Social Mobility and the Evolution of Tuberculosis

Elmer De La Pava - Salgado
Corporación Autónoma de Occidente
Cali, Colombia

Beatriz Salguero - Rivera
Corporación Autónoma de Occidente
Cali, Colombia

Juan Pablo Aparicio
Cornell University
Ithaca, NY

August 2001

Abstract

The explicit causes of the historic decline of mortality and morbidity rates of tuberculosis (TB) have not yet been clearly understood. Two different hypothesis have been proposed: a) The influence of public health programs against transmission of tuberculosis. b) The improvement of living standards, which decreased the likelihood of progression to active-TB; and Hypothesis (b) was already tested for the simple case of homogeneous population ([1]. Nevertheless, it is known that there exists a strong positive correlation between incidence of active-TB and poverty. In this work we introduce some degree of population heterogeneity. Population is divided in two classes: One is below the poverty level and the other one is above it is assumed that risk of progression to active-TB is greater in the population living below poverty level. United States data on poverty levels (measured by annual household income) [2] is used in order to approximate the time evolution of the size of the population living below poverty levels.

1

1 Introduction

Two centuries ago, TB was a major cause of death in most of now developed nations. However, mortality and morbidity rates begun a marked declining trend almost a century before the introduction of antibiotic treatment. In this work we explore some causes that could explain these historical trends. Tuberculosis (TB), is a disease caused by *Mycobacterium tuberculosis*. The bacteria can attack any part of the body, but they usually attack the lungs. TB is spread through the air (air-borne) from one person to another. In most people who become infected, the body is able to fight the bacteria to stop them from growing. The bacteria become inactive, but they remain alive in the body and can become active later. This is called TB latent infection. Most of the people who have latent TB infection never develop TB disease. But in other cases, especially people who have weak immune systems, the bacteria become active and cause TB disease. Malnutrition, alcoholism, drug abuse, concurrence of other infectious diseases, and even psychological stress may be a causes for a decrease in the immune response level [3].

Aparicio *et al.* [1] have shown that abrupt reductions in the risk of progression to active-TB together with the urbanization process may explain most of the the observed patterns. However they considered homogeneous populations which certainly is not the case. The high incidence of active-TB is strongly linked with some degree of social deprivation. The incidence of active TB rates for New York City as a whole is very low at present time, but rise to almost 100 in some poor neighborhoods. For instance, in 1990, the incidence tuberculosis in Central Harlem was the highest in New York City and was 32 times that of the more affluent, neighboring Kips Bay-Yorkville sections of Manhattan [6, 7].

Therefore, a better understanding of the evolution of TB patterns should be

achieved by incorporating population heterogeneities. In this work we consider the most simple case in which population is divided in two classes. One class is composed by the subpopulation living in some degree of poverty, while the other subpopulation is composed by the rest of the population. Our main assumption is that the risk of progression to active-TB of infected individuals is higher for people living in the first subpopulation.

Although there is a known link between risk of TB disease and social deprivation indicators it is not clear how to characterize such high risk populations, as well as its time evolution. Definition of poverty level is rather arbitrary and different authors used different definitions [10].

> "Poverty in America is of a far different order from poverty in most of the rest of the world and from the kind of poverty that most history has recorded. In arriving at a concept of poverty and in prescribing solutions, it should be recognized that poverty in the United States involves specific people, families, and groups and is not a mass phenomenon. There are two types of poverty. One identifies the poor as those falling within certain income levels at the bottom of the national income scale. The second type measures the poor as those living below some minimun - decency standart of living".

> *Chamber of Commerce of the United States.*

A significant declining trend was observed for the U.S. as a whole. That is there is a strong empirical evidence that the proportion of people living in risky populations have been declined through last century. In this work we look at the impact of such shift in the long-term dynamics of tuberculosis.

This work is organized as follow. In Section 2 we consider a two group a TB model with time-independent parameters and basic reproductive number is computed. In section 3 we considerer a more general model with time-dependent parameters which incorporates social mobility and population growth, section 4 are presented the methods and section 5 are presented results. The discussion and conclusions are presented in section 6.

2 A Simple Two-group TB Model

Here we analyze a two group TB model with constant parameters which is a modification of the model presented in [1]. Total population is subdivided in two subpopulations with different per-capita risk of progression to active-TB. Sub-population with higher risk of progression to active-TB will be called *high risk* population (with size N_1), while the other will be denoted as *low risk* population (of size N_2).

Individuals in each subpopulation may belong to one four epidemiologic classes: Susceptible, $S_i(t)$; individuals in a (high risk) latent class, $E_i(t)$; infectious individuals, $I_i(t)$; individuals in a (low risk) latent class $L_i(t)$, and $N_i = S_i + E_i + L_i + I_i$ for $i = 1, 2$.

In each subpopulation individuals are recruited at the per-capita rate Λ_i and die at the per capita rate μ. Individuals in the latent class E_i progress to active-TB (class I_i) at the per - capita rate k_i. Individuals who do not progress to active-TB are moved from the class E_i to a permanent low risk latent class L_i at the per-capita rate α. Recovery individuals are aggregated into the L_i class.

We assume that most of the contacts of an individual are within individuals in the same subpopulation. An infectious individual in the i subpopulation may produce (in an entirely susceptible population) Q_0^{ii} secondary infections in his/her subpopulation and Q_0^{ji} secondary infections in the other subpopulation. When only some of the contacts are already susceptible those numbers are reduced by the corresponding

4

susceptible proportions. We assume that mean infectious period is the same for both subpopulations. Then our model becomes:

$$\frac{dS_1}{dt} = \Lambda_1 - \gamma Q_0^{11} \frac{S_1}{N_1} I_1 - \gamma Q_0^{12} \frac{S_1}{N_1} I_2 - \mu S_1 \tag{1}$$

$$\frac{dE_1}{dt} = \gamma Q_0^{11} \frac{S_1}{N_1} I_1 + \gamma Q_0^{12} \frac{S_1}{N_1} I_2 - \mu E_1, \tag{2}$$

$$\frac{dI_1}{dt} = k_1 E_1 - \gamma I_1, \tag{3}$$

$$\frac{dL_1}{dt} = r_1 I_1 + \alpha_1 E_1 - \mu L_1, \tag{4}$$

$$\frac{dS_2}{dt} = \Lambda_2 - \gamma Q_0^{22} \frac{S_2}{N_2} I_2 - \gamma Q_0^{21} \frac{S_2}{N_2} I_1 - \mu S_2, \tag{5}$$

$$\frac{dE_2}{dt} = \gamma Q_0^{22} \frac{S_2}{N_2} I_2 + \gamma Q_0^{21} \frac{S_2}{N_2} I_1 - \mu E_2, \tag{6}$$

$$\frac{dI_2}{dt} = k_2 E_2 - \gamma I_2, \tag{7}$$

$$\frac{dL_2}{dt} = r_2 I_2 + \alpha_2 E_2 - \mu L_2, \tag{8}$$

where $1/\gamma$ is the mean infectious period $(\gamma = r + \mu)$, $N_1 = S_1 + E_1 + I_1 + L_1$ and $N_2 = S_2 + E_2 + I_2 + L_2$ are the sizes of the subpopulations.

The basic reproductive number indicates whether a disease may invade a population. The basic reproductive number of the system (1)-(8), defined as the spectral radius of the next generation operator [4], (see also [5]) is given by

$$R_0 = \frac{f_1 Q_o^{11} + f_2 Q_o^{22} + \sqrt{(f_1 Q_o^{11} - f_2 Q_o^{22})^2 + 4 f_1 Q_o^{12} f_2 Q_o^{21}}}{2}, \tag{9}$$

where, $f_1 = \frac{1}{k_1+\alpha_1+\mu}$, and $f_2 = \frac{1}{k_2+\alpha_2+\mu}$ are the fractions of infected individuals who progress to active TB in both classes respectively. Expression (9) reduced to expression $R_0 = Q_0 f$ in [1] for homogeneous population ($Q_o^{11} = Q_0^{22} = Q_0, f_1 = f_2 = f$). By definition, one infectious individual from class i placed in a susceptible population produces Q_0^{ii} secondary infections.

Then one infectious in class i produces:

$$R_0^{ij} + R_0^{ji} = R_0^i$$

secondary cases, where $R_0^{ij} = Q_o^{ij} f_j$. Regardless of the size of the N_1 population, we have that if $R_0^{ii} > 1$ then $R_0 > 1$, that is TB may survive at population level taking advantage of small risky pockets.

3 Adding Social Mobility and Population Growth

In this section we consider a more general model with time-dependent parameters which incorporates social mobility and population growth.

Individuals are recruited into the susceptible (and uninfected) classes S_i, at the per-capita rates $\Lambda_i(t)$, $i = 1, 2$ and die at the per-capita $\mu(t)$. All newborns are susceptible and we also assume that an individual may become infected only through contact with infectious individuals.

Individuals in the high risk population N_1 move to the low-risk population N_2 at the per-capita rate $p(t)$.

Then Model (1)-(8) becomes:

$$\frac{dS_1}{dt} = \Lambda_1(t) - \gamma_1 Q_0^{11} \frac{S_1}{N_1} I_1 - \gamma_1 Q_0^{12} \frac{S_1}{N_1} I_2 - \mu(t)S_1 - p(t)S_1, \qquad (10)$$

$$\frac{dE_1}{dt} = \gamma_1 Q_0^{11} \frac{S_1}{N_1} I_1 + \gamma_1 Q_0^{12} \frac{S_1}{N_1} I_2 - \mu(t)E_1 - p(t)E_1, \qquad (11)$$

$$\frac{dI_1}{dt} = k_1(t)E_1 - \gamma_1 I_1 - p(t)I_1, \tag{12}$$

$$\frac{dL_1}{dt} = r(t)I_1 + \alpha_1 E_1 - \mu(t)L_1 - p(t)L_1, \tag{13}$$

$$\frac{dS_2}{dt} = \Lambda_2(t) - \gamma_2 Q_0^{22}\frac{S_2}{N_2}I_2 - \gamma_2 Q_0^{21}\frac{S_2}{N_2}I_1 - \mu(t)S_2 + p(t)S_1, \tag{14}$$

$$\frac{dE_2}{dt} = \gamma_2 Q_0^{22}\frac{S_2}{N_2}I_2 + \gamma_2 Q_0^{21}\frac{S_2}{N_2}I_1 - \mu(t)E_2 + p(t)E_1, \tag{15}$$

$$\frac{dI_2}{dt} = k_2(t)E_2 - \gamma_2 I_2 + p(t)I_1, \tag{16}$$

$$\frac{dL_2}{dt} = r_2 I_2 + \alpha_2 E_2 - \mu L_2 + p(t)L_1, \tag{17}$$

where

$$N_1 = S_1 + E_1 + I_1 + L_1, \tag{18}$$

and

$$N_2 = S_2 + E_2 + I_2 + L_2, \tag{19}$$

4   Methods

Records on population size, life expectancy at birth, general mortality, and mortality by tuberculosis in United States are available from 1850 [9]. Data on the incidence of active TB (new cases per year) are available from 1915 [10]. Also we use data on incidence of poverty (from 1900)[2, 9, 11] to estimate possible time-evolution of the high risk population.

4.1 Estimation of Demographic Parameters

As in [1] we consider the total U.S. urban population (N). Per capita mortality rate is estimated as the inverse of life-expectancy at birth (see Appendix). The number of new susceptible individuals recruited (by birth or immigration) in a short interval δt is estimated as

$$\Lambda(t)\delta t = N(t + \delta t) - N(t) + \mu N \delta t.$$

The values of the U.S. urban population $N(t)$ are estimated from census data using linear interpolation, and the observed time-evolution of the proportion of urban population [1]. Here we assume that recruitment in each subpopulation are proportional to their respective sizes, that is

$$\Lambda_1(t) = \Lambda(t)\frac{N_1}{N}$$

and

$$\Lambda_2(t) = \Lambda(t)\frac{N_2}{N}.$$

Size and time-evolution of the high risk population N_1 is unknowm. At least in the United States there is a strong correlation between incidence of active-TB and median income,[6]. Also comunity unemployment level and incidence of active-TB rates positevely correlates [6]. Therefore we used data on incidence of poverty as a sumgate of the population at high risk.

In figure 1 we show date on incidence of poverty (measured by income) together with three possibles time-evolutions of the high risk population proportion

$$h \equiv \frac{N_1}{N}$$

In all cases we used the same family of parametric models

$$h(t) = h_f + \frac{h_i - h_f}{1 + \exp[(t_i - t_{h1/2})/\Delta_h]} \tag{20}$$

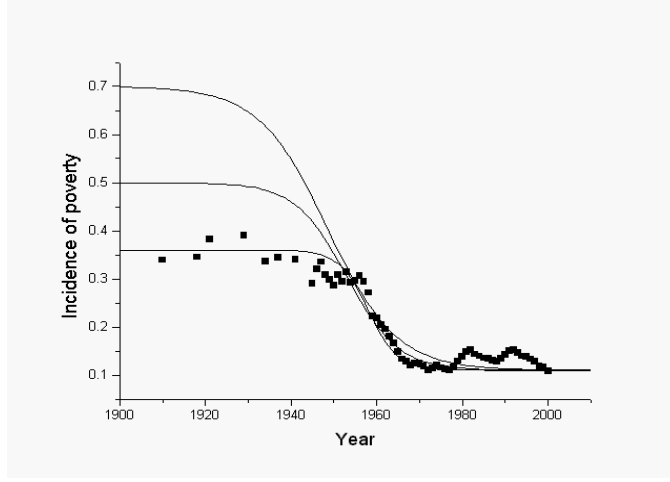and in each case the values of the parameters are listed in table 1.

8

Figure 1: The three cases considered for the time evolution of the high risk population fraction $h(t)$, and the data (o) are built with the data given by the census and the continuous curve is adjusted for the Model with Social Mobility and Population Growth

$t_{1/2}$	h_i	h_f	Δ_h
1945	0.7	0.11	10
1948	0.6	0.11	10
1953	0.5	0.11	8
1957	0.4	0.11	8

Table 1: Values used in the simulations estimate the incidences shown in figure 2

Using

$$\frac{dN_1}{dt} = \Lambda_1(t) - \mu(t)N_1(t) - p(t)N_1(t),$$

$$\frac{dN_2}{dt} = \Lambda_2(t) - \mu(t)N_2(t) + p(t)N_1(t),$$

and

$$\frac{dN}{dt} = (\Lambda_1(t) + \Lambda_2(t)) - \mu(t)N(t)$$

together with $N(t) = h(t)N_1(t)$ we obtain

$$p(t) = \frac{\Lambda_1(t) - \mu(t)N_1 - (dh/dt)N(t) + h(t)(\Lambda_1(t) + \Lambda_2(t) - \mu(t)N(t))}{h(t)N(t)}.$$

4.2 Epidemiological parameters

Trough this work we have considered constant values of Q_0^{ij}. Moreover we assumed that the potential per-infectiuos transmission rate is the same in both populations, that is

$$Q_0^{ij} = Q_0^{ji}, \quad \text{for all } i, j.$$

Also we assumed that most of the contacts of an individual takes place with members of his/her own subpopulation. In our simulations we set Q_0^{ii}=10 and $Q_0^{ij} = 1$

We assume that risks of progression to active-TB vary with time as in [1]

$$f(t) = f_f + \frac{f_i - f_f}{1 + \exp[(t_i - t_{1/2})/\Delta]} \tag{21}$$

and that $f_2 = \rho f_1$ with $\rho < 1$ and $f_1 = f$ (see appendix). In the simulations we considered $\rho = 0.75$ and $\rho = 0.5$.

4.3 Simulating Incidence of active-TB and Prevalence of infection

Incidence of active-TB (new active cases per year) in each subpopulation is estimated as $k_i(t)E_i(t)$, (setting one year as unite of time) because k_i and E_i can be considered constant during the short period of one year. Incidence of active-TB (rate per 100.000 individuals) is obtained as

$$\frac{k_1(t)E_1(t)100.000}{N_{tot}(t)} + \frac{k_2(t)E_2(t)100.000}{N_{tot}(t)}$$

where $N_{tot}(t)$ denotes total U.S. population at calendar year t which is estimated as $N_{tot} = P_U(t)^{-1}N(t)$ ([1], see also appendix). The first term of the sum represent the contribution to the high risk population., while the second term is the contribution corresponding to the low-risk population.

On the other hand, incidences of active-TB in the N_i populations are given by

$$\frac{k_i(t)E_i(t)100.000}{N_i(t)}.$$

Prevalence of infection is estimated as

$$\frac{E_1(t) + L_1(t)}{N_{tot}(t)} + \frac{E_2(t) + L_2(t)}{N_{tot}(t)}$$

where each term represent the contribution of each subpopulation, while prevalence of infection in each subpopulation is given by

$$\frac{E_i(t) + L_i(t)}{N_i(t)}.$$

5   Results

For the different cases considered for the time evolution of the high risk population proportion $h = \frac{N_1}{N}$. We found values f_i and f_f in (21) which allow for a good fit of model solutions to data (see figure 3).

Under the conservative assumptions $Q_0^{ij} = Q_0^{ji}$ and $f_2 = 0.75f_1$ we obtained that the actual contribution of the small high-risk population.

To the total incidence of active-TB is about 6 per 100.000. *That is 10% of the population produces 90% of all cases.* Furthermore the incidence in such subpopulation reach as high levels that those observed in many developing countries.

Despite the fact that differences in per-capita TB progression rates is small, population heterogeneity significantly reduce the ratio $\frac{f_i}{f_f}$, of the asymptotic values values of f needed to achieve a good match of the data, when compared with the value obtained using a homogeneous population ($\frac{f_i}{f_f} = 2$ in this work and $\frac{f_i}{f_f} = 3.5$ in the homogeneous population case [1]. Results for the three posible time-evolution of the fraction $h(t)$ considered are almost identical (see Table 1 and figure 1).

Today predicted prevalence of latent infection is about 3% for the general U.S. population (see figure 6). This result closely agree with the 4%-6% recent estimates [7], and improve the prediction of above 11% obtained with one group model. However, prevalence of latent infection in high risk subpopulation N_1 is above 10%.
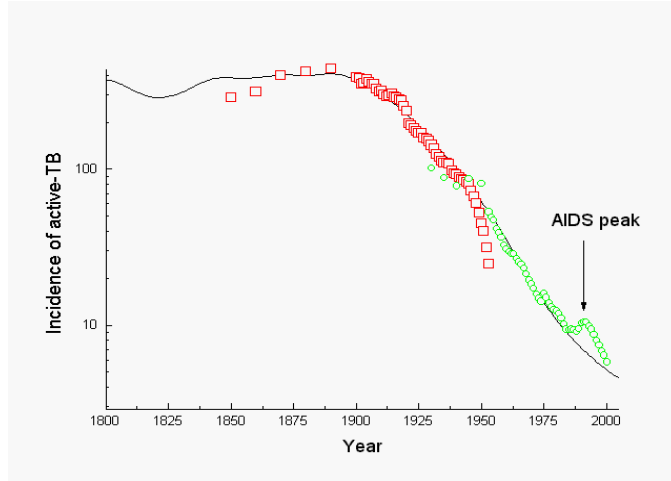
11

Figure 2: Simulated incidence of active-TB (rate per 100,000 individuals). Model solutions almost coincides for every of the cases considered for the time evolution of $h(t)$.
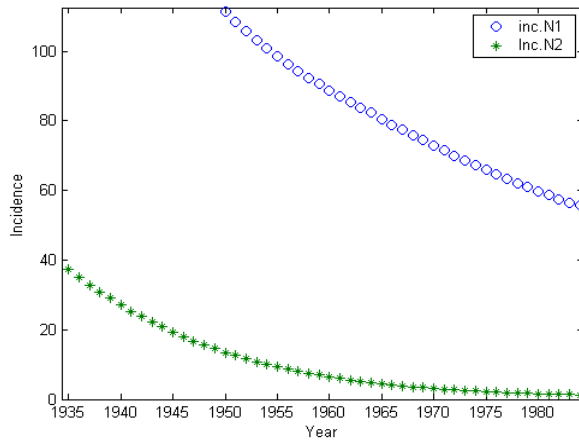

Figure 3: United States incidence of active-TB rates per 100.000 individuals obtained from model (10-19) the point o denote incidence of active-TB in subpopulation N_1. The points in $*$ describes the incidence to total population.
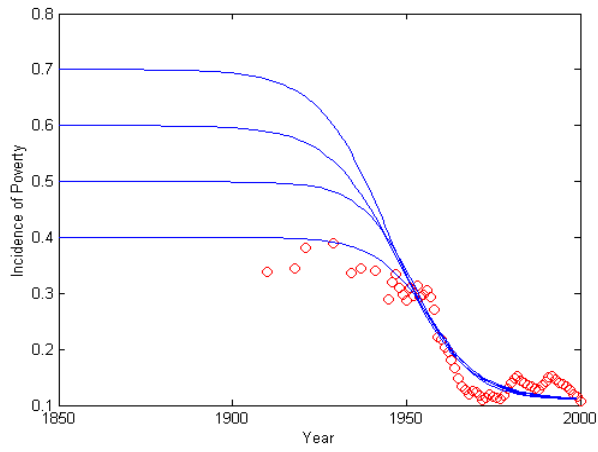
Figure 4: Adjustment of incidence of poverty for different values of $h(t)$. The points o are built with data given by the Census and the continous curve is adjusted for the Model with Social Mobility and Population Growth.
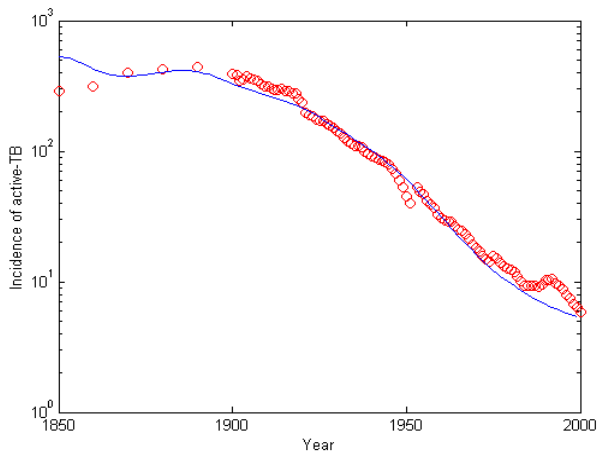

Figure 5: Incidence of active-TB for values $h = 0.5$ $h_f = 0.11$, $f_i = 0.32$, $f_f = 0.115$, $\Delta_h = 8$, $t_{\frac{1}{2}=1953}$

Figure 7 show the prevalence of a latent infection trend between years of 1800 and 1999
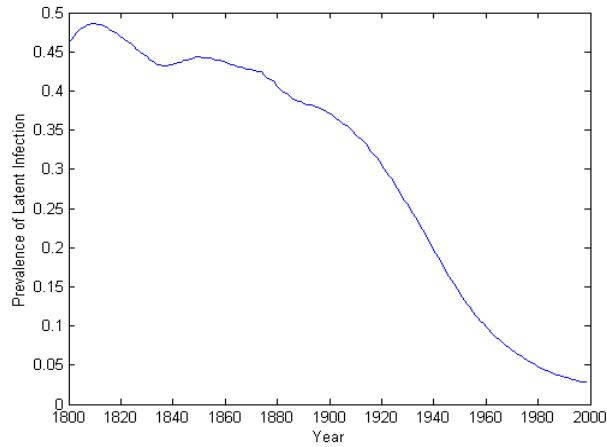

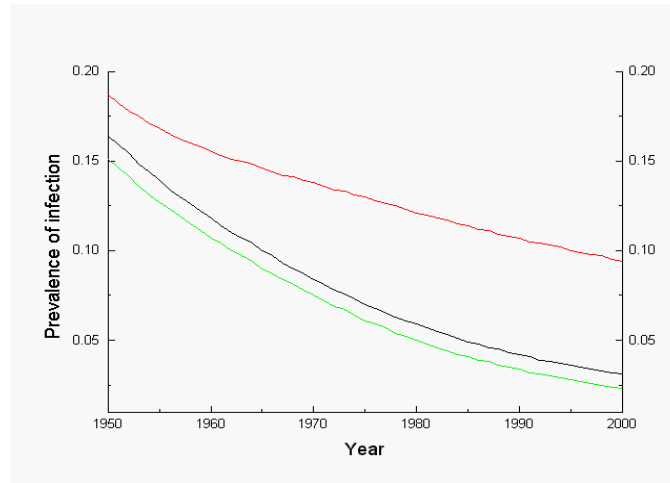Figure 6: The United States prevalence of active-TB (rates by 100.000 obtained from Model (10)-(19)).

Figure 7: Predicted prevalence of infection levels. Today values are about 3% but prevalence in high risk population rise to above 10%.

6   Discussion and conclusions

Many factors determine which populations are at risk of TB infection and TB disease, but most of them are associated with poor standard of living. It is clear that available data may only provide some rough estimate of the size and time evolution

14

	h_i	h_f	Δ_h	$t_{h1/2}$
Case A	0.36	0.11	3.42	1958
Case B	0.5	0.11	6	1953
Case C	0.7	0.11	8	1948

Table 2: Values used in the simulations estimate the incidences shown in figure 2.

of this population. In this work we used data on incidence of poverty as such indicator. Poverty may measured in several ways and defining poor people as those living under some standard of living is the most appropriate definition to our purposes but unfortunately is not available before 1959. However a significant positive correlation between income level and standard of living is expected. Data show a well defined declining trend of poverty incidence in United States during the last century. This fact strongly support our view that the proportion at risk of TB disease of the total population was declining during last century. Incidence of poverty, at the turn of the Twentieth century, was over 40%, but it is likely that many more people was already at significant risk of TB-disease. Incidence of poverty fell down to around 10% at present time. Our results shows that this relatively small population is what sustain tuberculosis in the whole population. Today predicted incidence of active-TB in population at high risk is above 30 per 100000 population but fall to a 6 per 100000 when considering the whole population. Most of the U. S. population is almost free of TB, but in the small disadvantaged population living in poverty TB incidence reach as high levels like in most developing countries.

At first approximation we have considered the simplest case in where individuals may belong to only two sub-populations. We assumed that individuals in both populations are identical except in the risk of progression to active-TB following infection. Estimations on the value of the between risks ratio f_1/f_2 could be obtained by evaluating the ratio prevalence-of-infection to incidence-of-active-TB in different

sub-populations. In this work we have considered the case $f_2 = 0.75f_1$. As in [1] we have considered that, besides the incidence of poverty, risks of progression declined with time. Fitting model solutions to data requires to select two asymptotic values (f_i and f_f in (21)) of the risk $f(t)$. When population heterogeneity is considered the difference between these asymptotic values is significantly smaller than the values obtained for homogeneous populations. For example we have obtained good fit to the data using $f_i = 0.22$ and $f_f = 0.1075$ for $Q_0 = 11$ while for the homogenous case these values resulted 0.34 and 0.1075 respectively[1]. A two fold decrease is necessary to explain the observed TB trends while homogenous population model required more than a three fold decrease in the risk of progression to achieve similar results.

Predicted prevalence of infection in total population is about 4% well within the estimated range 4%-6%, while homogenous population model predict a value of about 11% (JPA personal communication).

7   Acknowledgements

References

[1] Aparicio, J.P.,A.Capurro and C. Castillo- Chávez (2001) Markers of disease evolution:the case of tuberculosis . Submitted.

[2] U.S.Bureau of the Census(1959-2000). Statistical Abstracts of the United States. See, also http://www.census.gov/population.

[3] Aparicio, J.P., A. Capurro and C. Castillo- Chávez (2001). The mathematical modeling of tuberculosis epidemics .em Submitted.

[4] Diekmann O, Heesterbeek JAP and Metz JAJ. On the definition and the computation of the basic reproduction ratio $\Re_0$ in models for infectious diseases in heterogeneous populations. Journal of Mathematical Biology 28:365–382 (1990).

[5] C. Castillo-Chavez, Z. Feng and W. Huang, On the computation of $\Re_0$ and role on global stability. Mathematical Approaches for emerging and reemerging infectious diseases, PartI: An Introduction to Models, Methods and Theory. C.Castillo-chavez, S. Blower, D. Krischner, and A-A Yakubu IMA Volume Springer Verlag, Forthcoming 125:229–250.

[6] Felton CP, Ford JG. Tuberculosis in the Inner City. In: Reichman LB, Hershfield ES (eds.) Tuberculosis. A comprehensive International approcah. Marcel Dekker Inc, New York, NY, 483-504.

[7] Salpeter EE, Sh R. Salpeter.1998. Mathematical Model for the Epidemiology of Tuberculosis, with Estimates of the Reproductive Number and Infection - Delay Function. American Journal of Epidemiology .

[8] CDC (1999). Reported Tuberculosis in the United States. 1999, order $\#(99 - 6583)$. Available from http://www.cdc.gov/nchstp/tb/sur/surv99/surv99.htm

[9] U.S.Bureau of the Census(1975). Statistical Abstracts of the United States. See, also http://www.census.gov/population.

[10] Task Force on Economic Growth and Opportunity. 1965. The concept of poverty. Chamber of Commerce of the United States.

[11] . Chevalier JM, (1971). La Pauvrete Aux Etats - Unis . Presses Universitaries de France. Pag 113.

[12] Gordon M. Fisher. Octubre 10, 1999. How many Americans Were Really in Poverty in 1947? Estimates of the U.S. Poverty Population between 1947 and 1963 Under Two Contemporary (1949 and 1959) Definitions of Poverty.

Appendix

Since Aparicio and et al. in [1] defined life expectancy at birth , we used the following flexible parametric model:

$$\tau(t) = \tau_f + \frac{\tau_i - \tau_f}{1 + \exp[(t_i - t_{1/2})/\Delta]} \tag{22}$$

The parameters τ_i and τ_f model asymptotic values; $t_{1/2}$ denotes the time at which life-expectancy at birth reaches its half value, $\tau(t) = (\tau_i + \tau_f)/2$; and Δ is the shape parameter which determines the width of the sigmoid shape function. Parameter estimates are obtained from historical data Bureau Census [2]. A usual measure of the risk of progression to active-TB is given by the fraction $f = \frac{k}{k+\alpha+\mu}$ which roughly estimates the proportion of infected people who develop active-TB during their life-spans. The rate of progression to active-TB is obtained from f when $k = (\alpha + \mu)f/(1 - f)$. The value of α, which controls the average time spent in the high risk latent class $E_i, i = 1, 2$, was set to be equal to $2/3yr^{-1}$ [1]. The time-evolution

of f is fitted to the same family of parametric models (22), that is, to

$$f(t) = f_f + \frac{f_i - f_f}{1 + \exp[(t_i - t_{\frac{1}{2}})/\Delta]} \tag{23}$$

The parameter values of $t_{1/2}$ and Δ that determine the timing and abruptness of the transition of f between its asymptotic values, are those obtained from the best fit of (22) to the life-expectancy data. Fixing $t_{1/2}$ and Δ leaves only two free parameters in Model (23), the asymptotic values f_i and f_f [1].