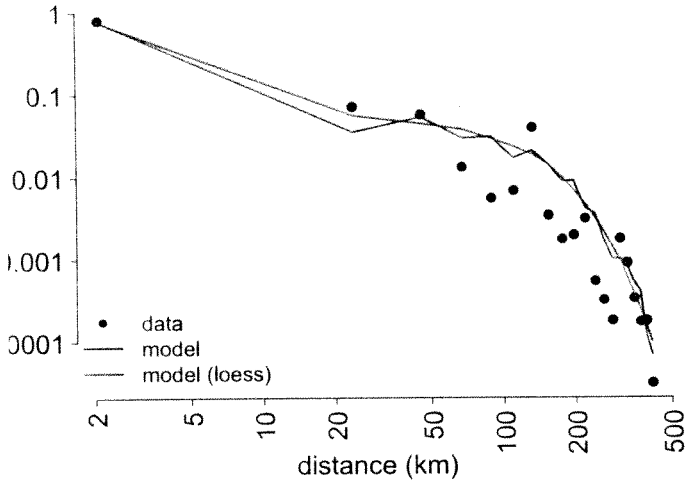

Figure 4: **Commuting probability over distance.** The graph represents the probability of commuting over a given distance in Arkansas. Dots represent observed data for 20 distance classes (logarithmically spaced) from all possible donor-receiver combination, and are placed at the geometric mean distance between the class lower and upper bounds. The blue line is the probability value predicted by the gravity model [12] for similarly grouped distance classes, while the yellow line is a LOESS smoothing of the model predictions. The first point represents commuting within counties; in this plot, an arbitrary distance of 2 km from house to school/work has been assigned for this point.

suitable members age. Finally, it is determined whether newborns are enrolled in a day care center or not, according to specific data [15].

Individuals who have not already generated their own family and are at least 15 years old can generate a new household, possibly a single member household, on the basis of the probability of getting married by age. The new household forms in the same cell of origin of the older member of the couple.

Divorce is also modeled. In this case, the older member of the couple forms a new single-person household in the same cell of the original household. Children and the aggregated elderly (if any) stay in the original family.

Every time the population is updated the age of each individual is increased and an employment category (possibly "unemployed") is associated to each individual on the basis of the employment rates by age. If this category is the same that she/he had before the update she/he keeps her/his place, otherwise a new school/workplace is randomly chosen among the existing ones. In the latter case, the commuting destination is determined by employing the same gravity model employed for the initialization of the network of contacts. Regarding students, the school type (day care center, kindergarten, primary, middle, high school, university) is deterministically assigned on the basis of the age of the individual. Since the age of the individuals is kept updated, the previously described procedure allows students to follow an educational career.

1.4.1 Validation of model evolution

Both model B and C were initialized with demographic data from Arkansas in 2000 on age-structure, age-specific mortality and women's fertility, publicly available from the U.S. Bureau of Census [1] and the CDC National Vital Statistics [16]. Model C also used data for the population density on the Arkansas territory in 2000 [17], the age-specific probabilities of marriage, divorce and employment, and the distributions of types and sizes for households schools and workplaces [1, 11, 18] in the same year. The evolution of the socio-demographic structure was validated against the Arkansas age structure as estimated in 2005 [2] and as projected for 2030 [19]. Model predictions for the total population and the total number of births, deaths, marriages and divorces was compared with vital statistics from 2001-2005, as made publicly available by the Arkansas Department of Health [20].

The upper panel of Figure 5 compares the initialization of model age structure with Census data from 2000 [1]. Almost perfect agreement is found, except for a small disalignment at younger ages (5-15 years). This is due to the classification of households by the U.S. Bureau of Census, which does not explicitly state the number of children under age for the corresponding household type: the model may underestimate such numbers, possibly replacing missing children with an excessive number of aggregated elderly person in such households (as suggested by the slight overestimation of elderly people in the rightmost part of the graph). The lower panel of Figure 5 compares the distribution of household sizes for year 2000 in the model and in the data. The slight overestimation of larger households (of sizes 6 and 7) can be partly explained by noting that 'size 7' in the data actually indicates households with at least 7 individuals; thus model simulations include the excess individuals in overpopulated households. In Figure 6 (top-left panel) we show a similar linear trend of growth for the model population as for that estimated and projected for Arkansas by the U.S. Bureau of Census. The slight underestimation of population growth is due to immigration: in fact, if we compare

the demographic balance without considering immigration (i.e., the total number of births minus the total number of deaths) in the first five years of simulation, as in the top-right panel of Figure 6, the model slightly overestimates the data. The breakdown for deaths and births is given in the two lower panels of Figure 6. Overall, the effect of immigration is very low in the demographic dynamics and can be neglected. Figure 7 compares the age structures as predicted by the model and that estimated in 2005 on a sample population and projected by the U.S. Bureau of Census in 2030. There is a general good agreement, except for an underestimation by the model for ages 50-70 in 2030. We hypothesize that this discrepancy derives from not accounting for immigration, since the age groups involved correspond to working ages in the period 2000-2030.

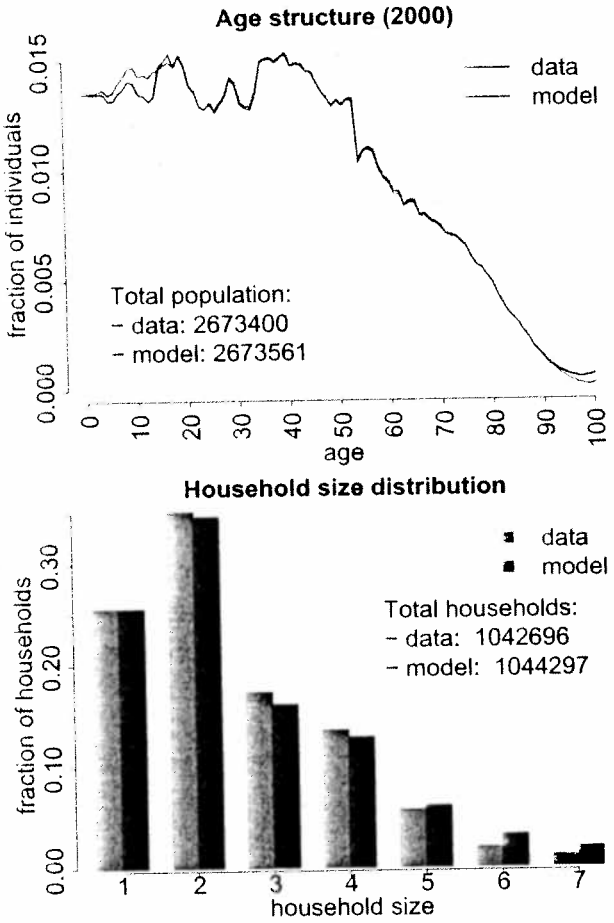

Figure 5: **Evaluation of the model's initialization.** Upper panel: age structure. Lower panel: household size distribution.

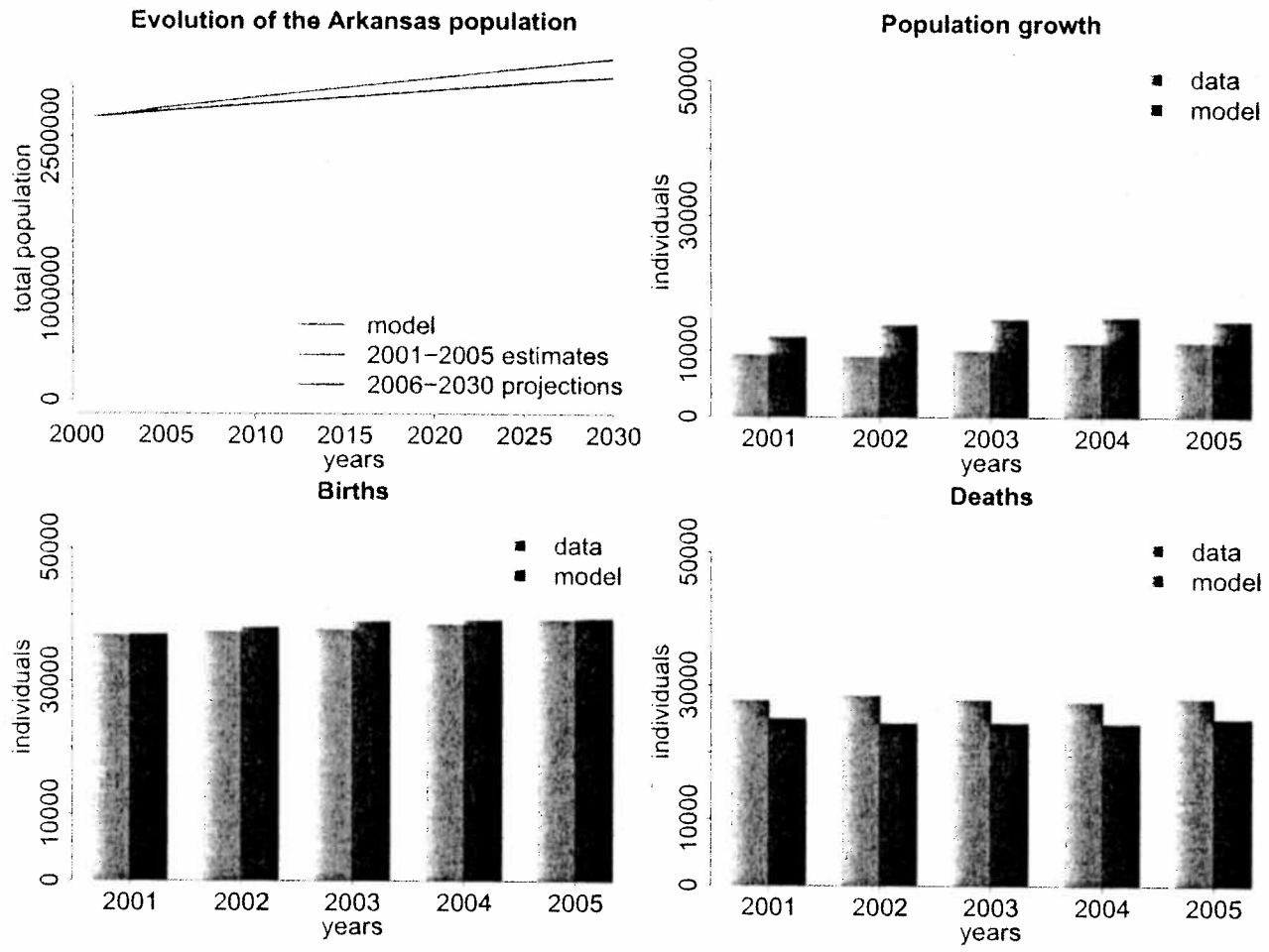

Figure 6: Comparison of total population and demographic balance.

1.5 Epidemiological rules for models B and C

The epidemiological rules are the same for models B and C, with differences only in the the setting and distance-dependent transmission in model C. The epidemiological flow is schematized in Figure 8. In this section we describe such rules, highlighting specificities of models B and C as appropriate. Once the population is initialized, at any time t of the simulation (time step $\Delta t = 1$ week) individuals from each compartment can change their epidemiological status.

1. Individuals exposed (i.e. who had adequate contact) for the first time (E_0) progress to one of the following possible outcomes of TB infection at a rate $k\Delta t$:

   - development of primary TB (I_p) with probability $p(a)g(t-\tau)$; this probability is composed of an age-dependent risk, $p(a)$ (a representing age), and a function of the time from exposure, $g(t-\tau)$, where τ is the time at which the adequate contact

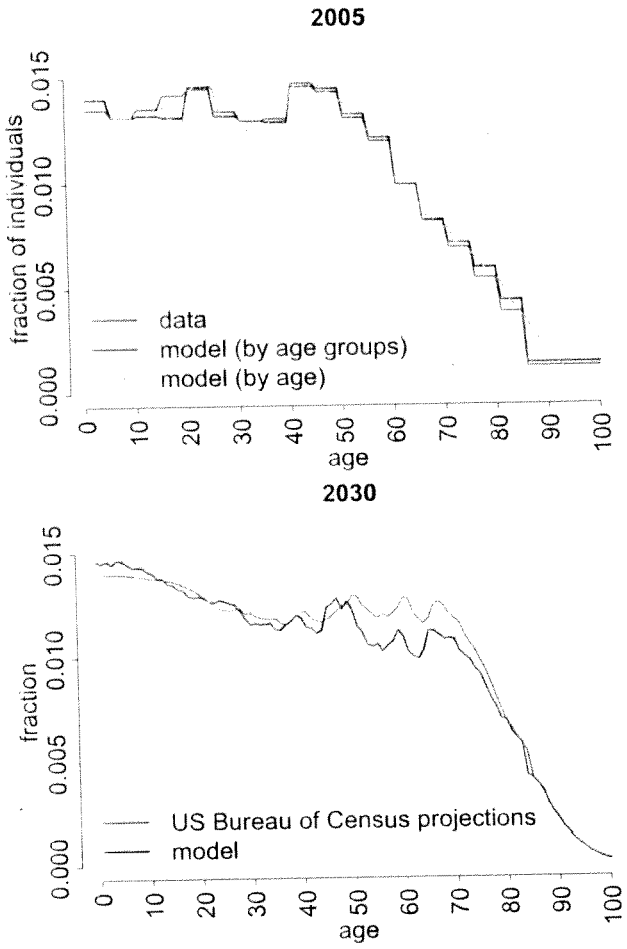

Figure 7: **Comparison of age structures.** Age structures predicted by the IBM against estimates (2005) and projections (2030) by the U.S. Bureau of Census.

with an infectious case has occurred. Function $g(t-\tau)$ accounts for the decreasing, non-zero risk of TB up to 5 years from the episode of exposure [21].

- clearance of the infection with probability χ; these individuals revert to susceptibility (S);
- containment of infection and progress to the latent compartment (L).

2. Re-exposed individuals (E_s) also exit from their state at a rate $k\Delta t$. They can develop active disease (exogenous reinfection, I_x) with a reduced probability as compared to first-time exposed [21]. We indicate the age-dependent protection factor with $\sigma(a)$, so that the risk of exogenous re-infection is $p(a)g(t)(1-\sigma(a))$. Those who do not progress to the I_x compartment revert to the latent state.

3. Latently infected individuals can progress to endogenous reactivation with an age-

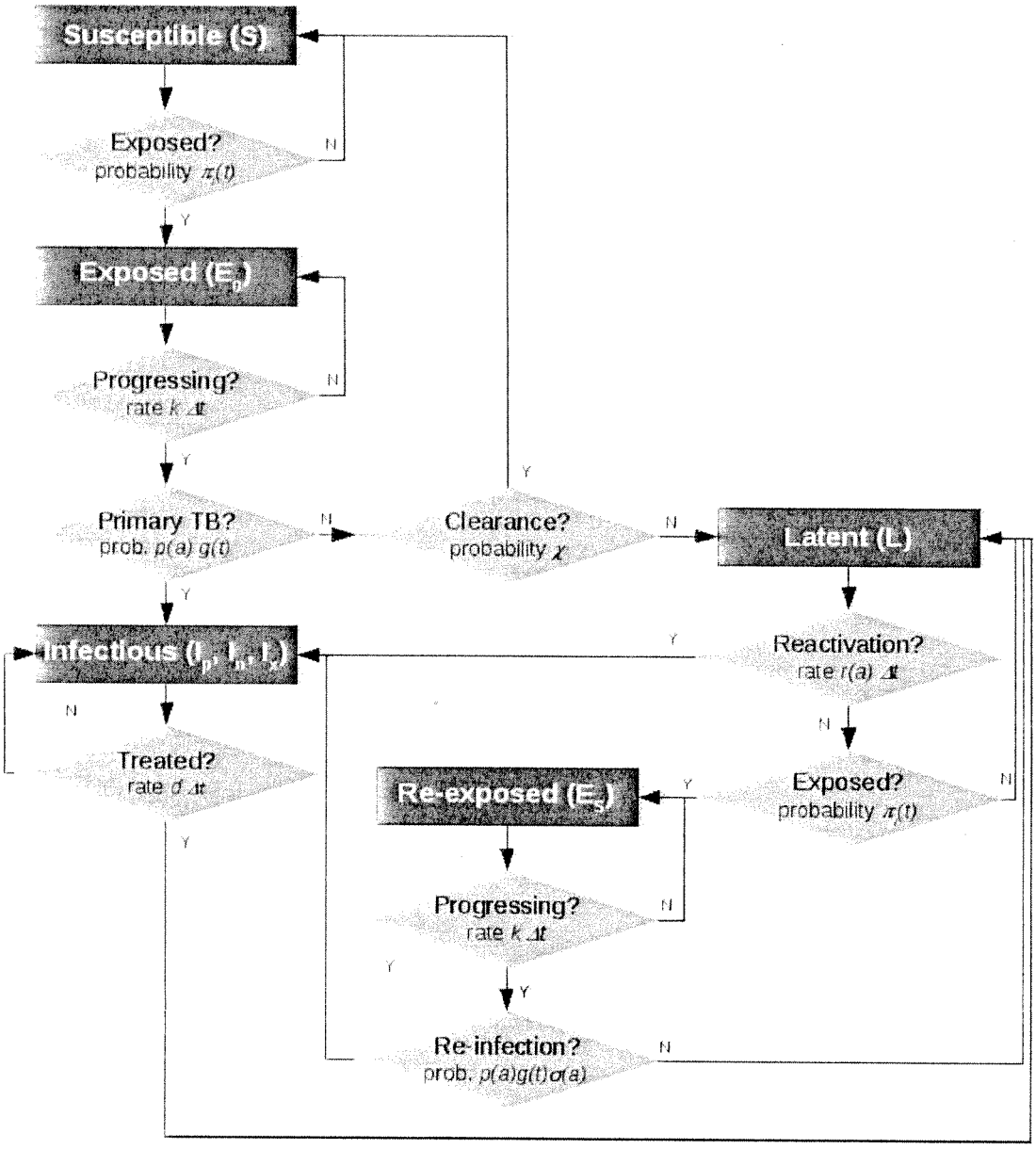

Figure 8: **Schematization of the epidemiological rules for Models B and C.** At each time step and for each individual, we start from one of the blue squares representing the current epidemiological status of the individual. The occurrence of one of the events described in the yellow diamonds is decided stochastically on the basis of the indicated rate or probability. The algorithm stops when it reaches a new (or possibly the same) blue square, representing the next epidemiological status of the individual.

dependent rate $r(a)\Delta t$.

4. All individuals who develop any form of active TB (I_p, I_n, I_x) may have either a positive or negative smear status, affecting their degree of infectiousness. The age-dependent probability of smear positive TB is denoted with $f^+(a)$.

5. Any individual with active TB (with no distinction between primary, endogenous and exogenous) can recover from the infection at a rate $d\Delta t$ or die from TB disease at a rate $\mu_T \Delta t$.

6. Any susceptible or latently infected individual i has a probability $p_i = 1 - e^{-\Delta t \cdot \lambda_i(t)}$ of inhaling the pathogen. The probability of this event depends on the instantaneous risk of infection $\lambda_i(t)$, computed at any time of the simulation (as in [6, 7, 8]).

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

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

     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 a smear-negative [22, 23].

   - 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 infections considered, namely:
     (a) contacts with infectious members of the household (first term in Eq. 2),
     (b) contacts with infectious individuals attending the same school or workplace, if any (second term in Eq. 2),
     (c) random contacts in the population (third term in Eq. 2).

$$\begin{aligned} \lambda_i &= \sum_{\{k=1,\ldots,N | H_k = H_i\}} \beta_H \frac{I_k s_k}{n_i} \\ &+ \sum_{\{k=1,\ldots,N | P_k = P_i\}} \beta_P \frac{I_k s_k}{m_i} \\ &+ \sum_{\{k=1,\ldots,N\}} \beta_R^{(C)} \frac{I_k s_k f(d_{ik})}{\sum_{\{k=1,\ldots,N\}} f(d_{ik})} \end{aligned} \tag{2}$$

     The terms in Eq. 2 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. It makes the casual transmission of TB in the general community explicitly dependent on distance through patterns of commuting ([24, 6, 7, 8]).

1.5.1 Functional forms for age and time dependent risks

To account for age and time-dependent risks of TB infection , general functional forms have been adopted, whose actual profile is determined by two parameters. Here we discuss choices on such forms (once again, we refer to only models B and C, since model A lacks an age structure).

- The functional form adopted for the age-dependent probability $p(a)$ of developing primary TB is a piece-wise linear function inspired by [21]. It can be described analytically as:

$$p(a) = \begin{cases} p_c & a \leq 10\ yrs \\ a\frac{(p_a - p_c)}{10} + 2p_c - p_a & 10\ yrs < a < 20\ yrs \\ p_a & a \geq 20\ yrs \end{cases}$$

An example of the function is reported in Figure 9. The linear segment in ages 10 to 20 is chosen in such a way that the function is continuous. p_c and p_a represent the risk of primary disease in children and adults respectively.

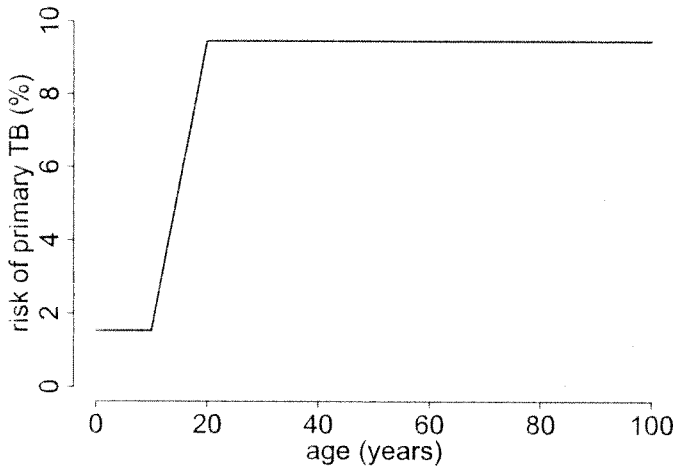

Figure 9: Age-dependent risk of developing primary TB upon exposure.

- The time-from-exposure-dependent relative risk of primary TB $g(t)$ proposed in [21] is a step-wise function of the year from exposure. It declined from a maximum of 1 in the first year to 0 after the fifth year from exposure. Rather than assuming a sudden decline in the risk at the end of each year from exposure, we used a negative exponential with equation: $g(t) = Ae^{-\gamma t}$

- For the risk of reactivation, $r(a)$ the same piece-wise functional form adopted for $p(a)$ has been previously proposed in [21]. However, such form can not account for the increased risk of reactivation in the elderly, observed in a recent US study [25]. This increased risk is also suggested by the trend of growth of the age-specific incidence data in Arkansas [26] (shown in Figure 3 in the main text). Therefore, we proposed a different functional form, having a linear growth in the first 50 years and a parabolic growth in the last 50 (Figure 10). The functional form for $r(a)$ can be analytically expressed as:

$$r(a) = r_m a + r_M (a - 50)^2 H(a - 50) \tag{3}$$

where $H(x)$ is the Heaveside unit step function:

$$H(x) = \begin{cases} 0 & x \leq 0 \\ 1 & x > 0 \end{cases}$$

and r_m, r_M are parameters of the function. Given the average risk of reactivation over ages r and one of the two parameters, the other can be back-calculated by inserting Eq. 3 in the definition of r:

$$r = \frac{1}{100} \sum_{a=1}^{100} r(a)$$

Therefore, we used r as model parameter, since reliable estimates of this quantity are available from data [25]. We chose the slope of the reactivation risk in younger ages, r_m,

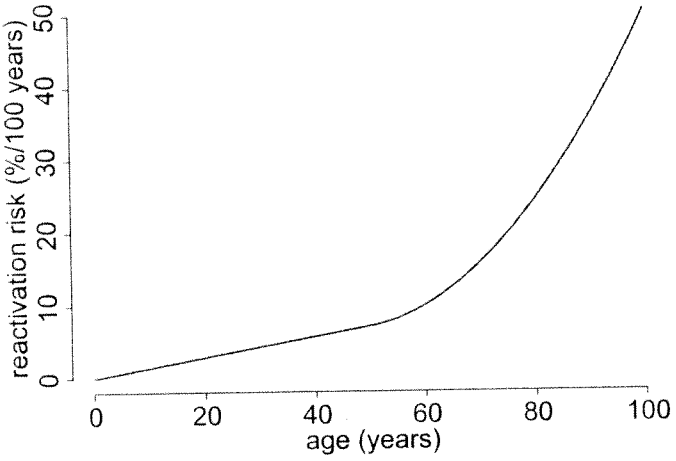

Figure 10: **Age-dependent risk of reactivating latent TB infection.**

as the second model parameter related to this functional form. Given r, higher values of r_m force the concavity in older ages, r_M, to be lower, and vice-versa: thus, r_m acts as a shape factor for $r(a)$.

- The functional form adopted for the age-specific protection from re-infection, $\sigma(a)$ is defined in the same way as $p(a)$, following [21]:

$$\sigma(a) = \begin{cases} \sigma_c & a \leq 10\ yrs \\ a\frac{(\sigma_a - \sigma_c)}{10} + 2\sigma_c - \sigma_a & 10\ yrs < a \leq 20\ yrs \\ \sigma_a & a \geq 20\ yrs \end{cases}$$

- Another piece-wise linear functional form was used for the age-specific risk of smear positive TB, as suggested by [21] and reported in Eq. 4.

$$f^+(a) = \begin{cases} f_c^+ & a \leq 5\ yrs \\ a\frac{(f_a^+ - f_c^+)}{10} + 2f_c^+ - f_a^+ & 5\ yrs < a \leq 25\ yrs \\ \sigma_a & a \geq 25\ yrs \end{cases} \quad (4)$$

2 Parametrization of the models

In this section we discuss choices of parameter values and ranges for the three models (Sections 2.1 and 2.2) and give further details on the parametrization process (Section 2.4).

2.1 Model A

- μ (natural death rate): we used a value of $1.25 \cdot 10^{-2}\ yrs^{-1}$, so to have an average lifetime of $\mu^{-1} = 80\ yrs$ (Table 1 in the main text);

- b (birth rate): we used a value of 33,750, so to obtain a constant population of $b/\mu = 2,700,000$ individuals, close to Arkansas values in the considered period [4] (Table 1 in the main text);

- μ_T (TB death rate): this quantity was calculated as the ratio between the yearly TB-induced deaths and the prevalent cases. The ratio is remarkably constant for USA data in 1990-2007 [27], with an average value of $0.133\ yrs^{-1}$ (standard deviation $7.38 \cdot 10^{-4}$, range $0.1320 - 0.1346$). The average value was used for all models (see Table 1 in main text).

- k (rate of progression to outcome): it is the inverse of the average time needed by an individual to resolve the outcome of its immune response. We used 4 months as the lower bound and 5 years as the upper bound for this time (as suggested by [28, 21]), corresponding to the values of k given in Table 5.

- d (treatment rate): its range was estimated from data, given the approximate relationship (TB-related death rate is neglected):

$$P(y+1) = P(y)(1 - d(y)) + I(y)$$

where P is the observed prevalence, I the observed incidence, and y represents any given year. Back-calculating d for all years from 1990-2007 data for the US [27], we obtain an average of 1.38 yrs^{-1} (standard deviation 0.0276, range $1.30 - 1.42$). Therefore, we let the parameter free to vary between 1.2 and 1.5 (Table 5).

- r (reactivation risk): we used the estimate of 10-16% per 100 years, as suggested in the literature [25] (Table 5).

- p (risk of developing primary TB): it is consistently suggested in the literature as 10 to 15% (e.g. [29, 21]); we used a minimum of 3% and a maximum of 20% (Table 5).

- β_R (transmission rate through casual contacts): Realistic values of this parameter can be estimated, for each value of χ, from the following formula:

$$ARI = \beta_R(1 - \chi)I$$

using as ARI the annual risk of infection, estimated for 1995 by [30, 31] at 0.03% and the corresponding prevalence I in the same year (6.11 per 100,000 individuals [26]). However, we let the value of β_R free to explore a broad range to give more flexibility to the model (Table 5).

- χ (proportion of cleared infections): since no information is available from data, the full range of possible values (1 to 99%) was adopted (Table 5).

- σ (protection from re-infection): the full range of possible values (1 to 99%) was also adopted (Table 5).

2.2 Models B and C

Fixed parameters.

- μ_T (TB death rate): see estimates for model A in Section 2.1;

- A, γ (parameters of the time-from-exposure-dependent risk of developing primary TB): the two parameters were estimated from the average risk by year reported in [21]. Using $A = 1.54$ and $\gamma = 0.92$ (Table 1 in main text) produces a good approximation of those figures, as shown in Figure 11.

- χ (proportion of cleared infection): No information is available in the literature on the proportion of cleared infections. This parameter affects the number of latently infected individuals and therefore the number of reactivations; however, in the time-span of 7 years considered in this study, very few recently infected individuals will eventually

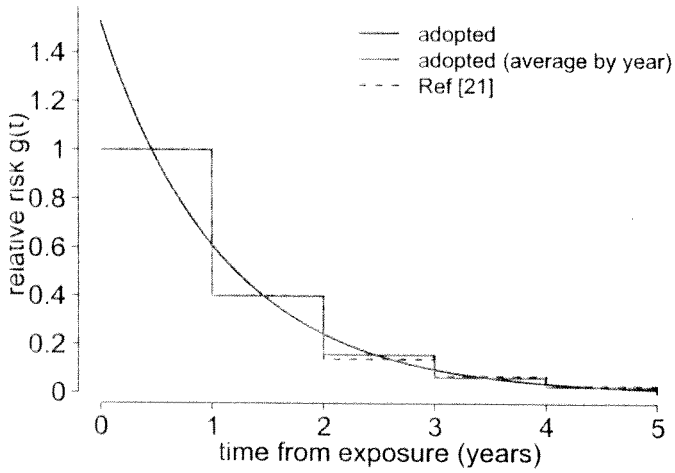

Figure 11: **Time-from-exposure-dependent risk of developing primary TB upon exposure.** The adopted curve (blue solid line) is a continuous version of the relative risk curve used in [21] (gold dashed line). Averaging its value by year from exposure (green solid line), we obtain estimates very close to values from [21].

reactivate. We expect this parameter to have a negligible effect: therefore, we set it to the best-fitting value of 68% suggested by calibration of Model A (Table 1 in the main text).

- σ_c, σ_a (protection from exogenous re-infection in children and adults): although few estimates are available for these parameters, re-exposure in our simulations is very low, due to the low prevalence of both TB cases and latently infected individuals. We expect that the uncertainty on this parameter will not significantly impact the model output, at least in low burden settings, and therefore we fix these parameters. We used $\sigma_c = 0\%$, $\sigma_a = 40\%$, based on the risks of exogenous re-infection estimated in [21] from UK data in the 1950s [32] (Table 1 in the main text).

- f_c^+, f_a^+ (age-specific proportion of smear positive cases in children and adults): these parameters were estimated from data on the proportions of smear positive cases [33], averaged in age groups [0,5] and [25,100], respectively (Table 1 in the main text). Figure 12 shows a comparison between the adopted function and data. Comparing to the corresponding estimate by [21], based on Norwegian data between 1951 and 1969, the observed proportion of smear positive cases is at least halved at all ages. This reduction can be explained with the earlier detection of TB in modern days, which decreases the number of patients diagnosed after smear conversion.

Parameters with ranges estimated from data (subset P1).

- k, d, r: see estimates for model A in Section 2.1 and Table 6.

- p_c, p_a (proportions of primary disease in children and adults): we chose a broad range that included estimates from [21] (respectively 4.1 and 13.8%) (Table 6).

Parameters with unknown ranges (subset P2).

- $\beta_H, \beta_P, \beta_R$ (transmission rates in households, school/workplaces and random contacts): since few data are available from literature, these parameters were left free to vary over a broad range. In the case of Model B, $\beta_R^{(B)}$ had to account for all transmission events, and therefore had an even larger range of variability with respect to the transmission rates in Model C (Table 6).

- r_m (slope of linear growth for the reactivation function): we chose the range of this parameter to span from very low values to the maximum value that guaranteed r_M to be positive with any value of r (Table 6).

2.3 Initialization of exposed individuals

Since reinfection is rare in Arkansas, we assume that all exposed individuals at the end of 1999 (the beginning of our simulations) are exposed for the first time (E_0). For each parameter set, we calculate the number of initially exposed individuals by inverting the approximated formula:

$$I(2000)N = rL(2000) + pkE_0(1999);$$

Thus, we obtain

$$E_0(1999) = \frac{I(2000)N - rL(2000)}{pk};$$

This choice of the initial number of exposed individuals forces the simulated initial incidence, $I(2000)$, to be close to actual epidemiological data for each given instance of parameters r, p and k. In the case of models B and C, p represents the average over ages of $p(a)$; we also use data on the age group distribution of initial latently infected individuals $L(a, 2000)$ and $r(a)$ in the place of its average value r.

Table 5: **Ranges of free parameters for model A.**

Par.	Unit	Description	Min	Max
k	yrs^{-1}	rate of progression to outcome	0.2	3
d	yrs^{-1}	treatment rate	1.2	1.5
r	yrs^{-1}	average reactivation rate	$1 \cdot 10^{-3}$	$1.6 \cdot 10^{-3}$
p	%	proportion of primary TB	3	20
$\backslash$	%	proportion of cleared infections	1	99
σ	%	protection from re-infection	1	99
β_R	yrs^{-1}	transmission rate in casual contacts	$1 \cdot 10^{-3}$	200

2.4 Parametrization procedure for models B and C

As described in Section 2.4 of the main text, the parameter search for models B and C was composed of two steps: a global search with LHS sampling size $Q = 10,000$, and a local

Table 6: **Ranges of free parameters for models B and C.**

Par.	Model	Unit	Description	Min	Max
			Subset P1		
k	B,C	yrs^{-1}	rate of progression to outcome	0.2	3
d	B,C	yrs^{-1}	treatment rate	1.2	1.5
r	B,C	yrs^{-1}	average reactivation rate	$1 \cdot 10^{-3}$	$1.6 \cdot 10^{-3}$
p_e	B,C	%	proportion of primary TB in ≤ 10 years	1	10
p_a	B,C	%	proportion of primary TB in ≥ 20 years	6	20
			Subset P2		
r_m	B,C	yrs^{-2}	slope of reactivation rate	$1 \cdot 10^{-6}$	$2 \cdot 10^{-5}$
β_R	B	yrs^{-1}	transmission rate in casual contacts	0.5	250
	C			0.5	50
β_H	C	yrs^{-1}	transmission rate in households	1	100
β_P	C	yrs^{-1}	transmission rate in schools and workplaces	1	100

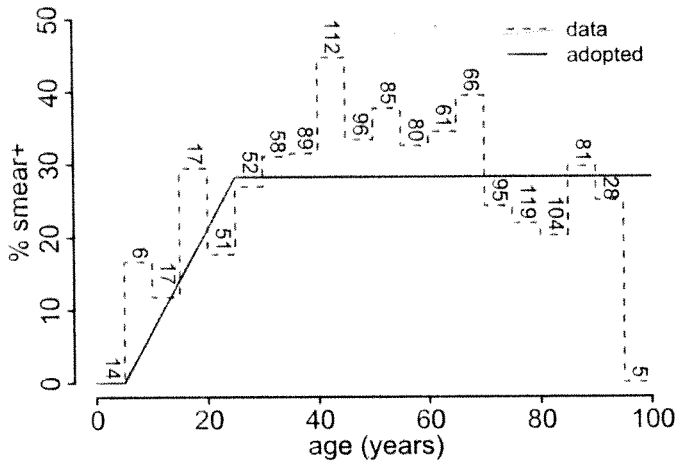

Figure 12: **Proportion of smear positive cases in Arkansas data and adopted in the model.** Data from TB cases in 1996-2003 [33] have been divided by 5-years age groups, and the corresponding proportion of smear positive cases has been displayed with a yellow dashed line. The sample size for each age class is reported with a number above each data-points. The blue solid line represents the probability adopted in Models B and C.

search around the best minimum ($Q^{loc} = 500$). In Figure 13 we plot the values of the score function for each of the Q experiments on model B, sorted by increasing value of F. We then selected the 65 parameter sets yielding a score function below a threshold F_{thr} (represented by the dashed red line in Figure 13), corresponding to the first percentile of the distribution of F. We analyzed the corresponding distributions of best-fitting parameters, and found that those belonging to subset P_1 all span the whole variability range proposed in Table 6: this suggests that any value in the estimated range could equally produce a good fit of the data. In contrast, the distributions of parameters from subset P_2 revealed a restricted range of suitable

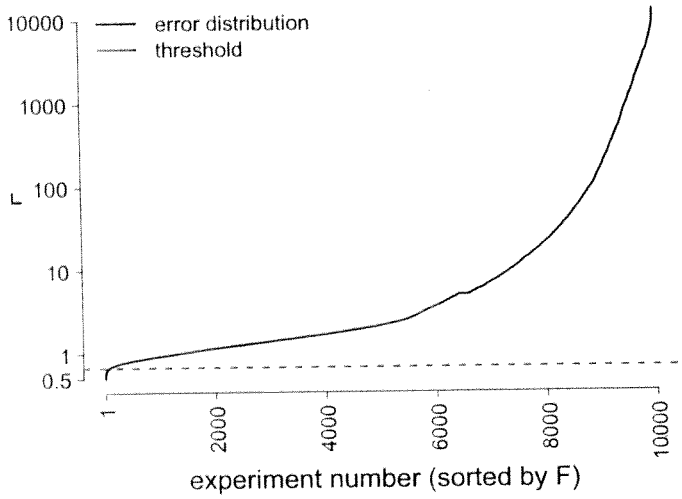

Figure 13: **Distribution of the score F by parameter set in the global search.** Experiments (each corresponding to a different parameters set) are sorted by increasing F scores; the dashed red line indicates the reference threshold for the analysis of best-fitting parameters.

values in the 65 best-fitting simulations, as shown in Figures 14 and 15. Similar results were obtained from a global search with $Q = 10,000$ on model C, leading to Figures 16, 17 and 18 (the histogram for r_m was very similar to Figure 15 and is not reported).
These distributions were used to restrict the ranges of variability of parameters from subset P_2. For β_R (in both models B and C), β_H and β_P, the lower limit of the range was kept identical to the global search, and the upper limit was set to the 90-percentile of the corresponding distribution of best-fitting values; for r_m (in both models B and C) we sampled values starting from the 10- percentile of the distribution, keeping the upper limit unchanged. See Table 7 for restricted parameter ranges in the local searches. The only significant correlations among best fitting parameters were found between k and β_R (correlation coefficient $R = -0.59$) and between β_R and p ($R = -0.49$).

Table 7: **Ranges of P_2 parameters in the local search for models B and C.**

Par.	Model	Unit	Description	Min	Max
r_m	B,C	yrs^{-2}	slope of reactivation rate	$1.3 \cdot 10^{-5}$	$2.0 \cdot 10^{-5}$
β_R	B	yrs^{-1}	transmission rate in casual contacts	0.5	50
	C			0.5	15
β_H	C	yrs^{-1}	transmission rate in households	1	35
β_P	C	yrs^{-1}	transmission rate in schools and workplaces	1	25

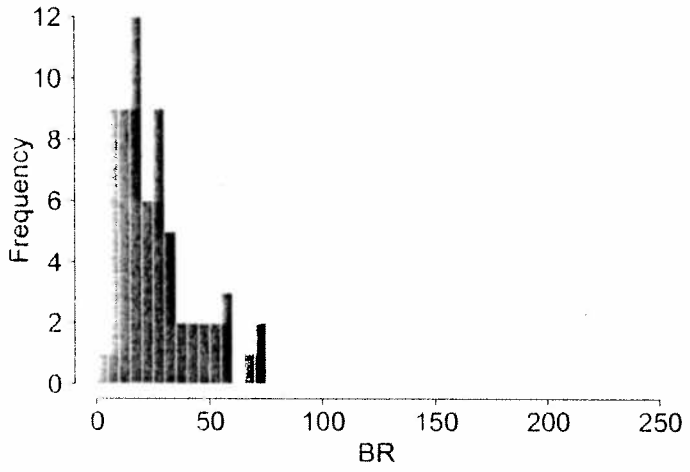

Figure 14: Distribution of $\beta_R^{(B)}$ values from best-fitting parameter sets ($F \leq F_{thr}$) in the global search using model B.

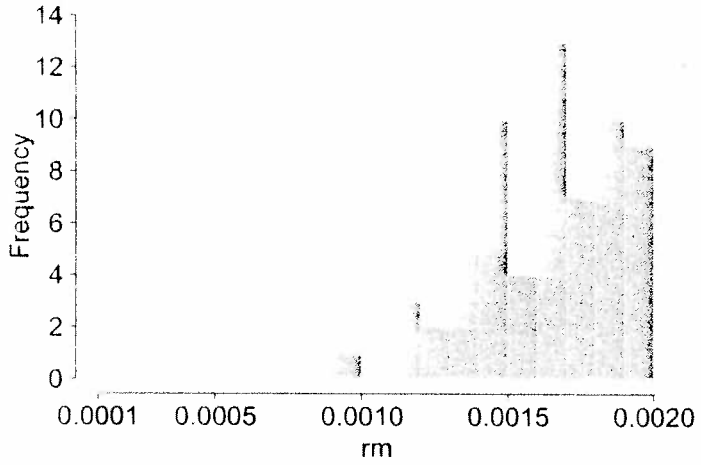

Figure 15: Distribution of r_m values from best-fitting parameter sets ($F \leq F_{thr}$) in the global search using model B.

3 Realizations and resources

The model was implemented in C language. For running a simulation on the Arkansas population (almost 3 million individuals explicitly represented in the model) the amount of RAM memory required is about 350 MB. A 10-year simulation takes about 1 minute on a single CPU of an Intel Xeon 2.67-GHz.

References

[1] U.S. Bureau of Census. Download Center. `http://factfinder.census.gov/servlet/`

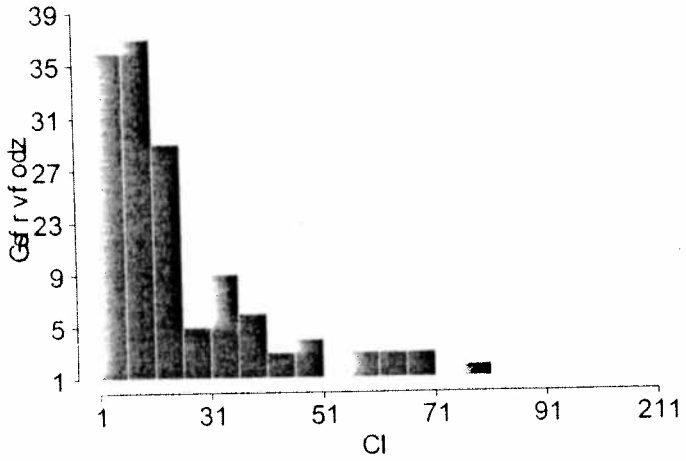

Figure 16: Distribution of λ_H values from best-fitting parameter sets ($F \leq F_{thr}$) in the global search using model C.

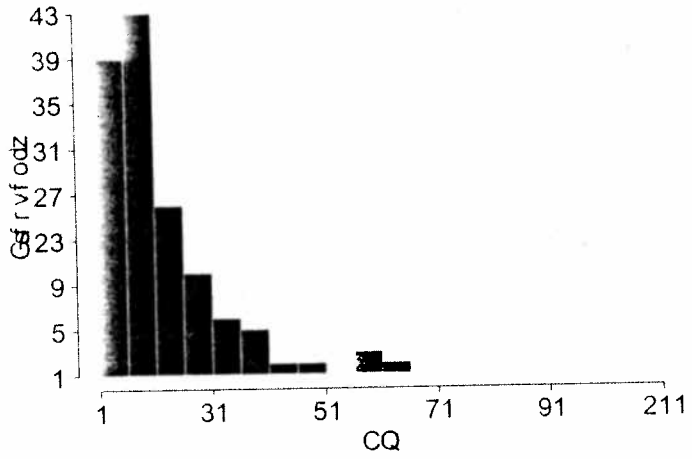

Figure 17: Distribution of λ_P values from best-fitting parameter sets ($F \leq F_{thr}$) in the global search using model C.

DownloadDatasetServlet.

[2] U.S. Bureau of Census and U.S. Bureau of Labor Statistics. Current Population Survey. http://www.census.gov/cps/.

[3] D. Balk and G. Yetman. The global distribution of population: evaluating the gains in resolution refinement. Technical report, Center for International Earth Science Information Network, 2004.

[4] U.S. Bureau of Census. Annual Population Estimates. http://www.census.gov/popest/states/tables/NST-EST2008-01.csv.

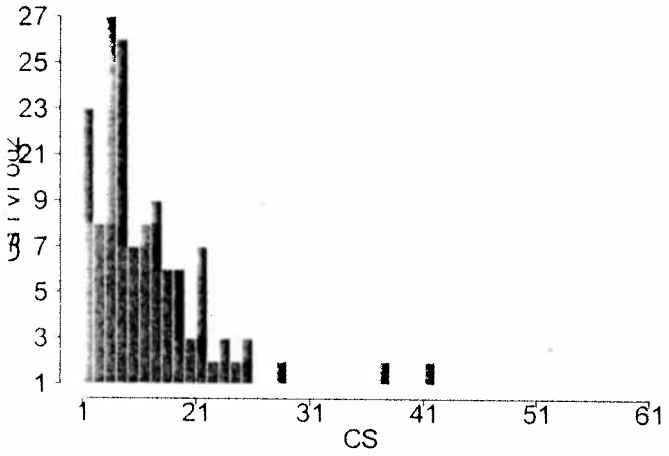

Figure 18: **Distribution of $\beta_R^{(C)}$ values from best-fitting parameter sets ($F \leq F_{thr}$) in the global search using model C.**

[5] D. L. Cohn. Treatment of Latent Tuberculosis Infection: Renewed Opportunity for Tuberculosis Control. *Clincal Infectious Diseases*, 31:120–124, 2000.

[6] M.L. Ciofi degli Atti, S. Merler, C. Rizzo, M. Ajelli, M. Massari, P. Manfredi, C. Furlanello, G. Scalia Tomba, and M. Iannelli. Mitigation measures for pandemic influenza in Italy: an individual based model considering different scenarios. *PLoS ONE*, 3(3):e1790, 2008.

[7] M. Ajelli and S. Merler. An individual-based model of hepatitis A transmission. *Journal of Theoretical Biology*, 259(3):478–488, 2009.

[8] S. Merler and M. Ajelli. The role of population heterogeneity and human mobility in the spread of pandemic influenza. *Proceedings of the Royal Society B*, 277(1681):557–565, 2010.

[9] U.S. National Center for Education Statistics. Main Website. `http://nces.ed.gov/`.

[10] U.S. Bureau of Census. North American Industry Classification System. `http://censtats.census.gov/cbpnaic/cbpnaic.shtml`.

[11] U.S. Bureau of Labor Statistics. Main Website. `http://www.bls.gov/`.

[12] C. Viboud, O. N. Bjrnstad, D. L. Smith, L. Simonsen, M. A. Miller, and B. T. Grenfell. Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science*, 312(5772):447–451, 2006.

[13] U.S. Bureau of Census. Year 2000 County-To-County Worker Flow Files. `http://www.census.gov/population/www/cen2000/commuting/index.html`.

[14] J. A. Nelder and R. Mead. A simplex algorithm for function minimization. *Computer Journal*, 7:308–313, 1965.

[15] Research Connection. Child Care and Early Education. http://www.researchconnections.org/childcare/datatools-02.jsp.

[16] U.S. Center for Disease Control and Prevention. National Vital Statistics System. http://www.cdc.gov/nchs/nvss.htm, accessed November 30, 2010.

[17] University of Columbia. Gridded Population of the World - version 3. http://sedac.ciesin.columbia.edu/gpw/, accessed November 30, 2010.

[18] U.S. National Center for Education Statistics. Main Website. http://nces.ed.gov/, accessed November 30, 2010.

[19] U.S. Bureau of Census - Population Division. Interim State Population Projections. http://www.census.gov/population/www/ projections/projectionsagesex.html, accessed November 30, 2010.

[20] Arkansas Department of Health. Vital Statistics Highlights. http://www.healthy.arkansas.gov/stats/2005vital/1-4.htm, accessed November 30, 2010.

[21] E. Vynnycky and P. E. M. Fine. The natural history of tuberculosis: the implications of age-dependent risks of disease and the role of reinfection. *Epidemiology and Infection*, 119:183–201, 1997.

[22] M. D. Behr, S. A. Warren, H. Salamon, P. C. Hopewell, A. Ponce de Leon, C. L. Daley, and P. M. Small. Transmission of Mycobacterium tuberculosis from patients smear-negative for acid-fast bacilli. *Lancet*, 353(9151):444–449, 1999.

[23] L. J. Abu-Raddad, L. Sabatelli, J.T. Achterberg, J.D. Sugimoto, I.M. Longini, C. Dye, and M.E. Halloran. Epidemiological benefits of more-effective tuberculosis vaccines, drugs, and diagnostics. *PNAS*, 106(33):13980–13985, 2009.

[24] N. M. Ferguson, D. A. T. Cummings, S. Cauchemez, C. Fraser, S. Riley, A. Meeyai, S. Iamsirithaworn, and D. S. Burke. Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature*, 437:209–214, 2005.

[25] C. R. Horsburgh, M. O'Donnell, S. Chamblee, J. L. Moreland, J. Johnson, B. J. Marsh, M. Narita, L. Scoles Johnson, and C. Fordham von Reyn. Revisiting rates of reactivation tuberculosis. *American Journal of Respiratory and Critical Care Medicine*, 182(420):425, 2010.

[26] U.S. Center for Disease Control and Prevention. Online Tuberculosis Information System. http://wonder.cdc.gov/tb.html.

[27] World Health Organization. Global Tuberculosis Control: Epidemiology, Strategy, Financing. Report 2009. Technical report, 2009.

[28] D. Sud, C. Bigbee, J. A. L. Flynn, and D. E. Kirschner. Contribution of CD8 T Cells to control of Mycobacterium tuberculosis infection. *The Journal of Immunology*, 176:4296–4314, 2006.

[29] W. W. Stead, J. P. Lofgren, E. Warren, and C. Thomas. Tuberculosis as an endemic and nosocomial infection among the elderly in nursing homes. *New England Journal of Medicine*, 312(1483):1487, 1985.

[30] T. M. Daniel and S. M. Debanne. Estimation of the Annual Risk of Tuberculous Infection for White Men in the United States. *Journal of Infectious Diseases*, 175(6):1535–1537, 1997.

[31] E. E. Salpeter and S. R. Salpeter. Mathematical Model for the Epidemiology of Tuberculosis, with Estimates of the Reproductive Number and Infection-Delay Function. *American Journal of Epidemiology*, 142(4):398–406, 1998.

[32] I. Sutherland. The ten-year incidence of clinical tuberculosis following "conversion" in 2,550 individuals aged 14 to 19 years. TRSU Progress Report. Technical report, KNCV, The Hague, The Netherlands, 1968.

[33] A. M. France, M. D. Cave, J. H. Bates, B. Foxman, T. Chu, and Z. Yang. 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*, 166(6):662–671, 2007.

Elenco dei report pubblicati

- ☐ **n. 52** Paolo Manca, Investimenti e finanziamenti generalizzati.
- ☐ **n. 53** Laura Lecchini - Odo Barsotti, Le rôle des immigrés extra- communautaires dans le marché du travail
- ☐ **n. 54** Riccardo Cambini, Alcune condizioni di ottimalità relative ad un insieme stellato.
- ☐ **n. 55** Gilberto Ghilardi, Uno schema di campionamento areale per le rilevazionninsulle famiglie in Italia.
- ☐ **n. 56** Riccardo Cambini, Studio di una classe di problemi non lineari: un metodo sequenziale.
- ☐ **n. 57** Riccardo Cambini, Una nota sulle possibili estensioni a funzioni vettoriali di significative classi di funzioni concavo-generalizzate.
- ☐ **n. 58** Alberto Bonaguidi - Valerio Terra Abrami, Metropolitan aging transition and metropolitan redistribution of the elderly in Italy.
- ☐ **n. 59** Odo Barsotti - Laura Lecchini, A comparison of male and female migration strategies: the cases of African and Filipino Migrants to Italy.
- ☐ **n. 60** Gilberto Ghilardi, Un modello logit per lo studio del fenomeno delle nuove imprese.
- ☐ **n. 61** S. Schaible, Generalized monotonicity.
- ☐ **n. 62** Vincenzo Bruno, Dell'elasticità in economia e dell'incertezza statistica.
- ☐ **n. 63** Laura Martein, Alcune classi di funzioni concave generalizzate nell'ottimizzazione vettoriale
- ☐ **n. 64** Anna Marchi, On the relationships between bicriteria problems and non-linear programming problems.
- ☐ **n. 65** Giovanni Boletto, Considerazioni metodologiche sul concetto di elasticità prefissata.
- ☐ **n. 66** Laura Martein, Soluzione efficienti e condizioni di ottimalità nell'otti-mizzazione vettoriale.

Anno: **1993**

- ☐ **n. 67** Maria Francesca Romano, Le rilevazioni ufficiali ISTAT della popolazione universitaria: problemi e definizioni alternative.
- ☐ **n. 68** Marco Bottai - Odo Barsotti, La ricerca "Spazio Utilizzato" Obiettivi e primi risultati.
- ☐ **n. 69** Marco Bottai - F Bartiaux, Composizione familiare e mobilità delle persone anziane. Una analisi regionale.
- ☐ **n. 70** Anna Marchi - Claudio Sodini, An algorithm for a non-differentiable non -linear fractional programming problem.
- ☐ **n. 71** Claudio Sodini - S.Schaible, An finite algorithm for generalized linear multiplicative programming.
- ☐ **n. 72** Alberto Cambini - Laura Martein, An approach to optimality conditions in vector and scalar optimization.
- ☐ **n. 73** Alberto Cambini - Laura Martein, Generalized concavity and optimality conditions in vector and scalar optimization.
- ☐ **n. 74** Riccardo Cambini, Alcune nuove classi di funzioni concavo-generalizzate.

Anno: **1994**

- ☐ **n. 75** Alberto Cambini - Anna Marchi - Laura Martein, On nonlinear scalarization in vector optimization.
- ☐ **n. 76** Maria Francesca Romano - Giovanna Nencioni, Analisi delle carriere degli studenti immatricolati dal 1980 al 1982.
- ☐ **n. 77** Gilberto Ghilardi, Indici statistici della congiuntura.
- ☐ **n. 78** Riccardo Cambini, Condizioni di efficienza locale nella ottimizzazione vettoriale.
- ☐ **n. 79** Odo Barsotti - Marco Bottai, Funzioni di utilizzazione dello spazio.
- ☐ **n. 80** Vincenzo Bruno, Alcuni aspetti dinamici della popolazione dei comuni della Toscana, distinti per ampiezza demografica e per classi di urbanità e di ruralità.
- ☐ **n. 81** Giovanni Boletto, I numeri indici del potere d'acquisto della moneta.
- ☐ **n. 82** Alberto Cambini - Laura Martein - Riccardo Cambini, Some optimality conditions in multiobjective programming.
- ☐ **n. 83** S. Schaible, Fractional programming with sume of ratios.
- ☐ **n. 84** Stefan Tigan - I.M.Stancu-Minasian, The minimun-risk approach for continuous time linear-fractional programming.
- ☐ **n. 85** Vasile Preda - I.M.Stancu-Minasian, On duality for multiobjective mathematical programming of n-set.
- ☐ **n. 86** Vasile Preda - I.M.Stancu-Minasian - Anton Batatorescu, Optimality and duality in nonlinear programming involving semilocally preinvex and related functions.

Anno: **1995**

- ☐ **n. 87** Elena Melis, Una nota storica sulla programmazione lineare: un problema di Kantorovich rivisto alla luce del problema degli zeri.
- ☐ **n. 88** Vincenzo Bruno, Mobilità territoriale dell'Italia e di tre Regioni tipiche: Lombardia, Toscana, Sicilia.
- ☐ **n. 89** Antonio Cortese, Bibliografia sulla presenza straniera in Italia
- ☐ **n. 90** Riccardo Cambini, Funzioni scalari affini generalizzate.
- ☐ **n. 91** Piero Manfredi - Fabio Tarini, Modelli epidemiologici: teoria e simulazione. (I)
- ☐ **n. 92** Marco Bottai - Maria Caputo - Laura Lecchini, The "OLIVAR" survey.Methodology and quality.
- ☐ **n. 93** Laura Lecchini - Donatella Marsiglia - Marco Bottai, Old people and social network.
- ☐ **n. 94** Gilberto Ghilardi, Uno studio empirico sul confronto tra alcuni indici statistici della congiuntura.
- ☐ **n. 95** Vincenzo Bruno, Il traffico nei porti italiani negli anni recenti.
- ☐ **n. 96** Alberto Cambini - Anna Marchi - Laura Martein - S. Schaible, An analysis of the falk-palocsay algorithm.
- ☐ **n. 97** Alberto Cambini - Laura Carosi, Sulla esistenza di elementi massimali.

Anno: **1996**

- ☐ **n. 98** Riccardo Cambini - S. Komlòsi, Generalized concavity and generalized monotonicity concepts for vector valued.
- ☐ **n. 99** Riccardo Cambini, Second order optimality conditions in the image space.
- ☐ **n. 100** Vincenzo Bruno, La stagionalità delle correnti di navigazione marittima.
- ☐ **n. 101** Eugene Maurice Cleur, A comparison of alternative discrete approximations of the Cox -I ngersoll - ross model.
- ☐ **n. 102** Gilberto Ghilardi, Sul calcolo del rapporto di concentrazione del Gini.
- ☐ **n. 103** Alberto Cambini - Laura Martein - Riccardo Cambini, A new approach to second order optimality conditions in vector optimization.
- ☐ **n. 104** Fausto Gozzi, Alcune osservazioni sull'immunizzazione semideterministica.
- ☐ **n. 105** Emilio Barucci - Fausto Gozzi, Innovation and capital accumulation in a vintage capital model an infinite dimensional control approach.
- ☐ **n. 106** Alberto Cambini - Laura Martein - I.M.Stancu-Minasian., A survey of bicriteria fractional problems.
- ☐ **n. 107** Luciano Fanti - Piero Manfredi, Viscosità dei salari, offerta di lavoro endogena e ciclo.
- ☐ **n. 108** Piero Manfredi - Luciano Fanti, Ciclo di vita di nuovi prodotti: modellistica non lineare.
- ☐ **n. 109** Piero Manfredi, Crescita con ciclo, gestazione dei piani di investimento ed effetti.

Elenco dei report pubblicati

- ☐ **n. 110** Luciano Fanti - Piero Manfredi, Un modello "classico" di ciclo con crescita ed offerta di lavoro endogena.
- ☐ **n. 111** Anna Marchi, On the connecttedness of the efficient frontier : sets without local maxima.
- ☐ **n. 112** Riccardo Cambini, Generalized concavity for bicriteria functions.
- ☐ **n. 113** Vincenzo Bruno, Variazioni dinamiche (1971-1981-1991)dei fenomeni demografici dei comuni (urbani e rurali) della Lombardia, in relazione ad alcune caratteristiche di mobilità territoriale.

Anno: **1997**

- ☐ **n. 114** Piero Manfredi - Fabio Tarini - J.R. Williams - A. Carducci - B. Casini, Infectious diseases: epidemiology, mathematical models, and immunization policies.
- ☐ **n. 115** Eugene Maurice Cleur - Piero Manfredi, One dimensional SDE models, low order numerical methods and simulation based estimation: a comparison of alternative estimators.
- ☐ **n. 116** Luciano Fanti - Piero Manfredi, Point stability versus orbital stability (or instability): remarks on policy implications in classical growth cycle model.
- ☐ **n. 117** Piero Manfredi - Francesco Billari, transition into adulthood, marriage, and timing of life in a stable popolation framework.
- ☐ **n. 118** Laura Carosi, Una nota sul concetto di estremo superiore di insiemi ordinati da coni convessi.
- ☐ **n. 119** Laura Lecchini - Donatella Marsiglia, Reti sociali degli anziani: selezione e qualità delle relazioni.
- ☐ **n. 120** Piero Manfredi - Luciano Fanti, Gestation lags and efficiency wage mechanisms in a goodwin type growth model.
- ☐ **n. 121** G.Rivellini, La metodologia statistica multilevel come possibile strumento per lo studio delle interazioni tra il comportamento procreativo individuale e il contesto
- ☐ **n. 122** Laura Carosi, Una nota sugli insiemi C-limitati e L-limitati.
- ☐ **n. 123** Laura Carosi, Sull'estremo superiore di una funzione lineare fratta ristretta ad un insieme chiuso e illimitato.
- ☐ **n. 124** Piero Manfredi, A demographic framework for the evaluation of the impact of imported infectious diseases.
- ☐ **n. 125** Alessandro Valentini, Calo della fecondità ed immigrazione: scenari e considerazioni sul caso italiano.
- ☐ **n. 126** Alberto Cambini - Laura Martein, Second order optimality conditions.

Anno: **1998**

- ☐ **n. 127** Piero Manfredi and Alessandro Valentini, Populations with below replacementfertility: theoretical considerations and scenarioes from the italian laboratory.
- ☐ **n. 128** Alberto Cambini - Laura Martein - E. Moretti, Programmazione frazionaria e problemi bicriteria.
- ☐ **n. 129** Emilio Barucci - Fausto Gozzi - Andrej Swiech, Incentive compatibility constraints and dynamic programming in continuous time.

Anno: **1999**

- ☐ **n. 130** Alessandro Valentini, Impatto delle immigrazioni sulla popolazione italiana: confronto tra scenari alternativi.
- ☐ **n. 131** K. Iglicka - Odo Barsotti - Laura Lecchini, Recent developement of migrations from Poland to Europe with a special emphsis on Italy K.Iglika - Le Migrazioni est-ovest: le unioni miste in Italia
- ☐ **n. 132** Alessandro Valentini, Proiezioni demografiche multiregionali a due sessi, con immigrazioni internazionali e vincoli di consistenza.
- ☐ **n. 133** Fabio Antonelli - Emilio Barucci - Maria Elvira Mancino, Backward-forward stochastic differential utility: existence, consumption and equilibrium analysis.
- ☐ **n. 134** Emilio Barucci - Maria Elvira Mancino, Asset pricing with endogenous aspirations.
- ☐ **n. 135** Eugene Maurice Cleur, Estimating a class of diffusion models: an evaluation of the effects of sampled discrete observations.
- ☐ **n. 136** Luciano Fanti - Piero Manfredi, Labour supply, time delays, and demoeconomic oscillations in a solow-typegrowth model.
- ☐ **n. 137** Emilio Barucci - Sergio Polidoro - Vincenzo Vespri, Some results on partial differential equations and Asian options.
- ☐ **n. 138** Emilio Barucci - Maria Elvira Mancino, Hedging european contingent claims in a Markovian incomplete market.
- ☐ **n. 139** Alessandro Valentini, L'applicazione del modello multiregionale-multistato alla popolazione in Italia mediante l'utilizzo del Lipro: procedura di adattamento dei dati e particolarità tecniche del programma.
- ☐ **n. 140** I.M.Stancu-Minasian, optimality conditions and duality in fractional programming-involving semilocally preinvex and related functions.
- ☐ **n. 141** Alessandro Valentini, Proiezioni demografiche con algoritmi di consistenza per la popolazione in Italia nel periodo 1997-2142: presentazione dei risultati e confronto con metodologie di stima alternative.
- ☐ **n. 142** Laura Carosi, Competitive equilibria with money and restricted partecipation.
- ☐ **n. 143** Laura Carosi, Monetary policy and Pareto improvability in a financial economy with restricted partecipation
- ☐ **n. 144** Bruno Cheli, Misurare il benessere e lo sviluppo dai paradossi del Pil a misure di benessere economico sostenibile, con uno sguardo allo sviluppo umano
- ☐ **n. 145** Bruno Cheli - Laura Lecchini - Lucio Masserini, The old people's perception of well-being: the role of material and non material resources
- ☐ **n. 146** Eugene Maurice Cleur, Maximum likelihood estimation of one-dimensional stochastic differential equation models from discrete data: some computational resultas
- ☐ **n. 147** Alessandro Valentini - Francesco Billari - Piero Manfredi, Utilizzi empirici di modelli multistato continui con durate multiple
- ☐ **n. 148** Francesco Billari - Piero Manfredi - Alberto Bonaguidi - Alessandro Valentini, Transition into adulthoold: its macro-demographic consequences in a multistatew stable population framework
- ☐ **n. 149** Francesco Billari - Piero Manfredi - Alessandro Valentini, Becoming Adult and its Macro-Demographic Impact: Multistate Stable Population Theory and an Application to Italy
- ☐ **n. 150** Alessandro Valentini, Le previsioni demografiche in presenza di immigrazioni: confronto tra modelli alternativi e loro utilizzo empirico ai fini della valutazione dell'equilibrio nel sistema pensionistico
- ☐ **n. 151** Emilio Barucci - Roberto Monte, Diffusion processes for asset prices under bounded rationality
- ☐ **n. 152** Emilio Barucci - P. Cianchi - L. Landi - A. Lombardi, Reti neurali e analisi delle serie storiche: un modello per la previsione del BTP future
- ☐ **n. 153** Alberto Cambini - Laura Carosi - Laura Martein, On the supremum in fractional programming
- ☐ **n. 154** Riccardo Cambini - Laura Martein, First and second order characterizations of a class of pseudoconcave vector functions
- ☐ **n. 155** Piero Manfredi and Luciano Fanti, Embedding population dynamics in macro-economic models. The case of the goodwin's growth cycle
- ☐ **n. 156** Laura Lecchini e Odo Barsotti, Migrazioni dei preti dalla Polonia in Italia
- ☐ **n. 157** Vincenzo Bruno, Analisi dei prezzi, in Italia dal 1975 in poi
- ☐ **n. 158** Vincenzo Bruno, Analisi del commercio al minuto in Italia
- ☐ **n. 159** Vincenzo Bruno, Aspetti ciclici della liquidità bancaria, dal 1971 in poi

Elenco dei report pubblicati

- ☐ **n. 160** Anna Marchi, A separation theorem in alternative theorems and vector optimization

Anno: **2000**

- ☐ **n. 161** Piero Manfredi and Luciano Fanti, Labour suppley, population dynamicics and persistent oscillations in a Goodwin-type growth cycle model
- ☐ **n. 162** Luciano Fanti and Piero Manfredi, Neo-classical labour market dynamics and chaos (and the Phillips curve revisited)
- ☐ **n. 163** Piero Manfredi - and Luciano Fanti, Detection of Hopf bifurcations in continuos-time macro- economic models, with an application to reducible delay-systems.
- ☐ **n. 164** Fabio Antonelli - Emilio Barucci, The Dynamics of pareto allocations with stochastic differential utility
- ☐ **n. 165** Eugene M. Cleur, Computing maximum likelihood estimates of a class of One-Dimensional stochastic differential equation models from discrete Date*
- ☐ **n. 166** Eugene M. Cleur, Estimatng the drift parameter in diffusion processes more efficiently at discrete times:a role of indirect estimation
- ☐ **n. 167** Emilio Barucci - Vincenzo Valori, Forecasting the forecasts of others e la Politica di Inflation targeting
- ☐ **n. 168** A.Cambini - L. Martein, First and second order optimality conditions in vector optimization
- ☐ **n. 169** A. Marchi, Theorems of the Alternative by way of Separation Theorems
- ☐ **n. 170** Emilio Barucci - Maria Elvira Mancino, Asset Pricing and Diversification with Partally Exchangeable random Variables
- ☐ **n. 171** Piero Manfredi - Luciano Fanti, Long Term Effects of the Efficiency Wage Hypothesis in Goodwin-Type Economies.
- ☐ **n. 172** Piero Manfredi - Luciano Fanti, Long Term Effects of the Efficiency wage Hypothesis in Goodwin-type Economies: a reply.
- ☐ **n. 173** Luciano Fanti, Innovazione Finanziaria e Domanda di Moneta in un Modello dinamico IS-LM con Accumulazione.
- ☐ **n. 174** P Manfredi, A Bonaccorsi, A Secchi, Social Heterogeneities in Classical New Product Diffusion Models. I: "External" and "Internal" Models.
- ☐ **n. 175** Piero Manfredi - Ernesto Salinelli, Modelli per formazione di coppie e modelli di Dinamica familiare.
- ☐ **n. 176** P Manfredi, E. Salinelli, A Melegaro, A Secchi, Long term Interference Between Demography and Epidemiology: the case of tuberculosis
- ☐ **n. 177** Piero Manfredi - Ernesto Salinelli, Toward the Development of an Age Structure Teory for Family Dynamics I: General Frame.
- ☐ **n. 178** Piero Manfredi - Luciano Fanti, Population heterogeneities, nonlinear oscillations and chaos in some Goodwin-type demo-economic models Paper to be presented at the Second workshop on "nonlinear demography" Max Planck Institute for demographic Research Rostock, Germany, May 31-June 2, 2
- ☐ **n. 179** E. Barucci - M.E. Mancini - Roberto Renò, Volatility Estimation via Fourier Analysis
- ☐ **n. 180** Riccardo Cambini, Minimum Principle Type Optimality Conditions
- ☐ **n. 181** E. Barucci, M. Giuli, R. Monte, Asset Prices under Bounded Rationality and Noise Trading
- ☐ **n. 182** A. Cambini, D.T Luc, L.Martein, Order Preserving Transformations and application.
- ☐ **n. 183** Vincenzo Bruno, Variazioni dinamiche (1971-1981-1991) dei fenomeni demografici dei comuni urbani e rurali della Sicilia, in relazione ad alcune caratteristiche di mobilità territoriale.
- ☐ **n. 184** F Antonelli, E.Barucci, M.E.Mancino, Asset Pricing with a Backward-Forward Stochastic Differential Utility
- ☐ **n. 185** Riccardo Cambini - Laura Carosi, Coercivity Concepts and Recession Functions in Constrained Problems
- ☐ **n. 186** John R. Williams, Piero Manfredi, The pre-vaccination dynamics of measles in Italy: estimating levels of under-reporting of measles cases
- ☐ **n. 187** Piero Manfredi, John R. Williams, To what extent can inter-regional migration perturbe local endemic patterns? Estimating numbers of measles cases in the Italian regions
- ☐ **n. 188** Laura Carosi, Johannes Jahn, Laura Martein, On The Connections between Semidefinite Optimization and Vector Optimization
- ☐ **n. 189** Alberto Cambini, Jean-Pierre Crouzeix, Laura Martein, On the Pseudoconvexity of a Quadratic Fractional Function
- ☐ **n. 190** Riccardo Cambini - Claudio Sodini, A finite Algorithm for a Particular d.c. Quadratic Programming Problem.
- ☐ **n. 191** Riccardo Cambini - Laura Carosi, Pseudoconvexity of a class of Quadratic Fractional Functions.
- ☐ **n. 192** Laura Carosi, A note on endogenous restricted partecipation on financial markets: an existence result.
- ☐ **n. 193** Emilio Barucci - Roberto Monte - Roberto Renò, Asset Price Anomalies under Bounded Rationality.
- ☐ **n. 194** Emilio Barucci - Roberto Renò, A Note on volatility estimate-forecast with GARCH models.
- ☐ **n. 195** Bruno Cheli, Sulla misura del benessere economico: i paradossi del PIL e le possibili correzioni in chiave etica e sostenibile, con uno spunto per l'analisi della povertà
- ☐ **n. 196** M.Bottai, M.Bottai, N. Salvati, M.Toigo, Le proiezioni demografiche con il programma Nostradamus. (Applicazione all'area pisana)
- ☐ **n. 197** A. Lemmi - B. Cheli - B. Mazzolli, La misura della povertà multidimensionale: aspetti metodologici e analisi della realtà italiana alla metà degli anni '90
- ☐ **n. 198** C.R. Bector - Riccardo Cambini, Generalized B-invex vector valued functions
- ☐ **n. 199** Luciano Fanti - Piero Manfredi, The workers' resistance to wage cuts is not necessarily detrimental for the economy: the case of a Goodwin's growth model with endogenous population.
- ☐ **n. 200** Emilio Barucci - Roberto Renò, On Measuring volatility of diffusion processes with high frequency data
- ☐ **n. 201** Piero Manfredi - Luciano Fanti, Demographic transition and balanced growth

Anno: **2001**

- ☐ **n. 202** E.Barucci - M. E. Mancini - E. Vannucci, Asset Pricing, Diversification and Risk Ordering with Partially Exchangeable random Variables
- ☐ **n. 203** E. Barucci - R. Renò - E. Vannucci, Executive Stock Options Evaluation.
- ☐ **n. 204** Odo Barsotti - Moreno Toigo, Dimensioni delle rimesse e variabili esplicative: un'indagine sulla collettività marocchina immigrata nella Toscana Occidentale
- ☐ **n. 205** Vincenzo Bruno, I Consumi voluttuari, nell'ultimo trentennio, in Italia
- ☐ **n. 206** Michele Longo, The monopolist choice of innovation adoption: A regular-singular stochastic control problem
- ☐ **n. 207** Michele Longo, The competitive choice of innovation adoption: A finite-fuel singular stochastic control problem.
- ☐ **n. 208** Riccardo Cambini - Laura Carosi, On the pseudoaffinity of a class of quadratic fractional functions
- ☐ **n. 209** Riccardo Cambini - Claudio Sodini, A Finite Algorithm for a Class of Non Linear Multiplicative Programs.
- ☐ **n. 210** Alberto Cambini - Dinh The Luc - Laura Martein, A method for calculating subdifferential Convex vector functions
- ☐ **n. 211** Alberto Cambini - Laura Martein, Pseudolinearity in scalar and vector optimization.
- ☐ **n. 212** Riccardo Cambini, Necessary Optimality Conditions in Vector Optimization.
- ☐ **n. 213** Riccardo Cambini - Laura Carosi, On generalized convexity of quadratic fractional functions.
- ☐ **n. 214** Riccardo Cambini - Claudio Sodini, A note on a particular quadratic programming problem.

Elenco dei report pubblicati

- ☐ **n. 215** Michele Longo - Vincenzo Valori, Existence end stability of equilibria in OLG models under adaptive expectations.
- ☐ **n. 216** Luciano Fanti - Piero Manfredi, Population, unemployment and economic growth cycles: a further explanatory perspective
- ☐ **n. 217** J.R.Williams,P Manfredi,S.Salmaso,M.Ciofi, Heterogeneity in regional notification patterns and its impact on aggregate national case notification data: the example of measles in Italy.
- ☐ **n. 218** Anna Marchi, On the connectedness of the efficient frontier: sets without local efficient maxima
- ☐ **n. 219** Laura Lecchini - Odo Barsotti, Les disparités territoriales au Maroc au travers d'une optique de genre.

Anno: 2002

- ☐ **n. 220** Gilberto Ghilardi - Nicola Orsini, Sull'uso dei modelli statistici lineari nella valutazione dei sistemi formativi.
- ☐ **n. 221** Andrea Mercatanti, Un'analisi descrittiva dei laureati dell'Università di Pisa
- ☐ **n. 222** E. Barucci - C. Impenna - R. Renò, The Italian Overnight Market: microstructure effects, the martingale hypothesis and the payment system.
- ☐ **n. 223** E. Barucci, P Malliavin, M.E.Mancino, R.Renò, A.Thalmaier, The Price-volatility feedback rate: an implementable mathematical indicator of market stability.
- ☐ **n. 224** Andrea Mercatanti, Missing at random in randomized experiments with imperfect compliance
- ☐ **n. 225** Andrea Mercatanti, Effetto dell'uso di carte Bancomat e carte di Credito sulla liquidità familiare: una valutazione empirica
- ☐ **n. 226** Piero Manfredi - John R. Williams, Population decline and population waves: their impact upon epidemic patterns and morbidity rates for childhood infectious diseases. Measles in Italy as an example.
- ☐ **n. 227** Piero Manfredi - Marta Ciofi degli Atti, La geografia pre-vaccinale del morbillo in Italia. I. Comportamenti di contatto e sforzi necessari all'eliminazione: predizioni dal modello base delle malattie prevenibili da vaccino.
- ☐ **n. 228** I.M.Stancu-Minasian, Optimality Conditions and Duality in Fractional Programming Involving Semilocally Preinvex and Related
- ☐ **n. 229** Nicola Salvati, Un software applicativo per un'analisi di dati sui marchi genetici (Genetic Markers)
- ☐ **n. 230** Piero Manfredi, J. R. Williams, E. M. Cleur, S. Salmaso, M. Ciofi, The pre-vaccination regional landscape of measles in Italy: contact patterns and related amount of needed eradication efforts (and the "EURO" conjecture)
- ☐ **n. 231** Andrea Mercatanti, I tempi di laurea presso l'Università di Pisa: un'applicazione dei modelli di durata in tempo discreto
- ☐ **n. 232** Andrea Mercatanti, The weak version of the exclusion restriction in causal effects estimation: a simulation study
- ☐ **n. 233** Riccardo Cambini and Laura Carosi, Duality in multiobjective optimization problems with set constraints
- ☐ **n. 234** Riccardo Cambini and Claudio Sodini, Decomposition methods for nonconvex quadratic programs
- ☐ **n. 235** R.Cambini and L. Carosi and S.Schaible, Duality in fractional optimization problems with set constraints
- ☐ **n. 236** Anna Marchi, On the mix-efficient points

Anno: 2003

- ☐ **n. 237** Emanuele Vannucci, The valuation of unit linked policies with minimal return guarantees under symmetric and asymmetric information hypoteses
- ☐ **n. 238** John R Williams - Piero Manfredi, Ageing populations and childhood infections: the potential impact on epidemic patterns and morbidity
- ☐ **n. 239** Bruno Cheli, Errata Corrige del Manuale delle Impronte Ecologiche (2002) ed alcuni utili chiarimenti
- ☐ **n. 240** Alessandra Petrucci-Nicola Salvati-Monica Pratesi, Stimatore Combinato r Correlazione Spaziale nella Stima per Piccole Aree
- ☐ **n. 241** Riccardo Cambini - Laura Carosi, Mixed Type Duality for Multiobjective Optimization Problems with set constraints
- ☐ **n. 242** O Barsotti, L.Lecchini, F.Benassi, Foreigners from central and eastern European countries in Italy: current and future perspectives of eu enlargement
- ☐ **n. 243** A. Cambini - L. Martein - S. Schaible, Pseudoconvexity under the Charnes-Cooper transformation
- ☐ **n. 244** Eugene M. Cleur, Piero Manfredi, and John R. William, The pre-and post-Vaccination regional dynamics of measles in Italy: Insights from time series analysis

Anno: 2004

- ☐ **n. 245** Emilio Barucci - Jury Falini, Determinants of Corporate Governance in Italy: Path dependence or convergence?
- ☐ **n. 246** R. Cambini - A. Marchi, A note on the connectedness of the efficient frontier
- ☐ **n. 247** Laura Carosi - Laura Martein, On the pseudoconvexity and pseudolinearity of some classes of fractional functions
- ☐ **n. 248** E. Barucci - R. Monte - B. Trivellato, Bayesian nash equilibrium for insider trading in continuous time
- ☐ **n. 249** Eugene M. Cleur, A Time Series Analysis of the Inter-Epidemic Period for Measles in Italy
- ☐ **n. 250** Andrea Mercatanti, Causal inference methods without exclusion restrictions: an economic application.
- ☐ **n. 251** Eugene M. Cleur, Non-Linearities in Monthly Measles data for Italy
- ☐ **n. 252** Eugene M. Cleur, A Treshold Model for Prevaccination Measles Data: Some Empirical Results for England and Italy
- ☐ **n. 253** Andrea Mercatanti, La gestione dei dati mancanti nei modelli di inferenza causale: il caso degli esperimenti naturali.
- ☐ **n. 254** Andrea Mercatanti, Rilevanza delle analisi di misture di distribuzioni nelle valutazioni di efficacia
- ☐ **n. 255** Andrea Mercatanti, Local estimation of mixtures in instrumental variables models
- ☐ **n. 256** Monica Pratesi - Nicola Salvati, Spatial EBLUP in agricultural surveys: an application based on italian census data.
- ☐ **n. 257** Emanuele Vannucci, A model analyzing the effects of information asymmetries of the traders
- ☐ **n. 258** Monica Pratesi-Emilia Rocco, Two-Step centre sampling for estimating elusive population size
- ☐ **n. 259** A.Lemmi, N.Pannuzi, P.Valentini, B.Cheli, G.Berti, Estimating Multidimensional Poverty: A Comparison of Three Diffused Methods°

Anno: 2005

- ☐ **n. 260** Nicola Salvati, Small Area estimation: the EBLUP estimator using the CAR model
- ☐ **n. 261** Monica Pratesi-Nicola Salvati, Small Area Estimation: the EBLUP estimator with autoregressive random area effects
- ☐ **n. 262** Riccardo Cambini-Claudio Sodini, A solution algorithm for a class of box constrained quadratic programming problems
- ☐ **n. 263** Andrea Mercatanti, A constrained likelihood maximization for relaxing the exclusion restriction in causal inference.
- ☐ **n. 264** Marco Bottai - Annalisa Lazzini - Nicola Salvati, Le proiezioni demografiche. Pisa 2003/2032
- ☐ **n. 265** Andrea Mercatanti, An exercise in estimating causal effects for non-compliers: the return to schooling in Germany and Austria
- ☐ **n. 266** Nicola Salvati, M-quantile Geographically Weighted Regression for Nonparametric Small Area Estimation
- ☐ **n. 267** Ester Rizzi, Alessandro Rosina, L'influsso della Luna sul comportamento sessuale

Elenco dei report pubblicati

- ☐ **n. 268** Silvia Venturi, Linda Porciani, Moreno Toigo, Federico Benassi, Il migrate nello spazio sociale transnazionale: tra integrazione nel Paese di destinazione e appartenenza al Paese di origine
- ☐ **n. 269** James Raymer, Alberto Bonaguidi, Alessandro Valentini, Describing and Projecting the Age and Spatial Structures of Interregional Migration in Italy
- ☐ **n. 270** Laura Carosi, Laura Martein, Some classes of pseudoconvex fractional functions via the Charnes-Cooper transformation
- ☐ **n. 271** Laura Carosi,Antonio Villanacci, Relative wealth dependent restricted participation on financial markets
- ☐ **n. 272** Riccardo Cambini, Claudio Sodini, A sequential method for a class of box constrained quadratic programming problems
- ☐ **n. 273** Riccardo Cambini, Rossana Riccardi, An approach to discrete convexity and its use in an optimal fleet mix problem
- ☐ **n. 274** Riccardo Cambini, Claudio Sodini, An unifying approach to solve a class of parametrically-convexifiable problems
- ☐ **n. 275** Paolo Manca, Misure di Rischio Finanziario
- ☐ **n. 276** Bruno Cheli e Gianna Righi, Rapporto sulle abitudini di consumo di acqua potabile nel Comune di Cecina
- ☐ **n. 277** Anna Marchi - Laura Martein, Pseudomonotonicity of an affine map and the two dimensional case
- ☐ **n. 278** Andrea Pallini, Bernstein-type approximation of smooth functions
- ☐ **n. 279** Ray Chambers, Monica Pratesi, Nicola Salvati, Nikos Tzavidis, Spatial M-quantile Models for Small Area Estimation

Anno: **2006**

- ☐ **n. 280** Franco Fineschi and Riccardo Giannetti, ADJOINTS OF A MATRIX
- ☐ **n. 281** Andrea Mercatanti, An ML procedure for partially identified Causal models
- ☐ **n. 282** Marco Geraci, Nicola Salvati, The geographical distribution of the consumption expenditure in Ecuador: Estimation and mapping of the regression quantiles
- ☐ **n. 283** Mauro Sodini, Labour supply in a polluted world
- ☐ **n. 284** Mauro Sodini, The Fragility of Social Capital: An Analytical Approach
- ☐ **n. 285** Mauro Sodini, An endogenous growth model with social capital
- ☐ **n. 286** Mauro Sodini, A two sectors growth model with social capital
- ☐ **n. 287** Monica Pratesi, M. Giovanna Ranalli, Nicola Salvati, Nonparametric M-quantile Regression using Penalized Splines
- ☐ **n. 288** Riccardo Cambini e Claudio Sodini, A computational comparison of some branch and bound methods for indefinite quadratic programs
- ☐ **n. 289** Riccardo Cambini, Multiobjective Problems with Set Constraints: from Necessary Optimality Conditions to Duality Results
- ☐ **n. 290** Il ruolo della complementarità stretta in programmazione matematica, Giorgio Giorgi
- ☐ **n. 291** Andrea Pallini, Bernstein-type approximation using the beta-binomial distribution
- ☐ **n. 292** Andrea Mercatanti, Identifiability and two-steps estimation procedures in casual models with ignorable assignments and non-ignorable compliance

Anno: **2007**

- ☐ **n. 293** Nikos Tzavidis, Nicola Salvati, Monica Pratesi, Ray Chambers, M-quantile Models with Application to Small Area Estimation and Poverty Mapping
- ☐ **n. 294** Andrea Pallini, Saturation and Superefficiency for some Approximation of the Bernstein Type
- ☐ **n. 295** Giorgio Guzzetta, Piero Manfredi, Estimation of the forces of infection in a complex epidemiological model for meningitis using genetic algorithms
- ☐ **n. 296** Emanuele Del Fava, Piero Manfredi, Strange phenomena in the most basic inferential procedure: interval estimation for a binominal proportion
- ☐ **n. 297** Odo Barsotti, Federico Benassi, Moreno Toigo, Migrants, employ et développement économique dans les provinces italiennes.
- ☐ **n. 298** Odo Barsotti, Federico Benassi, Linda Porciani, Moreno Toigo, Silvia Venturi, Trasmigrants, The Integration Process and Links with Country of Origin
- ☐ **n. 299** Riccardo Cambini
  Claudio Sodini, Global optimization of a generalized quadratic program
- ☐ **n. 300** Riccardo Cambini end Rossana Riccardi, Theoretical and algorithmic results for a class of hierarchical fleet mix problems

Anno: **2008**

- ☐ **n. 301** Riccardo Cambini and Claudio Sodini, A brannch and bound approach for a class of d.c. programs
- ☐ **n. 302** I.M. Stancu - Minasian and Andrea Madalina Stancu, SUFFICIENT OPTIMALITY CONDITIONS FOR NONLINEAR PROGRAMMING WITH MIXED CONSTRAINTS AND GENERALIZED p-LOCALLY ARCWISE
- ☐ **n. 303** Ray Chambers, Hukum Chandra and Nicola Salvati, Estimation of Proportions for Small Areas Using.Unit Level Models With Spatially Correlated population – An Application to Poverty Mapping.
- ☐ **n. 304** Andrea Mercatanti, Assessing the effect of debit cards on households' spending under the uncounfoundedness assumption
- ☐ **n. 305** Riccardo Cambini and Rossana Riccardi, On Discrete quasiconvexity concepts for single variable scalar functions
- ☐ **n. 306** Sara Biagini, Marco Frittelli, Matheus Grasselli, Indifference price with general semimartingales
- ☐ **n. 307** Sara Biagini, Paolo Guasoni, Relaxed Utility Maximization
- ☐ **n. 308** Monica Pratesi, Nonparametric Small Area Estimation via M-quantile Regression using Penalized Splines
- ☐ **n. 309** Angelo Antoci, Mauro Sodini, Indeterminacy, bifurcations and chaos in an overlaping generations model with negative environmental externalities
- ☐ **n. 310** A. Cambini L. Martein, On the maximal domains of pseudoconvexity of some classes of generalized fractional functions.
- ☐ **n. 311** A. Cambini L. Martein, On the generalized convexity of quadratic functions.
- ☐ **n. 312** Riccardo Cambini, Claudio Sodini, Global optimization of a generalized linear program.
- ☐ **n. 313** Cambini Alberto, Carosi Laura and Martein Laura, A new approach for regularity conditions in vector optimization
- ☐ **n. 314** Porciani Linda, Martin Pilar, La mediazione familiare: strumento di risoluzione dei conflitti

Anno: **2009**

- ☐ **n. 315** Federico Benassi, Linda Porciani, The dual profile of migration in Tuscany.
- ☐ **n. 316** Laura Carosi, Michele Gori, Antonio Villanacci, Endogenous Restricted Participation in General Financial Equilibrium-Existence Results
- ☐ **n. 317** Sara Biagini Mihai Sirbu, A note on investment opportunities when the credit line is infinite
- ☐ **n. 318** G. Giorgi, C. Zuccotti, Matrici a diagonale dominante:

principali definizioni, proprietà
e applicazioni

- ☐ **n. 319** Riccardo Cambini and Claudio Sodini, Global optimization of a generalized linear multiplicative program
- ☐ **n. 320** Riccardo Cambini end Francesca Salvi, Solving a class of low ranck d.c. programs via a branch and bound approach: a computational experience.
- ☐ **n. 321** Riccardo Cambini end Francesca Salvi, Solving a class of low ranck d.c. programs via a branch and reduce approach: a computational study.
- ☐ **n. 322** Riccardo Cambini end Francesca Salvi, A branch and reduce approach for solving a class of low ranck d.c. programs.
- ☐ **n. 323** Andrea Pallimi, On the asymptotic error of the bernstein-type approximations based on the beta-binominal distribution
- ☐ **n. 324** Sara Biagini - Ales Cerny, Admissible strategies in emimartingale portfolio selection
- ☐ **n. 325** Angelo Antoci, Ahmad Naimzada, Mauro Sodini, Strategic interaction and heterogeneity in a overlapping generation model with negative environmental externalities.
- ☐ **n. 326** Alessandra Coli, Francesca Scucces, La percezione della Solvay tra i residenti del Comune di Rosignano Marittimo: la progettazione di una indagine campionaria
- ☐ **n. 327** Bruno Cheli, Alessandra Coli, Barbara Burchi, Valutazione delle ricadute economiche della Solvay sul territorio della Val di Cecina

Anno: **2010**

- ☐ **n. 328** Ahmad Naimzada, Mauro Sodini, Multiple attractor and non linear dynamics in an Overlapping Generations Model with Environment.
- ☐ **n. 329** Data Envelopment Analysis with autputs uncertainty, Rossana Riccardi and Roberta Toninelli.
- ☐ **n. 330** Ahmad Naimzada, Mauro Sodini, Multiple attractor and non linear dynamics in an Overlapping Generations Model with Environment
- ☐ **n. 331** Ahmad Naimzada, Pierluigi Sacco, Mauro Sodini, Wealth-sensitive positional competition as a source of dynamic complexity in OLG models.
- ☐ **n. 332** Massimiliano Landi, Mauro Sodini, A dynamical Analysis of turnout with conformist citizens
- ☐ **n. 333** G. Oggioni, R. Riccardi, R. Toninelli, The cement industry: eco-efficiency country comparison using Data Envelopment Analysis
- ☐ **n. 334** Alessandra Coli, Francesca Tartamella, Income and consumption expenditure by households groupsin National accounts
- ☐ **n. 335** Alessandra Coli, Francesca Tartamella, Micro-macro integration: survey data on household income for theestimate of the Italian GDP
- ☐ **n. 336** Rossana Riccardi, Hierarchical Fleet Mix Problems with risk-aversion: a CVaR approach.
- ☐ **n. 337** Riccardo Cambini and Claudio Sodini, On solving a class of rank-two nonconvex programs by means of parametric uadratic semidefinite subproblems
- ☐ **n. 338** Massimiliano Landi, Mauro Sodini, Conformism and Turnout
- ☐ **n. 339** Alberto Cambini, Laura Martein, On the maximal domains of pseudoconvexity of a quadratic fractional function
- ☐ **n. 340** Stefano Marchetti, Claudia Dolci, Samantha Riccadonna and Cesare Furlanello, Bayesian Hierarchical Model for Small Area Disease Mapping: a Breast Cancer Study
- ☐ **n. 341** G. Oggioni R. Riccardiy R. Toninelliz, Eco-effciency of the world cement industry: A Data Envelopment
- ☐ **n. 342** Angelo Antoci, Fabio Sabatini, Mauro Sodini, See you on Facebook: the effect of social networking on human interaction

Anno: **2011**

- ☐ **n. 343** Stefano Marchetti, Nikos Tzavidis, Monica Pratesi, Non-parametric Bootstrap Mean Squared Error Estimation for M-quantile Estimators of Small Area Averages, Quantiles and Poverty Indicators
- ☐ **n. 344** G. Oggioni R. Riccardiy R. Toninelliz, Evaluating the e ciency of the cement sector in presence of undesirable output: a world based Data Envelopment Analysis
- ☐ **n. 345** Mauro Sodini, Local and global dynamics in an verlapping generations model with endogenous time discounting
- ☐ **n. 346** Daniela Ghio, Odo Barsotti, Silvia Venturi, Marc Termote, Lingua e cittadinanza: criteri alternativi per identificare una popolazione e le sue dinamiche?
  Un approccio demo-linguistico
- ☐ **n. 347** G. Giorgi and C. Zuccotti, Again on Regularity Conditions in Di erentiable Vector Optimization
- ☐ **n. 348** Annamaria Bianchi, Nicola Salvati, Asymptotic properties and variance estimators of the M-quantile regression coefficients estimators
- ☐ **n. 349** Alessandra Coli, Paolo Consolini, Marcello D'Orazio., Matching Eusilc and administrative data for studying poverty and social exclusion at the local level

Anno: **2012**

- ☐ **n. 350** Ricacardo Cambini e Laura Carosi, A note on semi-pseudoconvexity and "generalized" invexity
- ☐ **n. 351** Riccardo Cambini and Lura Carosi, A note on scalar "generalized" invexity
- ☐ **n. 352** Riccardo Cambini and Claudio Sodini, On the minimization of a class of generalized linear functions on a flow polytope
- ☐ **n. 353** Sara Biagini, Mustafa Pinar, Best gain-loss ratio is a poor performance measurement
- ☐ **n. 354** Giorgio Guzzetta, Marco Ajelli, Zenhua Yang, Stefano Merler, Cesare Furnalello, Denise Kirschner, Modelling socio-demography to capture tubercolosis transmission dynamics in a low burder setting.