JOURNAL OF THE ROYAL SOCIETY Interface

Tuberculosis reinfection rate as a proportion of total infection rate correlates with the logarithm of the incidence rate: a mathematical model

Pieter W Uys, Paul D van Helden and John W Hargrove

Supplementary data	"Data Supplement" http://rsif.royalsocietypublishing.org/content/suppl/2009/02/06/6.30.11.DC1.html
References	This article cites 18 articles, 8 of which can be accessed free http://rsif.royalsocietypublishing.org/content/6/30/11.full.html#ref-list-1
Subject collections	Articles on similar topics can be found in the following collections biomathematics (38 articles)
Email alerting service	Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click **here**

To subscribe to *J. R. Soc. Interface* go to: **http://rsif.royalsocietypublishing.org/subscriptions**

Tuberculosis reinfection rate as a proportion of total infection rate correlates with the logarithm of the incidence rate: a mathematical model

Pieter W. Uys[1,2,*], Paul D. van Helden[1] and John W. Hargrove[2]

[1]*DST/NRF Centre of Excellence for Excellence in Epidemiological Modelling and Analysis (SACEMA), Stellenbosch University, Stellenbosch, Western Cape 7602, Republic of South Africa*
[2]*Division of Molecular Biology and Human Genetics, MRC Center for Molecular and Cellular Biology, DST/NRF Centre of Excellence for Biomedical TB Research, Faculty of Health Sciences, Stellenbosch University, PO Box 19063, Tygerberg 7505, Republic of South Africa*

In a significant number of instances, an episode of tuberculosis can be attributed to a reinfection event. Because reinfection is more likely in high incidence regions than in regions of low incidence, more tuberculosis (TB) cases due to reinfection could be expected in high-incidence regions than in low-incidence regions. Empirical data from regions with various incidence rates appear to confirm the conjecture that, in fact, the incidence rate due to reinfection only, as a proportion of all cases, correlates with the logarithm of the incidence rate, rather than with the incidence rate itself. A theoretical model that supports this conjecture is presented. A Markov model was used to obtain a relationship between incidence and reinfection rates. It was assumed in this model that the rate of reinfection is a multiple, ρ (the reinfection factor), of the rate of first-time infection, λ. The results obtained show a relationship between the proportion of cases due to reinfection and the rate of incidence that is approximately logarithmic for a range of values of the incidence rate typical of those observed in communities across the globe. A value of ρ is determined such that the relationship between the proportion of cases due to reinfection and the logarithm of the incidence rate closely correlates with empirical data. From a purely theoretical investigation, it is shown that a simple relationship can be expected between the logarithm of the incidence rates and the proportions of cases due to reinfection after a prior episode of TB. This relationship is sustained by a rate of reinfection that is higher than the rate of first-time infection and this latter consideration underscores the great importance of monitoring recovered TB cases for repeat disease episodes, especially in regions where TB incidence is high. Awareness of this may assist in attempts to control the epidemic.

Keywords: *Mycobacterium tuberculosis*; reinfection; incidence; Markov model; transmission; susceptibility

1. INTRODUCTION

The incidence of tuberculosis (TB) found in various regions of the globe ranges from almost zero to more than 600 cases per 100 000 persons per annum (WHO Report 2006). Incidence double this rate has also been reported for a few communities (WHO Report 2006). The possibility of repeated infections leading to further disease episodes is real and contributes to such high rates. Indeed, mixed infections in a large number of individuals of a population in a high-incidence region have been reported (Richardson *et al.* 2002; Behr 2004; Warren *et al.* 2004) and the phenomenon of reinfection resulting in second episodes of disease in patients has been documented (van Rie *et al.* 1999). In a Dutch setting with a low tuberculosis incidence, approximately one in six new disease episodes among patients with previous tuberculosis infection or disease may be attributable to recent reinfection (de Boer *et al.* 2003).

There appears to be a correlation between the proportion of cases that can be attributed to reinfection and the logarithm of the incidence of TB (Wang *et al.* 2007). Various factors could influence this correlation.

*Author and address for correspondence: Division of Molecular Biology and Human Genetics, Faculty of Health Sciences, Stellenbosch University, PO Box 19063, Tygerberg 7505, Republic of South Africa (edserve@iafrica.com).

The issue of whether or not a first episode of TB imparts some measure of immunity to further TB infections has not been unequivocally established, but may have an impact (Warren *et al.* 2004; Verver *et al.* 2005; Nagelkerke *et al.* 2006). On the other hand, a potential confounding effect is that a patient experiencing a second episode of TB may simply be demonstrating that he/she has a higher susceptibility, either innate or due to environmental conditions (Hoal 2002; Hanekom *et al.* 2007). It is also possible that an episode of TB actually renders the patient more susceptible to infection or likely to become ill. All these issues are surveyed in Verver *et al.* (2005).

Reinfection is thus an important event with largely unknown consequences; for example, it can be speculated that multiple exposures may even precipitate progression to disease. The influence of incidence on the reinfection rate thus appears to be a fundamental issue (Wang *et al.* 2007; van Helden *et al.* 2008), and to investigate this we constructed a Markov model that simulates the epidemiology of a TB endemic in a hypothetical region. This Markov model paradigm assumes that the population size is constant but we regard this to be a reasonable assumption, given a short period estimation. Immigration and emigration must also be excluded. However, these restrictions can be relaxed to a certain extent without seriously undermining the deductions made from this model. Our model was intended to assist in investigating reported data (Wang *et al.* 2007). Unfortunately, the HIV status of the patients involved in this study was not reported. For internal consistency across different incidence rates within our model we elected to exclude HIV considerations. However, assuming a constant proportion of HIV$^+$ among the TB-infected cases, their innate higher rate of progression to disease can be reflected by a higher rate of progression generally. This does not, of course, yield an accurate indication when comparing data from different settings subject to large disparities in HIV incidence. Deviations from model predictions could thus be anticipated in any community where there is a significantly increased presence of HIV or sudden population change.

Drug-resistant TB as a separate or specific entity is not considered in the model, which we do not think is a serious shortcoming because the incidence of drug-resistant TB is relatively low compared with susceptible TB and the basic principles are the same. Drug-resistant TB can be regarded as a component of the TB epidemic.

2. METHODS

2.1. The compartmental model and its representation as a Markov chain

A basic compartmental model was constructed (figure 1). This is essentially the model studied by Vynnycky & Fine (1997) except that the dependency on age has been removed. This, however, does not undermine the validity of our model since we shall be concerned with long-term steady states and the assumed condition of constant population size. This implies that each compartment maintains a constant

distribution with respect to the ages of its members and that therefore the effect of age on the transition probabilities is absent. The assumptions concerning fixed population size and the absence of emigration or immigration also mean that the number of births in a given time interval must be approximately the total number of deaths during the same time. This allows the compartmental model to be interpreted as a Markov chain. An iteration time period must be specified: six months is deemed suitable in this model as this equals the duration of most of the therapy protocols. This is also the typical mean delay to diagnosis (Asch *et al.* 1998; Sherman *et al.* 1999) and hence the time spent by a member in the compartment of infectious cases, P. Thus, all the people in the P compartment move out (commence therapy or die) at a simulation step and, to maintain the constant compartment sizes, they must be replaced by infected people who have now progressed to disease. In 1 year, this transition occurs twice so that the disease incidence is given by $2 \cdot p$ (where p is the number of people in the population with active disease).

The Markov chain representation of the compartmental model implies a matrix T, called the transition matrix, for which the entries are the rates of flow from one compartment to another. These correspond, among others, to the rates of mortality, infection or reinfection, progression to disease (endogenous and exogenous) and transition into the latent class (see the electronic supplementary material).

Using standard matrix notation, if at some time the distribution of the population is given by the row vector $\boldsymbol{u}^1$ then, one time step later, the distribution vector is given by $\boldsymbol{u}^2 = \boldsymbol{u}^1 T$. After many consecutive time steps, it may be the case that a steady state, $\boldsymbol{v}$, is achieved, where the relative proportions of the population in the various compartments no longer undergo further changes. This state $\boldsymbol{v}$ is then given by $\boldsymbol{v} = \lim \boldsymbol{u}^1 T^n$, and generally a matrix, T^∞, called the long-term transition matrix, can be calculated so that $\boldsymbol{v} = \boldsymbol{u}^1 T^\infty$ (see the electronic supplementary material).

This $\boldsymbol{v}$ corresponds exactly to the determination of the constant compartment sizes found in the long-term steady state and has the form,

$$b \quad m \quad s \quad i_m \quad i \quad p \quad l_p \quad l_s \quad e_n \quad i^* \quad e_x.$$

The values of the elements of $\boldsymbol{v}$ are the proportions of the population in each of the compartments (see the electronic supplementary material). For convenience, the notation here for the various proportions relate to the labels of the respective compartments. Thus, $\boldsymbol{v}$ is an appropriate representation of the epidemiological conditions prevalent in a region where TB has been endemic for an extended period and such that the compartment sizes have achieved approximate stability. We refer to such conditions as stationary epidemiological conditions.

2.2. Finding the proportion of disease cases that are due to reinfection

The number of susceptible people who become infected for the first time, per annum, is given by $\lambda \cdot s$, where λ is the force of infection, also referred to as the annual
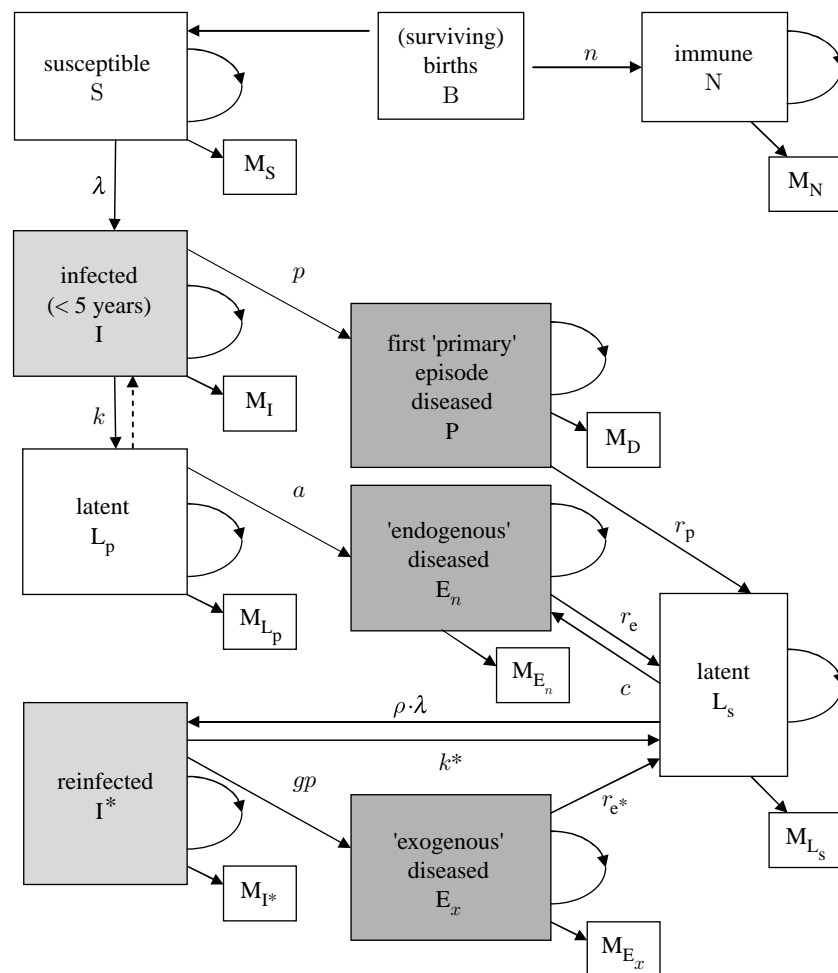
Figure 1. Markov chain flowchart. The number of births equals the total number of mortalities that are represented by boxes labelled with 'M'. The loops represent members of the compartments, which remain in their respective compartments during a time step. The probability k for the transition I–L_P takes into account the counter-transition represented by the dotted arrow and this counter transition, therefore, needs not be explicitly included in the model (see the electronic supplementary material for further discussion of infection rates).

risk of infection, and s is the number of susceptibles. Thus, the transition probability from S to I is λ. Similarly, the transition probability from L_s, the first-time latent compartment to i^*, the reinfected compartment, is given by $\rho \cdot \lambda$, where ρ is the factor by which the reinfection rate for people who have recovered from a TB disease episode differs from the first-time infection rate. The factor ρ will be referred to as the reinfection factor.

The steady-state distribution vector $\boldsymbol{v} =$

$$\boldsymbol{b} \quad \boldsymbol{m} \quad \boldsymbol{s} \quad \boldsymbol{i_m} \quad \boldsymbol{i} \quad \boldsymbol{p} \quad \boldsymbol{l_p} \quad \boldsymbol{l_s} \quad \boldsymbol{e_n} \quad \boldsymbol{i^*} \quad \boldsymbol{e_x}$$

can be evaluated for any chosen input value of λ, the force of infection (see table 1 in the electronic supplementary material) and, typically, simulations are performed with a range of selected values of λ.

We require the proportion of currently infected people who have had a previous disease episode and who have recovered compared with the total number of infected (including reinfected) people. This proportion is given by $\boldsymbol{i^*}/(\boldsymbol{i} + \boldsymbol{i^*})$ and is to be compared with the incidence of disease for various rates of incidence.

To this end, the transition matrix was constructed for a range of values for λ, the force of infection. For each instance of T, the steady-state transition matrix T^∞ was computed and from this $\boldsymbol{v}$ was obtained. The vector $\boldsymbol{v}$ has among its entries the incidence, infection and reinfection rates. (Thus, λ is an *input* parameter and incidence, infection and reinfection rates are *output* values.) This enabled the compilation of a collection of incidence rates together with their matching reinfection proportion percentages. These data points can then be plotted (figures 2–4). These graphs can be prepared for a variety of values of ρ, the reinfection factor (figures 2 and 3) and p, the rate of progression to infectious disease (see the electronic supplementary material) as well as the mortality rate μ (see the electronic supplementary material).

3. RESULTS

If the rate of infection for previous and cured cases is assumed to be less than that for new cases, then the graph of the proportion of cases due to reinfection versus the incidence rate is approximately linear and it does not show the high proportions of reinfection cases that have been reported (van Rie *et al.* 1999; Verver *et al.* 2005; figures 2 and 3). This is the case
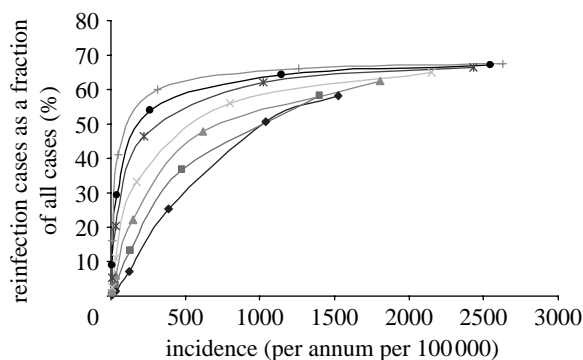
Figure 2. The reinfection rate as a function of incidence with $p=0.05$ and various values for the reinfection factor ρ (diamonds, 0.25; squares, 0.5; triangles, 1; crosses, 2; asterisks, 4.5; circles, 8; pluses, 16).
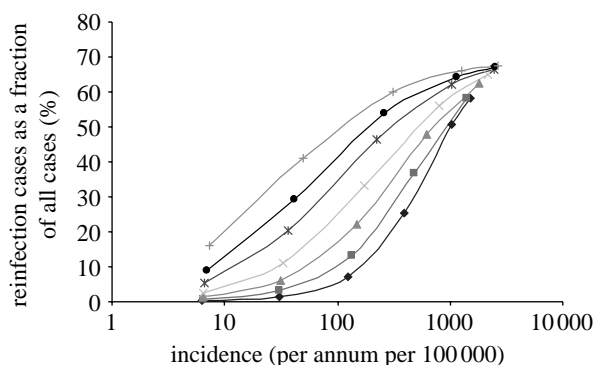

Figure 3. The reinfection rate as a function of incidence with $p=0.05$ and various values for the reinfection factor ρ (diamonds, 0.25; squares, 0.5; triangles, 1; crosses, 2; asterisks, 4.5; circles, 8; pluses, 16) plotted with log scale.
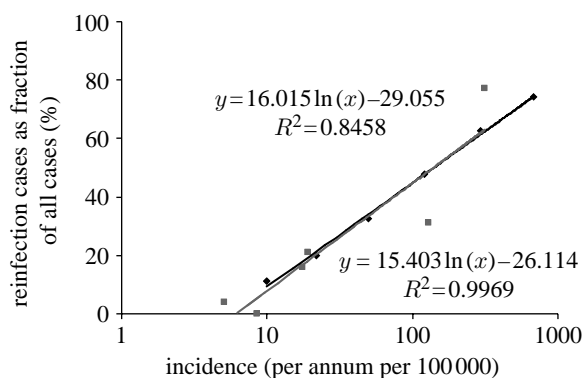

Figure 4. The reinfection rate as a function of incidence. A logarithmic fit to the actual data points (Wang *et al.* 2007) and the best theoretical fit as generated by the model are shown. A 10% CI for slope of line is 16.02 ± 7.29, and this is hence a 10% CI for the slope of the required best-fitting theoretical line. From this, a 10% CI for the ρ values for the theoretical lines corresponding to this range is found to be 7 ± 4. Diamonds, theoretical; squares, data; black line, log (theoretical); grey line, log (data).

irrespective of the value used for p, the rate of progression to infectious disease, and remains so even if differential rates of progression to disease are applied. The qualitative nature of the results does change, however, when differential rates of infection are applied, namely when the rate of infection for recovered cases is assumed to be greater by some factor (the

reinfection factor, ρ) than that for people who have not previously had symptoms of active TB disease. If ρ has a value in the range from 2 to 16, the graph does not have a linear trend but rather has a logarithmic appearance (figure 2).

The graph that best fits the empirical data is obtained when ρ has the value of 7 (figure 4). A 10% confidence interval (CI) for slope of the linear regression line fitted to the data is 16.02 ± 7.29. This in turn leads to a 10% CI 7 ± 4 for the value of ρ. It is striking that a value of 7 has been estimated from empirical data elsewhere (Verver *et al.* 2005).

4. DISCUSSION

Reinfection (van Rie *et al.* 1999) or multiple infection (Richardson *et al.* 2002) has been shown to be an important phenomenon in TB dynamics and occurs in both high- and low-incidence societies (Richardson *et al.* 2002; Lambert *et al.* 2003; Behr 2004; Chiang & Riley 2005; Rodrigues *et al.* 2007). Intuitively, one may expect that reinfection would be lower in low-incidence (e.g. developed) regions, which appears to be the trend seen in publications (García de Viedma *et al.* 2002; Wang *et al.* 2007). In this paper, we propose a quantitative model for reinfection, which demonstrates a log-law relationship with incidence.

The results discussed so far arise from simulations where the intercompartmental dynamics are driven exclusively by transition probabilities, and, except for the rates of infection, these are the same for all simulations with a specified value for p, the rate of progression to active disease.

In reality, conditions are likely to be different from setting to setting. In developing countries, compared with developed countries, mortality and birth rates will be higher. Lower socio-economic living conditions common in developing countries will likely result in higher incidence rates with greater opportunity for reinfection and this is especially so where HIV occurs at high prevalence. Actual data from communities experiencing a high incidence could therefore logically be expected to show higher reinfection rates than those due to simple constant probabilities. The reverse applies to communities where the incidence is low. For such communities, the possibility of transmission and the rate of progression to disease will be diminished on account of the population comprising people with immune systems that have not been weakened by malnutrition and living conditions. The likelihood of reinfection leading to active disease will be smaller owing to better health care, contact tracing and monitoring of people who have recovered from an earlier TB disease episode. These considerations are countered by the observation that the higher mortality rates expected in high-incidence regions result in depressed reinfection rates (see figure 2 in the electronic supplementary material).

The empirical observation of a positive correlation between the logarithm of the incidence and reinfection proportion percentage thus seems to have theoretical support and suggests that the reinfection factor is of the order of 7. This indicates that the risk of a second,

disease-causing infection for a person who has recovered from an episode of disease is approximately seven ($\pm$4) times greater than the risk of a first-time infection that leads to disease. This estimate correlates well with the actual minimum estimates made from data from a high TB incidence community (Verver *et al.* 2005), where the risk of developing a second episode of TB after infection was estimated to be four to seven times higher than a first episode.

It is evident then that recovered TB cases need follow-up monitoring particularly in communities where the incidence of TB is high. In view of this elevated risk, we suggest that TB cases require regular follow-up, and that this should be more intense for the first 3 years after an episode, but that, even thereafter, follow-up is indicated.

REFERENCES

Asch, S., Leake, B., Anderson, R. & Gelberg, L. 1998 Why do symptomatic patients delay obtaining care for tuberculosis? *Am. J. Respir. Crit. Care Med.* **157**(Pt 1), 1244–1248.

Behr, M. A. 2004 Tuberculosis due to multiple strains: a concern for the patient? A concern for tuberculosis control? *Am. J. Respir. Crit. Care Med.* **169**, 554–555. (doi:10.1164/rccm.2401001)

Chiang, C. Y. & Riley, L. W. 2005 Exogenous reinfection in tuberculosis. *Lancet Infect. Dis.* **5**, 629–636. (doi:10.1016/S1473-3099(05)70240-1)

de Boer, A. S., Borgdorff, M. W., Vynnycky, E., Sebek, M. M. & van Soolingen, D. 2003 Exogenous re-infection as a cause of recurrent tuberculosis in a low-incidence area. *Int. J. Tuberc. Lung Dis.* **7**, 145–152.

García de Viedma, D., Marín, M., Hernangómez, S., Díaz, M., Ruiz Serrano, M. J., Alcalá, L. & Bouza, E. 2002 Tuberculosis recurrences: reinfection plays a role in a population whose clinical/epidemiological characteristics do not favor reinfection. *Arch Int. Med.* **162**, 1873–1879. (doi:10.1001/archinte.162.16.1873)

Hanekom, M., van der Spuy, G. D., van Pittius, N. C., McEvoy, C. R., Ndabambi, S. L., Victor, T. C., Hoal, E. G., van Helden, P. D. & Warren, R. M. 2007 Evidence that the spread of *Mycobacterium tuberculosis* strains with the Beijing genotype is human population dependent. *J. Clin. Microbiol.* **45**, 2263–2266. (doi:10.1128/JCM.02354-06)

Hoal, E. G. 2002 Human genetic susceptibility to tuberculosis and other mycobacterial diseases. *IUBMB. Life* **53**, 225–229. (doi:10.1080/15216540212644)

Lambert, M. L., Hasker, E., Van Deun, A., Roberfroid, D., Boelaert, M. & Van der Stuyft, P. 2003 Recurrence in tuberculosis: relapse or reinfection? *Lancet Infect. Dis.* **3**, 282–287. (doi:10.1016/S1473-3099(03)00607-8)

Nagelkerke, N. J. D., de Vlas, S., Mahendradhata, Y., Ottenhoff, T. H. M. & Borgdorff, M. W. 2006 The search for a tuberculosis vaccine: an elusive quest? *Tuberculosis* **86**, 41–46. (doi:10.1016/j.tube.2005.06.001)

Richardson, M. *et al.* 2002 Multiple *Mycobacterium tuberculosis* strains in early cultures from patients in a high-incidence community setting. *J. Clin. Microbiol.* **40**, 2750–2754. (doi:10.1128/JCM.40.8.2750-2754.2002)

Rodrigues, P., Gomes, M. G. & Rebelo, C. 2007 Drug resistance in tuberculosis—a reinfection model. *Theor. Popul. Biol.* **71**, 196–212. (doi:10.1016/j.tpb.2006.10.004)

Sherman, L. F., Fujiwara, P. I., Cook, S. V., Bazerman, L. B. & Frieden, T. R. 1999 Patient and health care system delays in the diagnosis and treatment of tuberculosis. *Int. J. Tuberc. Lung Dis.* **3**, 1088–1095.

van Helden, P. D., Warren, R. M. & Uys, P. W. 2008 Predicting reinfection in TB. *J. Infect. Dis.* **197**, 172–173. (doi:10.1086/523829)

van Rie, A., Warren, R., Richardson, M., Victor, T. C., Gie, R. P., Enarson, D. A., Beyers, N. & van Helden, P. D. 1999 Exogenous reinfection as a cause of recurrent tuberculosis after curative treatment. *N. Engl. J. Med.* **341**, 1174–1179. (doi:10.1056/NEJM199910143411602)

Verver, S., Warren, R. M., Beyers, N., Richardson, M., van der Spuy, G. D., Borgdorff, M. W., Enarson, D. A., Behr, M. A. & van Helden, P. D. 2005 Rate of reinfection tuberculosis after successful treatment is higher than rate of new tuberculosis. *Am. J. Respir. Crit. Care Med.* **171**, 1430–1435. (doi:10.1164/rccm.200409-1200OC)

Vynnycky, E. & Fine, P. E. 1997 The natural history of tuberculosis: the implications of age-dependent risks of disease and the role of reinfection. *Epidemiol. Infect.* **119**, 183–201. (doi:10.1017/S0950268897007917)

Wang, J. Y., Lee, L. N., Lai, H. C., Hsu, H. L., Liaw, Y. S., Hsueh, P. R. & Yang, P. C. 2007 Prediction of the tuberculosis reinfection proportion from the local incidence. *J. Infect. Dis.* **196**, 281–288. (doi:10.1086/518898)

Warren, R. M., Victor, T. C., Streicher, E. M., Richardson, M., Beyers, N., Gey van Pittius, N. C. & van Helden, P. D. 2004 Patients with active tuberculosis often have different strains in the same sputum specimen. *Am. J. Respir. Crit. Care Med.* **169**, 610–614. (doi:10.1164/rccm.200305-714OC)

WHO. WHO Report 2006 *Global tuberculosis control: surveillance, planning, financing.* Geneva, Switzerland: World Health Organization.