MODELING TUBERCULOSIS IN AREAS OF HIGH HIV PREVALENCE

Georgina R. Hughes
Christine S.M. Currie
School of Mathematics
University of Southampton
Highfield, Southampton, SO17 1BJ, UK

Elizabeth L. Corbett
London School of Hygiene and Tropical Medicine
Biomedical Research and Training Institute (at NIHR)
Josiah Tongogara Avenue (Cnr Mazoes) Avenues
Harare, PO Box CY1753, ZIMBABWE

ABSTRACT

We describe a discrete event simulation model of tuberculosis (TB) and HIV disease, parameterized to describe the dual epidemics in Harare, Zimbabwe. TB and HIV are the leading causes of death from infectious disease among adults worldwide and the number of TB cases has risen significantly since the start of the HIV epidemic, particularly in Sub-Saharan Africa, where the HIV epidemic is most severe. There is a need to devise new strategies for TB control in countries with a high prevalence of HIV. This model has been designed to investigate strategies for reducing TB transmission by more efficient TB case detection. The model structure and its validation are discussed.

1 INTRODUCTION

Tuberculosis (TB) and HIV are the leading causes of death from infectious diseases among adults (Corbett et al. 2003; World Health Organization 1999). Estimates for 2003 put the number of incident TB cases at 8.8 million, up from an estimated 8.3 million in 2000, with HIV being the main driving force (World Health Organization 2003). Adult HIV prevalence rates are now 20% or higher in six Southern African countries. In these countries TB case notification rates have increased 2 to 5 fold since 1990 and now between 460 and 720 people develop active TB disease per 100,000 of population per year (World Health Organization 2003).

The international standard for TB control is the World Health Organization’s DOTS strategy (World Health Organization 2003), which aims to reduce the transmission of M. tuberculosis infection through prompt diagnosis and effective treatment of symptomatic TB patients who present at health care facilities, termed passive case-finding. Considerable progress was made during the last decade using this strategy in countries with small HIV epidemics but the effect of HIV on the African TB epidemic outweighs the gains being made in other regions.

In this paper we describe a discrete event simulation model that is being developed to evaluate the effects of more intensive case-finding strategies (so-called active case-finding) for TB control in a high HIV prevalence setting. In essence, active case-finding involves targeted testing of the population for active disease with one commonly used strategy being to target household members of TB patients. Those found to have active disease can be treated promptly, reducing the time spent with infectious TB and so cutting transmission rates. The current policy is that active case-finding for adults living in endemic TB settings is ineffective, because transmission events between casual contacts greatly outnumber household transmission events (Rieder 2003). This policy was developed in an era of low HIV prevalence and the impact of the HIV epidemic on the relative importance of household versus community transmission has not been fully assessed. The HIV epidemic will have some effect on transmission because the duration of infectiousness of HIV-related TB is much briefer than in HIV-negative TB patients (Corbett et al. 2004; Corbett et al. 2005).

The end goal of the modeling is a geospatial, discrete event simulation (DES) model of TB transmission in the district, which will allow a full assessment of the effectiveness of contact-tracing and case-finding strategies that make use of geographical information. The first stage, which we describe in this paper, is a DES model that simply differentiates between within-household transmission and random transmission of TB. Differentiating between these two modes of transmission is important as household transmission of M. tuberculosis has long been recognized as having very different dynamics from that in the wider community. Household exposure to infectious TB carries a high risk of infection (25 to 50%) (Rieder 2003) while random transmission occurs at a lower background rate of approximately 0.5 to 1% per year (Dye et al. 1999).

The model will be fitted to data from a large population-based trial in Harare, Zimbabwe, DETECTB, which will become available during 2006. Data is being collected on the size and location of every household in the study area, as well as the number of inhabitants, their ages and their TB and HIV status. When the baseline data are avail-

1-4244-0501-7/06/$20.00 ©2006 IEEE
able, we will use them to inform the model structure and parameters.

In the work presented here, the model has been parameterized so that it corresponds with earlier deterministic models of TB and HIV (Vynnycky and Fine 1997; Currie et al. 2003). Validation of its output has been achieved by comparison with country-wide statistics for Zimbabwe (World Health Organization 2006; HIV data). In the future, the output will be validated using the results of follow up surveys that are to be carried out in the same study area every 6 months until the end of 2008.

We discuss earlier models of TB and HIV in section 2 and also give some of the necessary background biology to tuberculosis. In Section 3, we describe the model in some detail, before going on to describe the validation process in Section 4. We conclude and discuss our plans for future work in Section 5.

2 BACKGROUND

2.1 Literature

The majority of previous studies have used deterministic compartment models (Blower 1996, 1998, Murray and Salomon 1998a, 1998b, Dye et al. 1998, Currie et al. 2003, 2005, Williams et al. 2005). Deterministic compartmental models are when the population is divided into different epidemiological states and the movements between the states are represented by a system of differential equations. These previous compartmental models have focused on modeling TB at the population level. The studies concentrated on the effect of interventions at a large scale and although many of them addressed the implications of reducing transmission, none of them were able to look at the actual mechanics behind it. These studies have been vital in understanding and quantifying TB disease progression in populations and will be used to inform our model. However it is felt that a discrete event simulation is more appropriate for investigating interventions at the household level and enables the more intricate details of transmission to be understood.

The importance of household versus random transmission has previously been studied using a deterministic compartment model developed to look at the importance of cluster size in TB disease dynamics (Aparicio et al. 2000, Song et al. 2002). Whilst this analysis was useful in understanding the impact of the evolution of TB virulence on TB dynamics, some aspects of the modeling were unrealistic and, critically, the population structure was ignored. The study also did not incorporate the effect of HIV.

Murray (Murray 2002) is the only example that we are aware of where discrete event simulation has been used to model TB. She used her model to investigate the clustering of different strains of TB. The use of this technique was successful and modeling TB at the individual level proved to be a powerful tool.

2.2 Tuberculosis

A person infected with Mycobacterium tuberculosis (mTB) bacilli may have a latent infection or active TB disease. If the infection is latent, the mTB bacilli are present in the body but that they are not active because the immune system has successfully “walled them off”. A person can retain a latent infection for many years and the infection often only develops into active TB disease when the immune system is weakened. The most common form of TB disease is pulmonary TB, which affects the lungs. Other forms of TB disease are referred to as extra-pulmonary disease and could affect the central nervous system, bones, joints and the lymphatic system. TB skin sores can develop when an infected lymph gland bursts, but this is very uncommon. TB is only infectious if the disease is in the lungs or if a TB skin sore is left uncovered. In this study, we concentrate on pulmonary TB as a means of TB transmission.

3 THE MODEL

The TB model is a discrete event simulation model which allows the population to belong to six epidemiological classes, dependent on TB status. Movements through the pathways of the model are determined by an individual’s attributes (age, gender, HIV status), with the distributions used to describe progression currently based on the medical literature. As data from the Harare survey become available, some of these distributions will be changed to reflect the survey results.

Transmission is currently modeled under the assumption that the majority of transmissions will occur outside of the household with only one household transmission occurring within the household per infectious person, per year (Rieder 2003). The numbers of each type of transmission event caused by an infectious person per year are therefore generated from Poisson distributions with means of 9 and 1 respectively. When a transmission event is generated, an individual will be selected at random to become infected. If transmission is within household, the individual to be infected will be selected from within the household, and if the transmission is outside the household, any member of the model population could become infected. Persons who have previously recovered from TB; or already have a latent infection can be “reinfected” with mTB bacilli.

The model uses a schedule of events that are executed in chronological order. Any new events that are created are inserted into the correct position on the schedule, which is held as a linked list. For example, when a person is born, their time of natural death is generated and this “Natural Death” event is scheduled onto the event list. If the indi-
Hughes, Currie and Corbett

Individual develops active TB disease and does not receive treatment before dying from TB, a new event, “Death from TB” will be scheduled at the appropriate time. If death from TB is scheduled to occur before their “Natural Death” event, the “Natural Death” event will be removed from the event list.

3.1 Modeling the Natural History of TB

The schematic for the model is given in Figure 1. Death can occur from any state, and death rates are higher for individuals with active disease. The structure is similar to that used for deterministic compartmental models in earlier studies (Debanne et al. 2000, Vynnycky 1996, Murray and Salomon 1998, Williams 2005 and Currie et al. 2005). We assume that the TB disease progression parameters for HIV-positive individuals only change when the individuals enter late-stage HIV, defined to be World Health Organization stages 3 and above, approximately 6 years after infection (Morgan et al. 2002, World Health Organization 1990).

Susceptible individuals are not infected with mTB. When they become infected, they enter the latent infection class. Individuals will then follow one of two routes: develop active disease within 5 years, termed fast progression to active disease; or retain a latent mTB infection, progressing to active disease at a rate of 0.001 per year for HIV negatives (Sutherland 1947, Horwitz et al. 1969, Barnett et al. 1971, Styblo 1991, Vynnycky 1996, Vynnycky and Fine 1997) and 0.1 per year for late-stage HIV-positives (Schulzer et al. 1992, Williams et al. 2005). Approximately 14% of HIV-negatives will exhibit fast progression (Sutherland 1968, Ferebee 1970, Comstock 1982, Sutherland et al. 1947, Styblo 1986, Krishnamurthy et al. 1976, 1990, Vynnycky and Fine 1997) and 67% of individuals in late-stage HIV (Di Perri et al. 1989, Daley et al. 1992, Edlin et al. 1992, Coronado et al. 1993).

Previous infection with mTB confers some immunity to developing active disease following reinfection with mTB. Only 4.9% of HIV-negatives and 50% of late-stage HIV-positives (World Health Organization stages 3 and above) who are reinfected with mTB will follow fast progression to active disease (Currie et al. 2005).

Active TB disease can be infectious or non-infectious, with 46% of individuals (30% of late-stage HIV-positives) developing infectious disease (Williams et al. 2005). Non infectious individuals are able to convert to infectious TB at rate 0.015 per year (Williams et al. 2005). When an individual develops active disease they will either be treated, and move to the “Treated” class, or they will die. The average time until treatment is 2 years (Currie et al. 2005) and the average time until death from TB is sampled from a Weibull distribution with mean of 3.3 years for infectious disease and 5 years for non infectious disease (Williams et al. 2005). For those with infectious disease, the time until death or treatment determines an individual’s duration of infectiousness, which in turn determines how many people the individual is likely to infect. We assume that individuals will transmit to an average of 10 other people per year (Currie et al. 2005). Children under fifteen rarely develop active TB disease but will retain a latent infection if infected with mTB bacilli.

Figure 1: Schematic of the discrete event simulation tuberculosis (TB) model. Death may occur from any state, but death rates are higher from active disease states.

The duration of treatment is assumed to be 6 months, at the end of which the individual will either fail or successfully complete treatment. Those that successfully complete the treatment course become “recovered” and those that fail will return to active disease. When a person has recovered they are again susceptible to reinfection from an infectious person, although they have an increased immunity compared to the “susceptible” population.

3.2 Model Initialization

The simulation generates a population of 10,000 people using the age structure and life expectancy data from the WHO life tables. The individuals are assigned to households with an average household size of 5.5 (Action Aid International 2005). The initial distribution of the population amongst the epidemiological classes is obtained from the steady state output of a deterministic compartmental
model with the same structure as the DES model but no HIV epidemic. The compartmental model was developed using Berkeley Madonna, <http://www.berkeleymadonna.com>. A warm up of approximately 300 years is then used before HIV is introduced.

3.3 Modeling HIV

We do not model HIV transmission explicitly but instead use a static model of HIV to generate a number of HIV infections in each time step. Reasonably good data are available for the proportion of the population infected with HIV, termed the HIV prevalence, over time in the countries of interest. This is generally measured by HIV-testing women attending antenatal clinics. The future course of the epidemic is much less certain, however, making it desirable to use a flexible model to describe the HIV epidemic, giving us the opportunity to consider scenarios in which HIV prevalence increases, decreases or remains constant in the future.

The model we use was first introduced by Salomon and Murray (Salomon and Murray 2001) and acts to derive estimates of the HIV prevalence from a given HIV incidence. The basic assumption is that the HIV prevalence in a time period is equal to the sum of the people who have previously been infected with HIV minus those who have died. Using this model, the prevalence at time \( t \) is given by,

\[
p(t) = \sum_{i=0}^{t-1} \text{Inc}(i) F(t-i),
\]

where \( F(t) \) is the probability of surviving \( t \) time periods and \( \text{Inc}(i) \) is the HIV incidence in time period \( i \). We use a Weibull function to describe the time from infection to death, matching that used in the simulation model. We assume that HIV incidence follows a gamma distribution with a multiplicative factor \( \gamma \) to allow for differences in scale, and an additive term that is used to describe the long-term incidence,

\[
\text{Inc}(t) = \left[ \gamma \beta^{-\alpha}(t-t_0)^{\alpha-1} e^{-(t-t_0)/\beta} \right]/\Gamma(\alpha)
\]

for \( t \leq t_0 + \beta(\alpha - 1) \) and

\[
\text{Inc}(t) = \gamma \beta^{-\alpha}\left[ (1-\theta)(t-t_0)^{\alpha-1} e^{-(t-t_0)/\beta} + \theta(\beta(\alpha - 1))^{\frac{\alpha}{1}} e^{\beta(\alpha - 1)/\beta} \right]
\]

for \( t > t_0 + \beta(\alpha - 1) \). This matches the function used in Salomon and Murray 2001.

The variables \( \alpha \) and \( \beta \) set the shape of the incidence curve, \( \gamma \) sets the scale of the curve and \( \theta \) sets the level of the long-term incidence, which is equal to \( \theta \) multiplied by the peak incidence. The parameter \( \theta \) can be used to define the scenario for long-term behavior. Here, we assume \( \theta \) equals 1, i.e. that the long-term HIV incidence is constant at the peak value, and fit the HIV model to HIV prevalence data from Zimbabwe <www.unaids.org> by varying the parameters \( \alpha \), \( \beta \) and \( \gamma \). We assume that the HIV prevalence data points have normal errors and therefore find the optimal set of parameters by minimizing the sum of the squared difference between the model’s estimate for HIV prevalence and the HIV prevalence data.

4 VALIDATION

In the absence of data from the household survey, the model was validated by comparing its output with TB incidence data for Zimbabwe and characteristics typical of epidemics of any infectious disease. We begin by considering the transitory behavior of the epidemic.

![Figure 2: TB cases per 100,000 of population.](image)

Figure 2 shows the average TB incidence output for 10 runs of the model over a 300 year period. The number of TB cases can be seen to exhibit damped oscillations. Damped oscillations in the number of new infections are typical of the simple deterministic differential equation models often used to describe infectious disease epidemics, the so-called SIR models (SIR models are differential equation models with three states: susceptible, infected and recovered). In these models, the number of infections will then either stabilize at an endemic level or the epidemic dies out. The graph also shows the definite amplifying affect of HIV on TB when it is introduced to the model during the early 1980s. Obtaining behavior with these characteristics from our stochastic model is encouraging.

Figure 3 shows the observed TB incidence and HIV prevalence in Zimbabwe from 1980 onwards. HIV prevalence is increasing in Zimbabwe from 1984 and the number of TB cases shows a corresponding increase after a time lag of about 6 years, the time from initial HIV infection to developing late-stage HIV.

The corresponding model output is shown in Figures 4 and 5. Figure 4 shows the average model output for TB Incidence along with the output from 10 runs. The variability in the model estimates of TB incidence is due to the small number of individuals in the model population.
Figure 3: TB incidence and HIV prevalence in Zimbabwe.

Figure 4: Model estimates of TB incidence in Zimbabwe. The graph shows the observed TB Incidence data from Zimbabwe and the output from 10 runs of the model, along with the average result.

Figure 5: Model estimates of HIV Prevalence in Zimbabwe. The graph shows the observed HIV Prevalence data from Zimbabwe and the output from 10 runs of the model, along with the average result.

Transmission of TB occurs with a relatively low probability and TB disease is relatively rare, therefore, these results are not surprising. As the model population is increased, the proportion of runs in which the TB epidemic dies out decreases. Future iterations of the design could consider two options for the model: first use a model population that is sufficiently large for TB disease to become endemic; or second include the effect of interactions of the model population with a background pool of TB infection. This background pool of TB infection should be representative of Zimbabwe as a whole and assumes the model population to be connected with the rest of the country rather than isolated from it.

Many of the parameters used within the model will not change following the calibration of the model to the survey data. Those that will change include the distribution of household size, the magnitude of the HIV epidemic. Additional information coming from the survey will include the distribution of the population across the different TB and HIV disease states. This could be used for model validation or for setting the initial conditions.

Experimentation within the model will focus on determining strategies for reducing TB transmission by targeting case-finding for TB disease. In its current form, experimentation with the model may be slow because of the small number of TB cases. The small number of transmission events will mean that a large number of runs will need to be made to obtain a reasonable level of variability. In situations such as this, importance sampling can often be used effectively as a variance reduction measure. This is something that we hope to investigate.

We have begun with a simple method for transmission, differentiating only between transmission inside the household and transmission in the general community. Despite its simplicity this will still allow us to determine the significance of household versus community transmission and to analyze interventions targeted at household contacts of TB patients. Ultimately, we would like to use a geospa-
tial model for transmissions, where transmission between people who live close to each other is more likely than transmission between people who live at opposite sides of the region being modeled. As well as making the model more accurate, this may also shed light on the effect of poverty (indicated by a high population density) on the clustering of TB cases.

ACKNOWLEDGMENTS

We thank the Engineering and Physical Sciences Research Council who helped fund this work as part of a PhD studentship for Georgina Hughes.

REFERENCES


Murray CJL, Salomon JA. 1998a. Modeling the impact of global tuberculosis control strategies. Proceedings of...
the National Academy of Sciences of the United States of America 95(23): 13881-13886

AUTHOR BIOGRAPHIES

GEORGINA R HUGHES is a PhD student in the School of Mathematics at the University of Southampton. She previously obtained a BSc in Mathematics and an MSc in Operational Research from the University of Southampton. She is interested in the mathematical modeling of epidemics. Her email address is <G.R.Hughes@soton.ac.uk>

CHRISTINE S M CURRIE is a lecturer of Operational Research in the School of Mathematics at the University of Southampton, where she also obtained her PhD. Her research interests include mathematical modeling of epidemics; bayesian statistics; variance reduction methods; the optimization of simulation models and dynamic pricing. Her e-mail address is <ccurrie@soton.ac.uk> and her URL <www.maths.soton.ac.uk/staff/currie>.

ELIZABETH L. CORBETT MB BCh, PhD, is a clinical epidemiologist and a Senior Lecturer in the Clinical Research Unit, London School of Hygiene and Tropical Medicine. She works full time in epidemiological research, based in the Biomedical Research and Training Institute, Harare, Zimbabwe. Her ongoing research projects are centred around the theme of TB epidemiology and control in high HIV prevalence areas, funded by the Wellcome Trust, UK. Her email address is <elc1@mweb.co.zw> or <liz.corbett@lshtm.ac.uk>