
1

OPTIMIZATION IN INFECTIOUS DISEASE CONTROL AND PREVENTION: TUBERCULOSIS MODELING USING MICROSIMULATION

SZE-CHUAN SUEN

Daniel J. Epstein Department of Industrial and Systems Engineering, University of Southern California, Los Angeles, CA, USA

Compared with many other optimization problems, optimization of treatments for national infectious disease control often involves a relatively small set of feasible interventions. The challenge is in accurately forecasting the costs and benefits of an intervention; once that can be evaluated for the limited set of interventions, the best one can be easily identified. Predicting the outcome of an intervention can be difficult due to the complexity of the disease natural history, the interactions between individuals that influence transmission, and the lack of data. It is therefore important to understand how a particular disease affects patients, spreads, and is treated in order to design effective control policies against it.

One such complex disease is tuberculosis (TB), which kills millions of people every year. It is transmitted through respiratory contacts, has a latent stage, and is difficult to diagnose and cure in resource-constrained settings, and treatment success varies by

demographic factors like age and sex. Moreover, the mechanisms of disease transmission are not fully known, making modeling of transmission difficult, and it is particularly prevalent in areas of the world where reliable disease statistics are hard to find.

All of these characteristics make TB a difficult disease to model in the settings where choosing an optimal control policy is most important. Traditional compartmental disease models may become intractable if all relevant demographic and treatment stratifications are specified (state space explosion), so a microsimulation may be a good alternative for modeling TB dynamics. In a microsimulation, individual health and treatment states are probabilistically simulated over time and averaged together to form population statistics. This allows for greater modeling flexibility and a more tractable model but may also result in problems of model stochasticity.

In this chapter, we first discuss the epidemiology of the disease, illustrating why TB modeling is necessary and highlighting challenging aspects of this disease. In the second section, we provide a brief overview of simulation and then discuss in depth a microsimulation model of TB to illustrate subtleties of using microsimulation to evaluate policies in infectious disease control.

1.1 TUBERCULOSIS EPIDEMIOLOGY AND BACKGROUND

In order to understand how to pick a model framework and implement a useful model, it is important first to understand the epidemiological characteristics and background of the disease. TB is caused by the bacteria *Mycobacterium tuberculosis*, which can attack the lungs (pulmonary TB) or other parts of the body (extrapulmonary TB). TB is a respiratory disease and transmitted through the air by coughing or sneezing. It has been declared a global public health emergency, killing 1.3 million people in 2012, while 8.6 million people developed the disease. The majority of cases were in Southeast Asia, African, or Western Pacific regions (Zumla et al. 2013). However, the disease varies by region and cannot be treated identically in all areas—for example, many African cases are concurrent with HIV, while in other regions, like India, HIV prevalence is low although TB prevalence is high (World Health Organization 2013). This means that models for one country may not be easily adapted to another, since comorbidities and the driving factors of the epidemic may be quite different.

Once contracted, TB may stay latent for many years and only activates in about 10% of cases. Latent TB is asymptomatic and cannot be transmitted. Activation rates depend on immunological health and have been observed to vary by demographic factors, like age (Horsburgh 2004; Vynnycky and Fine 1997), and behavioral factors, like smoking (Lin et al. 2007). Transmission of TB, which occurs through respiratory contact, may vary by age (Horby et al. 2011; Mossong et al. 2008), demographic patterns, and cultural trends but is poorly documented or understood.

Nondrug-resistant strains of TB, whether latent or active, are treatable using antibiotics, but misuse of first-line antibiotic regimens may lead to drug-resistant or

multidrug-resistant (MDR) TB, defined as strains that are resistant to at least isoniazid and rifampin, two first-line TB drugs. Premature treatment default or failure can result in the development of drug resistance, and drug-resistant strains may then be transmitted to other individuals. Drug-resistant TB can be treatable, depending on the level of drug resistance (pan-resistant TB strains have emerged), but require more expensive second-line antibiotic regimens of longer duration (drugs need to be taken many times a week for up to 2 years) with higher toxicity rates and lower cure rates. Therefore optimization of treatment policies needs to take imperfect treatment behavior and potential drug resistance into account. Drug-resistant TB prevalence varies by region, and this also contributes to the necessity of geographical specificity when evaluating potential TB control mechanisms.

Latent and active TB can be detected through a variety of different tests of varying sensitivity and specificity, and different tests may be preferred in different regions. For instance, Mantoux tuberculin skin test (TST) or interferon-gamma release assay (IGRA) blood test are used to detect TB infection in many areas with low TB prevalence, whereas sputum smear microscopy tests are commonly used to identify active TB cases in areas of high prevalence (Global Health Education 2014). While sputum smear tests have fast turnaround times and low costs, sputum smear tests have low sensitivity and active TB cases may be overlooked. Bacteriological culture may take up to several weeks but is a more accurate diagnosis method and can be used for drug susceptibility testing (it can be used to identify drug-resistant samples from susceptible TB samples). Initial diagnosis can also be passive (patients self-present at local clinics) or targeted (active case finding, contact tracing, etc.). After entering treatment, patients may undergo different tests sequentially to monitor treatment efficacy and determine if second-line treatment is necessary. The cost and the effectiveness of various screening policies vary by patient behavior, latent and active TB prevalence, and what treatment options are available. Identifying optimal region-specific timing and type of diagnosis is an area of active research (Acuna-Villaorduna et al. 2008; Winetsky et al. 2012).

TB infection and disease may be complicated by comorbidities. TB is often observed along with HIV, which can change the natural history of disease and complicate TB diagnosis and treatment. In 2012, 1.1 million of the 8.6 million new cases of TB were among people living with HIV (World Health Organization 2013). HIV patients have a higher risk of developing TB due to immune system compromise. Diabetes is another comorbidity that can change TB activation rates (World Health Organization 2011). While helping patients with multiple chronic diseases is an increasingly important part of TB control, modeling multiple diseases is challenging since the diseases interact and data to inform joint distributions on risks and rates may be scarce.

1.1.1 TB in India

India is the country with the largest number of TB cases—roughly 23% of the global total—despite large gains in the last few decades in decreasing TB mortality, incidence, and prevalence through TB treatment and diagnosis (World Health Organization 2015).

India has a federally funded TB treatment program called the Revised National Tuberculosis Control Program (RNTCP). This program offers the approved antibacterial drug regimens for treating TB, called Directly Observed Treatment, Short Course (DOTS), where health workers help patients administer their drugs to help ensure that they are taken correctly. These regimens require treatment for at least 6 months of treatment and may be longer for those patients who have previously been treated for TB (RNTCP 2010).

Despite this federally funded program, and unlike in many other countries with high TB burdens, many TB patients in India seek care in private sector clinics. Since the symptoms of TB can easily be mistaken for routine respiratory illnesses, many patients tend to first seek care from retail chemists or informal health providers in the private healthcare market. These private clinics may not have health practitioners trained in identifying and treating TB (Tu et al. 2010; Uplekar and Shepard 1991; Vandan et al. 2009), and patients using private clinics may use multiple clinics as they attain temporary relief from symptoms that then recur (Kapoor et al. 2012). This delay to getting appropriate TB care means that patients begin effective treatment at a later stage of their disease, may have infected others with TB, and may have been exposed to anti-TB drugs that can select for drug resistance.

Combating drug-resistant TB is a continuing challenge for India. More than half of the MDR-TB cases notified in 2014 occurred in India, China, and the Russian Federation (World Health Organization 2015). India started the federally funded DOTS-Plus MDR-TB treatment program in 2007, where MDR-TB patients can get access to the necessary 18–24 months of second-line TB antibiotics. However, the long treatment duration and drug toxicity make treating MDR-TB difficult, and patients may default from treatment, potentially generating more resistant disease strains. Cases of extensively drug-resistant TB (XDR-TB), where the MDR-TB strain is additionally resistant to fluoroquinolone and a second-line injectable antibiotic, have also been documented in India (Michael and John 2012).

While comorbidities can often complicate TB treatment, HIV comorbidity is relatively less common in India as in some other countries with high TB burdens: 4% of TB patients are HIV positive in India (as opposed to 61% in South Africa). For this reason we will be considering a simulation model of TB for India that does not specifically consider any comorbidities. We turn to simulations for disease control in the next section.

1.2 MICROSIMULATIONS FOR DISEASE CONTROL

A variety of model types can be used to model the diversity of issues in TB control and prevention. However, while natural disease dynamics and treatment policies can be approximated using difference equations, these may be difficult to solve analytically and require simulation to arrive at numerical answers. Simulations imitate the real system using a probability distribution to generate random events and obtain

statistical observations of system performance. Simulations can provide not only the epidemiological trends or costs of each treatment arm being considered but also disease trajectories over time.

One common disease modeling method is to use a compartmental model, where states are formed using health, treatment, or demographic status, depending on the complexity of the model. State transition probabilities can be estimated from the published literature or estimated from survey data, and the probability of an individual or population acquiring disease, incurring treatment costs, or any other outcome of interest, can be estimated by starting the model in one state and applying transition probabilities as time advances. The model would then provide the mean performance outcome at every time period. These models can be very useful and may be applicable for a variety of problems. They are discussed in detail in another chapter of this book.

Unfortunately, these models can quickly become intractable if the state space becomes too large, as can happen if a large number of stratifications are required. To illustrate this, suppose a hypothetical TB model included different transition probabilities for individuals of different ages (age 0–15, 16–45, 46–60, and 60+), sexes (male or female), and TB status (healthy, infected, or active disease). It would have $4 \times 2 \times 3 = 24$ states, and the modeler needs to specify transition probabilities to each of the other states. One can easily see that the number of states would quickly become very large if the model wished to use a finer age stratification (i.e., 1 year age bins) and include demographic characteristics about treatment status, TB strain (i.e., different strains of TB by drug resistance), and past treatment status. The model would potentially become difficult to work with. However, these patient characteristics may be important to capture to accurately reflect TB dynamics.

A microsimulation overcomes this issue by simulating an individual unit (in this case, an individual) over time instead of estimating the mean outcomes for a population. This allows the modeler to specify behavioral characteristics at a very detailed level if necessary—probabilities of disease progression or treatment can depend on the individuals' demographic characteristics and history. Using a random number generator, outcomes for each individual can then be probabilistically determined and recorded at each time period; a population of individuals of sufficient size should then generate the same average outcomes as those estimated in the compartmental model. However, in addition to providing the mean outcome measures of interest, the microsimulation can also provide the distribution of the measures of interest since every individual's health and treatment state is estimated at every time period. We will illustrate this in the second section of this chapter, in which we will describe in detail a microsimulation of TB in India.

However, before delving into that example, it may be useful to discuss and summarize the advantages and disadvantages of simulation and microsimulation in particular. Simulation methods are powerful and can help us find numerical estimates for outcomes that are analytically intractable. Unlike some analytical models, they easily allow the modeler to examine transient effects, not just model outcomes in

steady state. Microsimulation also allows for a great deal of modeling flexibility, since the modeler can easily add characteristics to individuals without specifying another set of states. Organizational structure in a microsimulation is more robust than a compartmental state transition model, since the modeled population can have as many characteristics as a modeler needs without the number of compartments becoming intractably large. However, the modeler also needs to be careful about stochastic fade-out, when no individuals have a certain characteristic just due to chance alone. For instance, since the number of individuals with TB compared with the total population of a country is small, it is likely that no individuals in the simulation will have TB if the total simulation population is too small. For example, consider a country where the TB prevalence is 0.1%; if the simulation only models 1000 individuals, and the individual with TB dies before transmitting the disease, it would look as if TB had been eradicated! Now imagine an analogous case where the modeler cared about TB in different subgroups of the population—then the number of individuals in these subgroups could be small, and the corresponding number of individuals with TB in that group would be even smaller. Therefore the larger the number of characteristics, the larger the modeled population is necessary so that individuals of all characteristics will be represented. If the number of individual characteristics is large, this may mean long computation times (if the simulation population is large) or noisy outcome measurements (if the simulated population is too small).

1.3 A MICROSIMULATION FOR TUBERCULOSIS CONTROL IN INDIA

To illustrate how microsimulation can be used for disease control and prevention, we are going to discuss in detail a microsimulation of the TB epidemic in India (Suen et al. 2014, 2015). This simulation was used to evaluate the impact of TB transmission prevention versus improving treatment as well as to evaluate the cost-effectiveness of treatment policies. The model uses a dynamic transmission model of TB that was calibrated to Indian demography and TB epidemiology from 1996 to 2010 and then projected into the future (until 2038 in the case of the transmission prevention analysis and until 2025 in the cost-effectiveness analysis) and includes health and treatment characteristics for nondrug-resistant and drug-resistant latent and active TB. Since TB dynamics and treatment trajectories can depend on age and sex, among other demographic factors, the model stratifies individuals these characteristics. This means that probabilities of mortality, transmission, activation, and treatment uptake and effectiveness vary by age and sex in the model. This level of detail precluded the model from using the more common compartmental model structure due to the large state space; it makes more sense as a microsimulation. Inclusion of these stratifications allows treatment policies to differentially influence different demographics, which may be of particular interest if there are particular demographics a modeler is interested in

(i.e., the elderly, school-aged children, etc.). Treatment availability and effectiveness were also included to estimate the effect of treatment policies that were already in effect.

Since not all TB dynamics parameters were known with certainty, this model was calibrated by changing activation and treatment uptake parameters until TB prevalence, incidence, and a variety of treatment demographic characteristics matched values from the literature. After validation, the model was used to estimate incidence, prevalence, and mortality from drug-resistant and nondrug-resistant strains of TB for projection into the future for scenarios where either treatment or diagnosis policies were improved. These estimates could then be used to provide insight on the effectiveness of intervention policies and the cost delaying such efforts. With this information, one could better optimize treatment versus prevention policies for controlling TB in India. We will discuss the model building and analysis process, from population inputs to calibration to treatment, in detail in the next sections.

1.3.1 Population Dynamics

In this microsimulation, individuals are simulated from birth to death as they pass through various health and treatment states. To accurately recreate the Indian population dynamics, the population growth rate in the microsimulation matches historic and projected trends. Non-TB mortality probabilities were calculated using World Health Organization (WHO) life tables for 1990, 2000, and 2009 (World Health Organization 2010a). All individuals in the microsimulation are exposed to age- and sex-specific background mortality, and those with active TB have an additional disease-specific risk of death. The resulting age structure in the model stabilizes during the “burn-in” period and shows population aging thereafter, especially over 2013–2038, as in reality. The “burn-in” period is where the model is run until population demographics and disease prevalence levels stabilize to steady-state levels that match observed pretreatment levels in India. This ensures that treatment policies implemented during the analysis period are due to the treatment arms of interest and not population dynamics effects.

The model uses a simulated population of 6.5 million people in 1996 that grows to 10 million by 2038. Because the model considers a lesser number of simulated individuals than in India in reality, these numbers were ultimately scaled to the total Indian population size to consider impacts on disease burden (thus, proportions of the modeled population are multiplied by the corresponding actual Indian population in a given year).

1.3.2 Dynamics of TB in India

After general population dynamics matched observed trends, the model needed to incorporate TB natural history. To do this, the model simulates individuals acquiring disease with some risk, activating from latent to active TB with some risk, and entering, defaulting, or failing treatment with some risk. Each of these probabilities is age and sex dependent to reflect medical and demographic data.

There are many ways to simulate different mixing patterns for disease transmission. It is common to assume homogeneous mixing, where all patients capable of transmission have equal probability of meeting, and infecting, a susceptible individual, as in classic susceptible–infected–susceptible (SIS) or susceptible–infected–recovered (SIR) models of infectious disease (Kermack and McKendrick 1927). In such models, individuals transition between disease-susceptible and infected (and in the case of SIR models, recovered) health states according to a set of differential equations. The probability of acquiring TB for a susceptible individual is calculated using the proportion of transmitting individuals multiplied by the proportion of susceptible individuals, scaled by a transmission factor.

With a microsimulation, however, we can add a more detailed representation of transmission as needed. If the data is available, individuals can have higher or lower risks of infection if an infected individual is in their family group or community, or vary by age, if individuals of certain ages are more social or susceptible to acquiring infection. Complex infection dynamics can be easily simulated—the only difficulty is to ensure that the parameters reasonably reflect reality. Since data is often scarce and simple models should be preferred to complex models, the most general modeling approach that still captures the relevant disease dynamics should be used. In this microsimulation example, the model used a “who-mixes-with-whom” transmission matrix to allow the probability of acquiring disease to vary by age while still assuming homogeneous mixing within age groups.

Figure 1.1 visually represents this matrix, where the colors represent the frequency of contacts of susceptible individuals across different ages (0–5, 6–10, etc.). Therefore the product of this matrix with the proportion of infectious individuals in each age group, multiplied by the per-contact probability of acquiring TB, should give a vector of probabilities of acquiring disease for susceptible of different age groups (the matrix need not be square, if different age brackets were used for infected and susceptible populations). During implementation, the microsimulation counts up the number of infected individuals in the population, calculates proportions, and applies this matrix to get the probability of infection and the number of newly infected individuals in this time period.

It is important to note that data should support the modeling assumptions made. In this case, the data about age-specific mixing was calculated from the published literature on respiratory contacts, and the per-contact infection probability, a scalar, was calibrated (Mossong et al. 2008). We discuss model calibration and validation further in the chapter.

1.3.3 Activation

Once an individual is infected with TB, TB may stay latent within the individual for years, or may never activate. Activation rates change over time from infection (Horsburgh 2004; Vynnycky and Fine 1997) and other factors like immune system

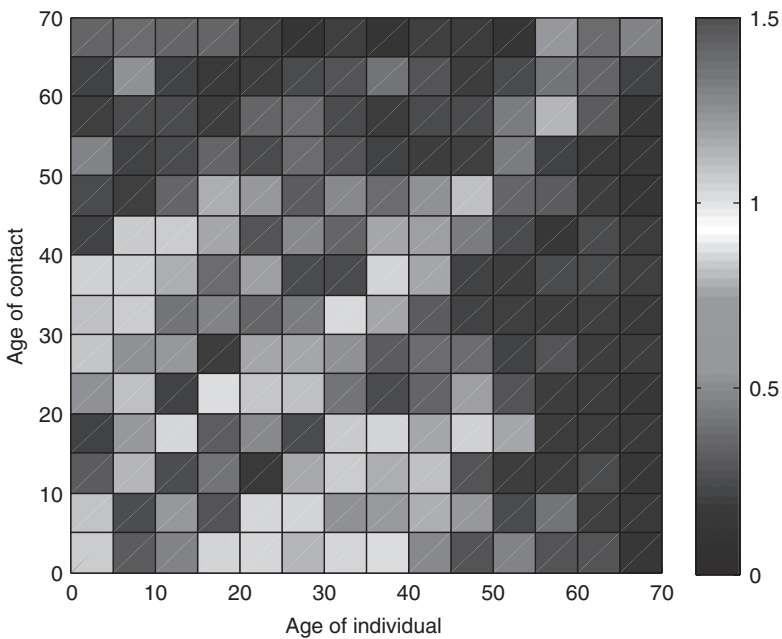


Figure 1.1 Who-mixes-with-whom matrix for a microsimulation. Source: Suen et al. (2014). <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0089822> Attribution 4.0 International (CC BY 4.0) <https://creativecommons.org/licenses/by/4.0/>.

compromise (as with HIV) or exposure to smoke (such as through smoking or cook fires (Lin et al. 2007)). Unlike active TB, latent TB cannot be transmitted and does not cause decreased quality of life, so accurate modeling of activation rates is important for capturing disease dynamics.

In the case of this microsimulation, activation rates varied over age as well as time from infection according to data from the literature (Horsburgh 2004). Generally, TB activation tends to be higher in the years right after infection and then declines (Horsburgh 2004; Vynnycky and Fine 1997), and this was reflected in the microsimulation activation rates. Since the simulation used data from a published medical study not conducted in India, the overall activation rate was calibrated in order to reflect the average activation rates in the country of interest (more on that in the calibration section).

1.3.4 TB Treatment

Before modeling treatment policies of interest, baseline treatment trends must be accurately captured—if treatment programs are already in effect, additional treatment policies must be evaluated compared with a baseline where these existent programs continue to act or risk overestimating the treatment policy. In India, a federal program

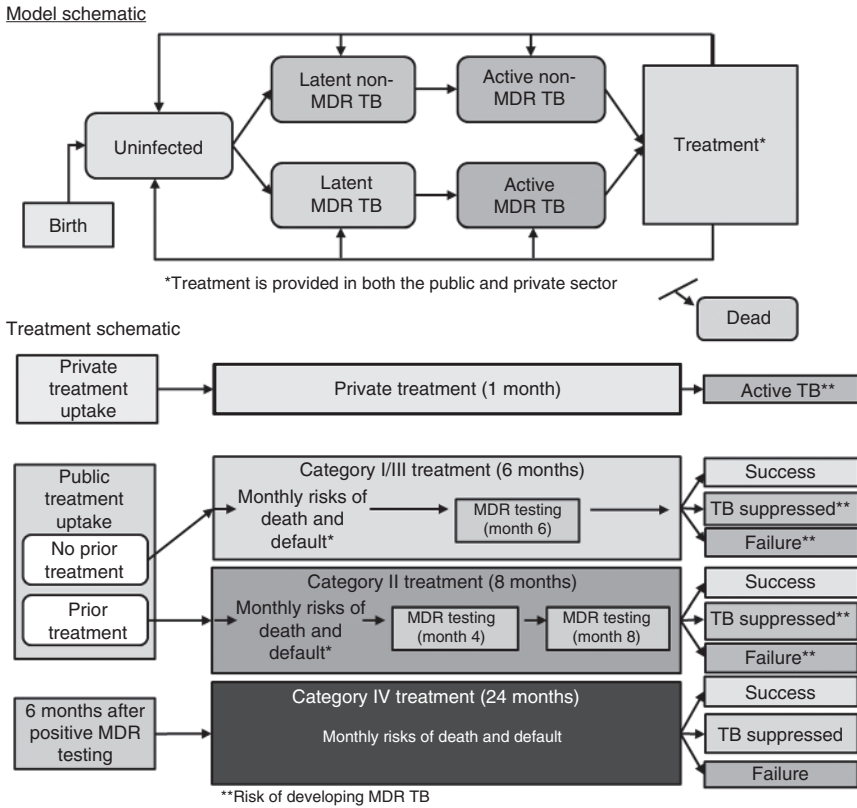


Figure 1.2 Heath state transitions and treatment schematic for a microsimulation. Source: Suen et al. (2014). <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0089822> Attribution 4.0 International (CC BY 4.0) <https://creativecommons.org/licenses/by/4.0/>.

that scaled up in the 1990s and 2000s is already in place. Since this microsimulation used 1990–2000 as its calibration period, it needed to simulate access to care, treatment uptake, treatment success, and default rates. In this case, since a key outcome parameter was MDR-TB prevalence and incidence, which can be caused by imperfect treatment adherence, these parameters were very important to model accurately. Therefore the microsimulation uses detailed representations of different treatment regimens used in India, shown in Figure 1.2, which vary by prior treatment status and whether the patient has tested positive for MDR-TB. Monthly default and death rates, which vary over treatment regimen, were calculated from survey data from Bihar. These regimens were incorporated in the model by exposing patients on TB treatment to the appropriate death, default, and cure rates for their treatment regimen for the appropriate duration. A schematic of the treatment module for the microsimulation is shown in Figure 1.2.

1.3.5 Probability Conversions

Clearly, the microsimulation relies heavily on the probabilities it uses—whether they describe activation, treatment default, or death, accurate measures are needed to get a valid estimation of disease measures. These probabilities can be calculated from data in the published literature or survey data, but often they will not be presented in a way that can be used directly. When calculating these probabilities, it is important to distinguish rates from probabilities and risk ratios from odds ratios.

Probabilities must lie between 0 and 1, and mutually exclusive, exhaustive events must sum to 1. They describe the possibility of an event occurring over a set time period. Rates, on the other hand, can be larger than 1 and are not associated with a time period—they are formally instantaneous rates. Usually rates are assumed to be constant over particular time periods (as in mortality tables, e.g., where death rates assumed to be constant for individuals between the ages of 0 and 1, 1–2, etc.)

Rates can be converted into probabilities using the following equation, where r is the instantaneous rate, p is the probability, and t is the time period. For instance, if one is converting an annual rate into a monthly probability, t would be 1/12:

$$p = 1 - \exp(-rt)$$

The microsimulation uses probabilities, not rates, and they should be scaled to the correct duration. Scaling should be done on rates, not probabilities (this is easy to remember if one tries to double a probability larger than 50%—it becomes larger than 1, which cannot be a valid probability).

The literature will often also report relative risk ratios and odds ratios, which also need to be converted into probabilities. A relative risk is a ratio of probabilities—the probability of an event happening to one group divided by the probability of that even happening to another group. An odds ratio, on the other hand, is a ratio of odds—the odds of an event happening to one group divided by the odds of it happening to another group, where odds are equal to the probability of the event happening divided by the probability of it not happening (a 2-to-1 odds of something happening, for instance, means 2/3 probability it will happen and 1/3 it won't). Relative risks and odds ratios cannot be converted into absolute probabilities in isolation, and a modeler will have to find the probability of that event happening to one of the groups and then solve for the probability of the event happening to the other group. To illustrate, if the relative risk of dying from TB for smokers to nonsmokers is 1.5, and the average probability of death for nonsmokers with active TB is 20%, then the probability of death is 1.5 times higher for smokers than nonsmokers (30%). In general, collecting and converting probabilities for all the demographics of interest may be a challenging task, so it is important to do carefully.

In our example microsimulation, the authors needed the probabilities of death, default, and failure stratified by age and sex. To do this, they used data from the literature that provided the odds ratio of male to female defaults, the proportion of males in treatment, the overall default probability, and the total number of people in treatment.

They then solved a system of equations using the definition of odds ratio to find the age- and sex-specific probabilities of death and default. The system of equations for stratifying default by age is shown as follows, where the unknowns to solve for were A , B , C , and D (taken from Suen et al. 2014):

$$\begin{aligned} \text{Odds ratio of male to female defaults} &= (A/B)/(C/D) \\ \text{Proportion of males in treatment} &= (A+B)/(A+B+C+D) \\ \text{Overall default probability} &= (A+C)/(A+B+C+D) \\ \text{Total number of people} &= A+B+C+D \end{aligned}$$

where

- A = number of males defaulting
- B = number of males not defaulting
- C = number of females defaulting
- D = number of females not defaulting

1.3.6 Calibration and Validation

Some probabilities may not be known, however, and that is where calibration comes in. In our example microsimulation, the overall transmission probability and the activation rate as well as overall treatment uptake probabilities were calibrated since little is known about these parameters in the published literature. Calibration involves identifying model parameters that allow model outputs to best fit certain output targets, aptly called calibration targets. Models may also do this for parameters that are known but have wide uncertainty ranges. In this microsimulation, the calibration targets were WHO estimations of incidence and prevalence over the 1990–2010 period.

There are many methods to go about calibrating a model. Essentially, this is just an optimization problem where the modeler tries to minimize some measure of distance between the calibration targets and model outputs by varying the uncertain or unknown model parameters over reasonable ranges. Since microsimulations are generally complex, it is usually not possible to represent this as an analytical problem. However, traditional algorithms for searching over the feasible space can be used (i.e., Nelder–Mead, etc.). In our example microsimulation, the modelers used a grid search to explore the feasible space since the feasible space was small and could be reasonably explored using this method. It was also relatively easy to implement and performed well (see Figure 1.3).

In the microsimulation, the calibration process was used to find two parameters related to overall TB and one parameter related to treatment seeking behavior. The parameters were (i) an activation rate, which determines the average time to activation for individuals with latent TB infections; (ii) the effective contact rate, a parameter that determines the average probability of TB transmission given a

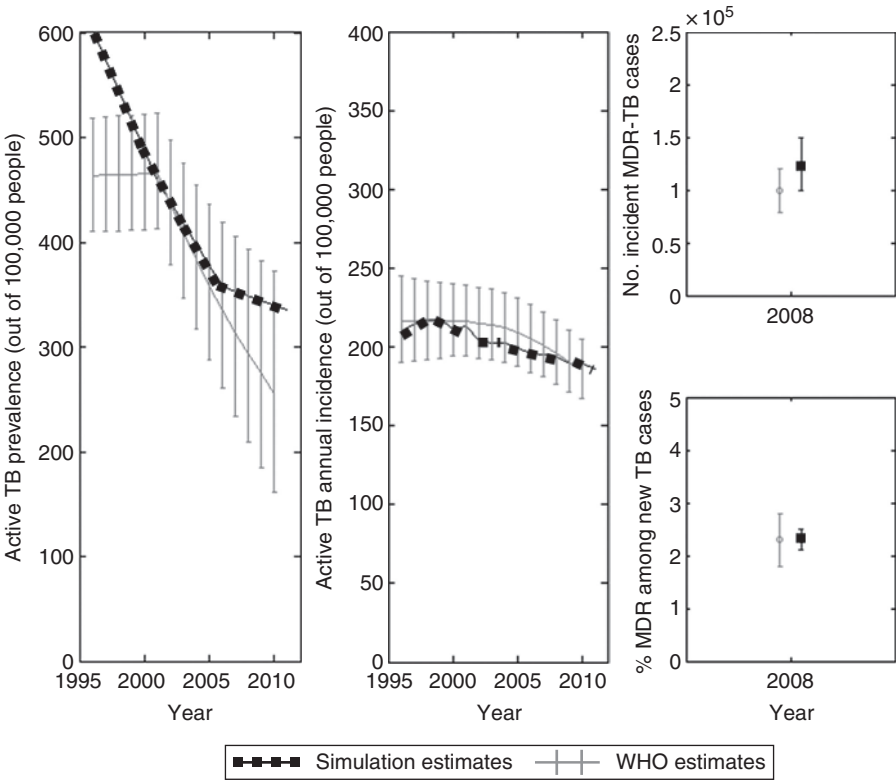


Figure 1.3 Calibration targets for microsimulation model of TB. WHO estimates are compared against model outputs. Source: Suen et al. (2014). <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0089822> Attribution 4.0 International (CC BY 4.0) <https://creativecommons.org/licenses/by/4.0/>.

contact between a susceptible and infectious individual; and (iii) the average probability of undergoing TB testing among individuals.

Figure 1.3 provides a visual representation of how a microsimulation model outputs' might look compared with empirical data. In this case, the model outputs were calibrated to WHO data for Indian TB prevalence and incidence over the 1995–2010 period and measures drug resistance in 2008. During the calibration period, the model includes treatment present during that period, such as the federal treatment program. This is what drives the decrease in disease prevalence in Figure 1.3 (leftmost panel), as the federal treatment program was scaled up over that period.

While a model cannot exactly match all empirical statistics of the population of interest—and attempting to do so would result in an overly complex model—it is important that the model behaves realistically enough to provide reasonable and useful projections of the future. After calibration, it is important that the model is

validated against external epidemiological and demographic measures to confirm that the population and disease dynamics are consistent with reality. These are measures that were not calibrated to (which is why they are called external validation measures). The model should be validated against epidemiological measures that are important for the analysis (ensure that baseline prevalence is consistent with the literature, if the main outcome measure is prevalence, for instance).

Table 1.1 provides validation measures for the microsimulation. These include demographic measures such as life expectancy, for males and females, as well as disease and treatment metrics. These values depend on the rate of transmission and activation, so validating that they match observed values provides evidence that the calibration was reasonable.

1.3.7 Intervention Policies and Analysis

After the model is validated, we can finally begin our policy analysis. The treatment policies considered must first be translated into changes in model parameters—this is one reason that it is important to have clear treatment policies at the beginning, so the model can be built to capture pertinent treatment characteristics.

Microsimulations offer more flexibility than other types of models since there is no assumption of analytical form and individuals can behave differently. This can be particularly useful for examining subgroups (e.g., incidence of MDR-TB in children). In the case of our example microsimulation, the authors were interested in examining prevalence of transmitted versus acquired drug-resistant TB (Suen et al. 2014), and this could be done by counting the number of individuals activating with transmitted or acquired drug-resistant TB without adding a separate compartment to stratify these populations (as would be needed in a compartmental model).

The microsimulation was used to evaluate the cost-effectiveness of several policies that might improve diagnosis and treatment quality in India. The WHO approved Cepheid GeneXpert diagnostic systems (Cepheid, Sunnyvale, CA, USA) for TB in 2010, but while these systems may be able to provide faster and more accurate TB diagnoses, it was unclear at the time of the analysis that these expensive systems should be implemented in resource-constrained settings like India. Using scarce public health treatment funds for improvements in diagnosis could trade off with other policies to combat TB, like improving the quality of care. For instance, since clinic quality in India may vary widely, pilot programs have been testing whether it would be effective to refer patients in low-quality private clinics to federally sponsored clinics. However, at the time of the analysis, it was unclear such a public-private mix (PPM) program would be cost-effective at a national level.

The simulation was therefore used to evaluate the cost-effectiveness of PPM, whether GeneXpert diagnostic systems should be used for all TB diagnosis or only for diagnosis of MDR-TB (“drug sensitivity testing” (DST)), or whether GeneXpert and PPM should be used in combination. The six interventions evaluated were then (i) the status quo, (ii) GeneXpert for all TB diagnoses, (iii) GeneXpert for DST,

TABLE 1.1 Example Simulation Validation Measures

	Empirical Estimate (95% CI, If Available)	Simulation Estimate (95% CI)	Source
<i>Demographics</i>			
Male life expectancy, 1990 (years)	57.2	55.9 (55.6–56.2)	World Health Organization (2010a)
Female life expectancy, 1990 (years)	57.9	55.7 (54.3–57.1)	World Health Organization (2010a)
<i>Tuberculosis</i>			
Life expectancy post TB activation without treatment (years)	~3 years	3.11 (2.95–3.28)	Tiemersma et al. (2011)
Lifetime fraction of latent infections that activate	0.10–0.20	0.17 (0.15–0.18)	Horsburgh (2004)
<i>Treatment</i>			
Average delay from symptom onset to RNTCP treatment (months)	2–7	6.1 (5.63–6.57)	Kapoor et al. (2012), Pantoja et al. (2009), Rajeswari et al. (2002), Selvam et al. (2007), Tobgay et al. (2006)
Median delay from symptom onset to RNTCP treatment (months)	2–7	3.9 (3.72–4.08)	
Default rates			
Category I	0.06	0.054 (0.047–0.061)	Central TB Division, Directorate General of Health Services, and Ministry of Health and Family Welfare (2010)
Category II	0.14	0.151 (0.119–0.183)	Central TB Division, Directorate General of Health Services, and Ministry of Health and Family Welfare (2010)
<i>MDR-TB Incidence in 2008 Using WHO Calculation Methods</i>			
% MDR among new TB cases	2.3 (1.8–2.8)	2.3 (2.1–2.5)	World Health Organization (2010b)
Number of MDR-TB among incident new and relapse TB cases	55,000 (40,000–74,000)	52,000 (47,000–56,000)	
Number of incident acquired MDR-TB cases	43,000 (33,000–56,000)	73,000 (52,000–94,000)	
Number of MDR-TB among incident total TB cases	99,000 (79,000–120,000)	124,000 (99,000–150,000)	

Source: Suen et al. (2014) <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0089822> Attribution 4.0 International (CC BY 4.0) <https://creativecommons.org/licenses/by/4.0/>.

(iv) PPM, (v) PPM combined with GeneXpert for all diagnoses, and (vi) PPM combined with GeneXpert for DST.

These interventions were modeled in the simulation by changing input parameters (ex., using GeneXpert for diagnosis would increase diagnosis accuracy, so the probability of being correctly diagnosed in the simulation was increased to the appropriate level). The simulation was then run with these modified parameters in order to generate outcome measures associated with each intervention. In the next sections we discuss the outcome metrics of the simulation and how cost-effectiveness was evaluated.

1.3.8 Time Horizons and Discounting

Common treatment measures are deaths averted, life years gained, QALYs or DALYs gained, and epidemiological measures like prevalence and incidence. A QALY is a life year adjusted for quality of life, such that a year of perfect health is worth one QALY and a year living with some health compromise is worth less (Weinstein et al. 2009). A DALY is similar but measures life years lost (World Health Organization 2014a).

While epidemiological measures are commonly reported in number of cases per population of 100,000, QALYs/DALYs or treatment costs are often discounted and aggregated to values accumulated per lifetime. In a hypothetical case, a study looking at the effect of mammography screening may report that a particular screening strategy increases a 40-year-old woman's lifetime discounted QALYs by 0.2, indicating that the model follows the individual for their entire life span, counts the QALYs accumulated during that time, and discounts them to the net present value using a discount factor (conventionally 3%).

While the treatment period may be defined (say, at 10 or 25 years), the benefits of that policy may last much longer than that, particularly in the case of a transmissible disease. For instance, suppose a treatment policy increased cure of TB, which has the immediate benefits of reducing mortality and also preventing onward transmission. The policy has then also potentially saved the lives of those who would have been infected by that individual and the lives of those who would have been infected by *those* individuals even later on. Therefore each prevented transmission may have benefits far in the future, and it is important to capture those in the analysis. Perhaps it is easier to see why in handwashing, for instance, clearly, those who have the flu already are not benefited by washing their hands, but their friends are certainly happy that they did so when they themselves are not running a fever a few days after seeing them! In this example, only considering the immediate effects (inconvenience of handwashing) is outweighed by the future benefits (friends not getting sick later), and as a forward-looking society, we should promote handwashing to prevent disease transmission.

But in a dynamic transmission model of an entire population, as in our example microsimulation of TB in India, there is no clear period after which to stop counting

benefits and costs. Since the entire population of India is modeled, there is no clear “lifetime” at which to stop counting the QALYs accumulated. One method is to run the microsimulation until the costs and benefits are so far in the future that their net present value is negligible—mathematically saying that we don’t care about those effects since they are so far away. Depending on the discount factor and the magnitude of costs and benefits, this may take different amounts of time. However, if discount factors are low and benefits and costs are large, this approach may be unfeasible because it would require too much computation time or undesirable because it is unreasonable to expect that the disease/treatment behaves as we modeled so far into the future (since the model cannot take into account unexpected technological breakthroughs). Another approach is to stop the simulation at the end of the analysis period and ignore all costs and benefits accrued afterward. This may incur those disadvantages discussed earlier to different degrees depending on the length of the analysis duration. One could also take an intermediate approach, where transmission is not considered after a certain period, or the simulation considers only a certain cohort after some time. The necessity of these approaches depends on the computational intensity of the microsimulation, disease dynamics, and the magnitude of the treatment policy affects.

In our example cost-effectiveness analysis, the modelers used such an intermediate approach, where they used a time horizon of 10 years from when the analysis was conducted in 2015, and then considered the lifetime costs and QALYs associated with those still alive at the end of those 10 years without further disease transmission. This approach reduces computational time and does not make assumptions unrealistically far into the future (compared with running the simulation until the discount factor reduces costs and QALYs to essentially zero) and still captures some of the long-term costs and QALYs generated by the intervention (unlike the approach where all costs and QALYs after the time horizon are not considered).

1.3.9 Incremental Cost-Effectiveness Ratios and Net Monetary Benefits

Once calculated, the costs and benefits for each intervention are usually plotted on a cost–benefit plane, and the incremental cost-effectiveness ratio (ICER) is calculated for each strategy. This is given as the incremental cost divided by the incremental benefits of each strategy between it and the next cheapest strategy. In essence, this provides the marginal cost to gain the marginal benefit and has units of dollars per QALY gained (or life year gained or DALY averted). A policy is generally said to be “cost-effective” if it costs less than three times the GDP per capita to gain one QALY and is “very cost-effective” if it costs less than the GDP per capita (R. Hutubessy et al. 2003).

The efficient frontier in Figure 1.4 (interventions on the blue line) highlights policies that are cost-effective (those labeled in white boxes with ICER). Dominated policies are shown off the efficient frontier and labeled with gray boxes. Monte Carlo simulation sampling uncertainty for the costs and QALYs of each strategy is depicted

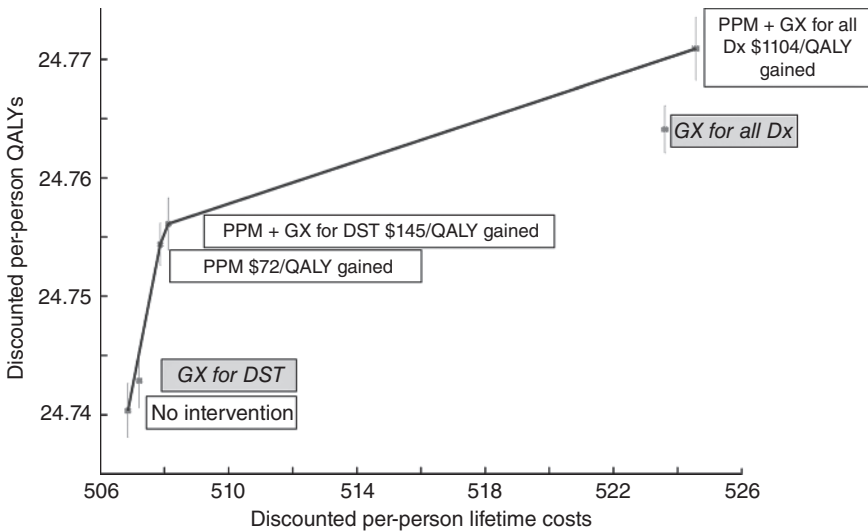


Figure 1.4 Cost-effectiveness frontier. Dx, diagnosis; GX, GeneXpert (Suen et al. 2015). Source: Reprinted with permission of the International Union Against Tuberculosis and Lung Disease. Copyright The Union.

as red 95% confidence intervals. In this analysis, even the most expensive policy, PPM with GeneXpert for all diagnosis, is cost-effective. It has an expected cost of \$1103.58 per QALY gained, which is less than one GDP per capita in India (\$1450). Even with sampling noise, this finding occurs with 99% probability.

Since it is often more intuitive to compare values in dollar units, the net monetary benefit is another way to represent intervention costs and benefits. It converts the total discounted lifetime benefits and costs into a single scalar value. It is calculated as the willingness-to-pay threshold times the total benefits minus the total costs. This willingness-to-pay threshold can therefore be thought of as the conversion factor between benefits (in units of QALYs or DALYs) to dollars; since there is no agreed-upon value for this conversion factor, the NMB can be calculated for a variety of willingness-to-pay thresholds. This can be useful for succinctly displaying uncertainty in total costs and QALYs; often probabilistic sensitivity analyses will report the probability of a strategy having the highest net monetary benefit for as a measure of how cost-effective it is.

While there is not a set number of microsimulation runs required for every simulation, the number of simulation runs should be large enough that the simulation outcomes are not obscured by Monte Carlo noise. This may vary by the number of individuals in the microsimulation—a larger number of simulated individuals may generate more robust outcome estimates as each individual's outcomes are averaged over a larger population. In the case of this model, the population size was large (6.5 million in the case of the cost-effectiveness results), and only 10 runs were needed to reduce Monte Carlo noise to levels where the results were clear (see error bars on Figure 1.5).

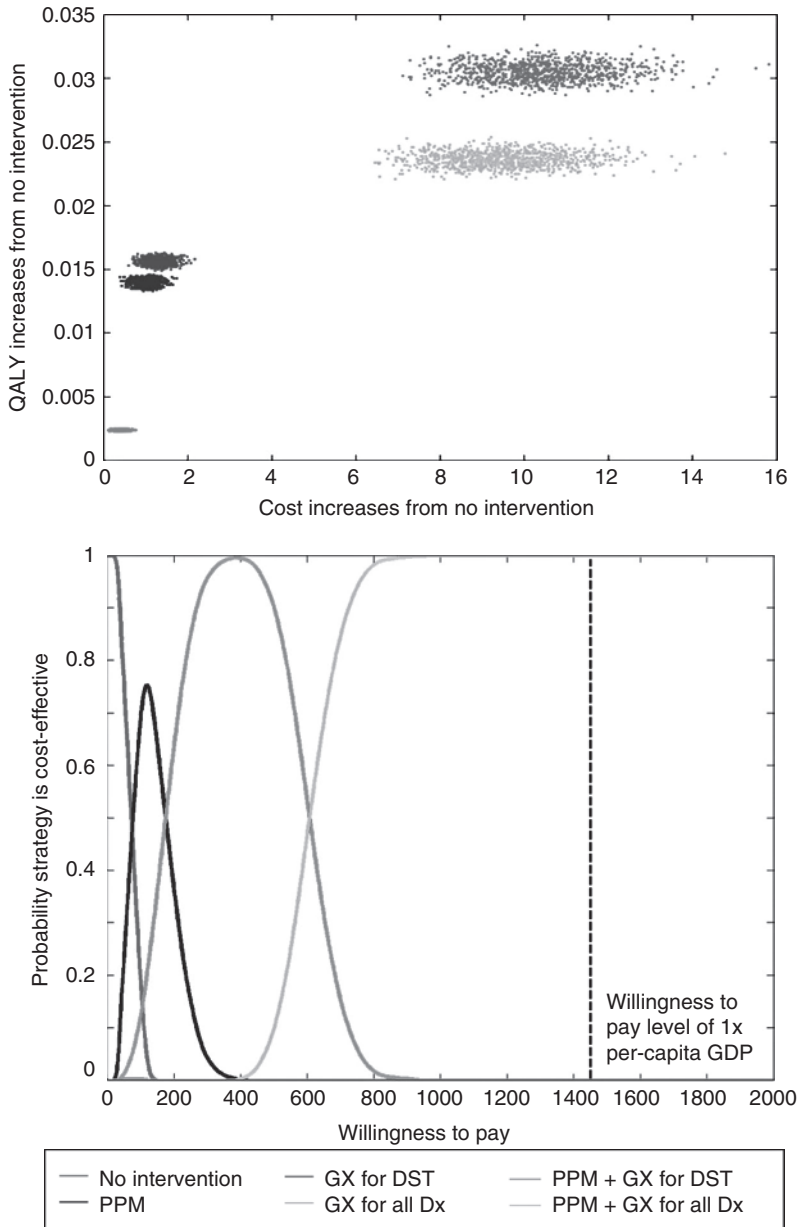


Figure 1.5 Probabilistic sensitivity analysis of healthcare costs and TB-related quality of life. At a willingness-to-pay threshold of India’s per capita GDP (\$1450), PPM with GeneXpert for all diagnosis is most likely cost-effective. Dx, diagnosis; GX, GeneXpert (Suen et al. 2015). Source: Reprinted with permission of the International Union Against Tuberculosis and Lung Disease. Copyright The Union.

1.3.10 Sensitivity Analysis

It is important to note that a microsimulation will be subject to sampling uncertainty or Monte Carlo noise if the simulated population is not large enough, and this uncertainty will further increase the uncertainty of cost and QALY outcomes of the modeled interventions. This type of uncertainty is due to the model itself and can therefore be reduced by increasing the number of individuals in the simulation or running additional simulation runs and averaging across runs to form an estimate of the average costs and benefits for each intervention. Estimates of the uncertainty of the mean costs and benefits can then be formed using a bootstrapping method over the multiple simulation runs—because cost and benefits cannot be assumed to be Gaussian, the modeler can empirically determine the distribution of mean costs and benefits by sampling from the simulation runs (with replacement) and calculating the average cost and benefits for each sample. Multiple resamples will generate a distribution of the costs and benefits, allowing the modeler to observe the distribution of average costs and benefits and quantify 95% confidence intervals (or other distribution characteristics) to describe the uncertainty due to Monte Carlo noise.

In contrast, parameter uncertainty reflects uncertainty in the outputs due to uncertainty in the input parameters and reflects unknowns about the true state of the world. The only way to reduce this type of uncertainty is to gather more information (incorporate additional study from the literature, run more clinical trials, interview more patients, etc.) and incorporate that information into the model. Accurate portrayal of this type of uncertainty in the model outputs allows decision makers to make informed decisions about how confident they should be in the model results.

We illustrate this with the example of uncertainty around the costs and QALY inputs. In Figure 1.5, the simulation has been run multiple times, each time sampling a costs and QALY input from their distributions (instead of treating them as point estimates, as in Figure 1.4). Each run then generates a different output, and this is reflected in the dispersion of points for each intervention—each point in Figure 1.5 panel a represents one simulation run. The results can be presented in a more compact form (panel b), which gives the probability the strategy is cost-effective over different willingness-to-pay levels. In this example, the strategy PPM+GeneXpert for all diagnosis is likely to be cost-effective even at very cost-effective willingness-to-pay levels (one GDP per capita).

1.4 CONCLUSION

Microsimulation can be a useful tool for evaluating the optimality of disease control policies within a set of intervention options, particularly for infectious diseases, where transmission dynamics and individual behavior may make disease dynamics difficult to represent analytically or in compartmental simulations. Microsimulations are flexible and can accommodate a high level of stratification, which may be important when modeling a disease with a complex natural history that is influenced by demographic and

behavioral factors. However, more stratifications make the model more complex, and joint distributions of risks may be unknown or uncertain. These aspects of the model may make communication of the analysis more difficult. Even so, microsimulations are an important tool for modeling infectious diseases, particularly when individual level or subgroup level analyses are desired. Once completed, they can be easily modified to include additional subgroups or treatment policies, provided the transition probability data is available. The output of these analyses can help policy makers assess the impact of different treatment programs on select subpopulations and help articulate the impact of difficult-to-predict behavioral changes on treatment effects. Used effectively, they may be an important tool in the fight against infectious disease.

REFERENCES

- C. Acuna-Villaorduna et al., “Cost-effectiveness analysis of introduction of rapid, alternative methods to identify multidrug-resistant tuberculosis in middle-income countries,” *Clin. Infect. Dis.*, vol. 47, no. 4, pp. 487–495, 2008.
- Central TB Division, Directorate General of Health Services, and Ministry of Health and Family Welfare, “RNTCP performance report, India: Third quarter, 2010,” New Delhi, India, 2010.
- Global Health Education, “TB tests: Skin test, sputum & other types of TB test serological tests for TB,” 2014. [Online]. Available: <https://www.tbfacts.org/tb-tests/>. Accessed on October 2014.
- P. Horby et al., “Social contact patterns in Vietnam and implications for the control of infectious diseases,” *PLoS One*, vol. 6, no. 2, p. e16965, 2011.
- R. Horsburgh, “Priorities for the treatment of latent tuberculosis infection in the United States,” *N. Engl. J. Med.*, vol. 350, no. 20, pp. 2060–2067, 2004.
- R. Hutubessy, D. Chisholm, and TT. Edejer, “Generalized cost-effectiveness analysis for national-level priority-setting in the health sector,” *Cost Eff. Resour. Alloc.*, vol. 1, no. (1), p. 8, 2003.
- S. K. Kapoor, A. V. Raman, K. S. Sachdeva, and S. Satyanarayana, “How did the TB patients reach DOTS services in Delhi? A study of patient treatment seeking behavior,” *PLoS One*, vol. 7, no. 8, p. e42458, 2012.
- W. Kermack and A. McKendrick, “A contribution to the mathematical theory of epidemics,” *Proc. R. Soc. Lond., Ser. A*, vol. 115, no. 772, pp. 700–721, 1927.
- H.-H. Lin, M. Ezzati, and M. Murray, “Tobacco smoke, indoor air pollution and tuberculosis: A systematic review and meta-analysis,” *PLoS Med.*, vol. 4, no. 1, p. e20, 2007.
- J. S. Michael and T. J. John, “Extensively drug-resistant tuberculosis in India: A review,” *Indian J. Med. Res.*, vol. 136, no. 4, pp. 599–604, 2012.
- J. Mossong et al., “Social contacts and mixing patterns relevant to the spread of infectious diseases,” *PLoS Med.*, vol. 5, no. 3, p. e74, 2008.
- A. Pantoja et al., “Economic evaluation of public-private mix for tuberculosis care and control, India. Part I. Socio-economic profile and costs among tuberculosis patients,” *Int. J. Tuberc. Lung Dis.*, vol. 13, no. 6, pp. 698–704, 2009.
- R. Rajeswari, V. Chandrasekaran, M. Suhadev, S. Sivasubramaniam, G. Sudha, and G. Renu, “Factors associated with patient and health system delays in the diagnosis of tuberculosis in South India,” *Int. J. Tuberc. Lung Dis.*, vol. 6, no. 9, pp. 789–795, 2002.

- Revised National Tuberculosis Control Programme (RNTCP), "Training module for medical practitioners," New Delhi, India, 2010.
- J. M. Selvam et al., "Health-seeking behaviour of new smear-positive TB patients under a DOTS programme in Tamil Nadu, India, 2003," *Int. J. Tuberc. Lung Dis.*, vol. 11, no. 2, pp. 161–167, 2007.
- S. Suen, E. Bendavid, and J. D. Goldhaber-Fiebert, "Disease control implications of India's changing multi-drug resistant tuberculosis epidemic," *PLoS One*, vol. 9, no. 3, p. e89822, 2014.
- S. Suen, E. Bendavid, and J. Goldhaber-Fiebert, "Cost-effectiveness of improvements in diagnosis and treatment accessibility for tuberculosis control in India," *Int. J. Tuberc. Lung Dis.*, vol. 19, no. (9), pp. 1115–1124, 2015.
- E. W. Tiemersma, M. J. van der Werf, M. W. Borgdorff, B. G. Williams, and N. J. D. Nagelkerke, "Natural history of tuberculosis: Duration and fatality of untreated pulmonary tuberculosis in HIV negative patients: A systematic review," *PLoS One*, vol. 6, no. 4, p. e17601, 2011.
- K. Tobgay, K. J. Sarma, and P. S. Thankappan, "Predictors of treatment delays for tuberculosis in Sikkim," *Natl. Med. J. India*, vol. 19e, no. 2, pp. 60–63, 2006.
- Z. F. Tu, L. M. Pinto, and M. W. Uplekar, "Tuberculosis management by private practitioners in Mumbai, India: Has anything changed in two decades?" *PLoS One*, vol. 5, no. 8, p. e12023, 2010.
- M. W. Uplekar and D. S. Shepard, "Treatment of tuberculosis by private general practitioners in India," *Tubercle*, vol. 72, no. 4, pp. 284–290, 1991.
- N. Vandan, M. Ali, R. Prasad, and C. Kuroiwa, "Assessment of doctors' knowledge regarding tuberculosis management in Lucknow, India: A public-private sector comparison," *Public Health*, vol. 123, no. 7, pp. 484–489, 2009.
- E. Vynnycky and P. E. Fine, "The natural history of tuberculosis: The implications of age-dependent risks of disease and the role of reinfection," *Epidemiol. Infect.*, vol. 119, no. 2, pp. 183–201, 1997.
- M. C. Weinstein, G. Torrance, and A. McGuire, "QALYs: The basics," *Value Health*, vol. 12 no. Suppl 1, pp. S5–S9, 2009.
- D. E. Winetsky et al., "Screening and rapid molecular diagnosis of tuberculosis in prisons in Russia and Eastern Europe: A cost-effectiveness analysis," *PLoS Med.*, vol. 9, no. 11, p. e1001348, 2012.
- World Health Organization, "WHO lifetables for India," Geneva, Switzerland, 2010a.
- World Health Organization, "Multidrug and extensively drug-resistant TB (M/XDR-TB): 2010 global report on surveillance and response," Geneva, Switzerland, 2010b.
- World Health Organization, "Collaborative framework for care and control of tuberculosis and diabetes," Geneva, Switzerland, 2011.
- World Health Organization, "Global tuberculosis report 2013," Geneva, Switzerland, 2013.
- World Health Organization, "Health statistics and information systems metrics: Disability-adjusted life year," 2014a. [Online]. Available: http://www.who.int/healthinfo/global_burden_disease/metrics_daly/en/. Accessed on August 10, 2017.
- World Health Organization, "Global tuberculosis report 2015," Geneva, Switzerland, 2015.
- A. Zumla, M. Raviglione, R. Hafner, and C. F. von Reyn, "Tuberculosis," *N. Engl. J. Med.*, vol. 368, no. 8, pp. 745–755, 2013.