Progression from latent infection to active disease in dynamic TB transmission models: a systematic review of the validity of modelling assumptions.

Authors

Nicolas A Menzies PhD[1,2,] Emory Wolf BSc[1], David Connors BA[1], Meghan Bellerose BA[1], Alyssa N Sbarra BS[3], Ted Cohen DPH[3], Andrew N Hill PhD[4], Reza Yaesoubi PhD[3], Kara Galer MPH[1], Peter J White PhD[5,6], Prof Ibrahim Abubakar PhD[7], Prof Joshua A Salomon PhD[1]

Affiliations

1. Department of Global Health and Population, Harvard T.H. Chan School of Public Health, Boston, USA

2. Center for Health Decision Science, Harvard T.H. Chan School of Public Health, Boston, USA.

3. Department of Epidemiology of Microbial Diseases, Yale School of Public Health, New Haven, USA.

4. Division of TB Elimination, U.S. Centers for Disease Control and Prevention, Atlanta, USA.

5. MRC Centre for Outbreak Analysis and Modelling and NIHR Health Protection Research Unit in Modelling Methodology, Imperial College London, London, UK

6. Modelling and Economics Unit, National Infection Service, Public Health England, London, United Kingdom.

7. Institute for Global Health, University College London, London, United Kingdom.

Corresponding author

Nicolas A Menzies, nmenzies@hsph.harvard.edu, (+1) 617 432 0492, Department of Global Health and Population, Harvard T.H. Chan School of Public Health, Boston, USA.

Manuscript word count: 5070

Disclaimer

The findings and conclusions in this paper are those of the author(s) and do not necessarily represent the views of the US Centers for Disease Control and Prevention, the UK Department of Health, MRC, National Health Service, NIHR, Public Health England, or the authors' other affiliated institutions.

Abstract

Mathematical modelling is commonly used to evaluate infectious disease control policy, and is influential in shaping policy and budgets. Mathematical models necessarily make assumptions about disease natural history, and if these assumptions are not valid the results of these studies may be biased. We conducted a systematic review of published TB transmission models, to assess the validity of assumptions about progression to active disease following initial infection (PROSPERO ID CRD42016030009). We searched PubMed, Web of Science, Embase, Biosis, and Cochrane Library, and included studies from the earliest available date (1962) to August 31st 2017. We identified 312 studies that met inclusion criteria. Predicted TB incidence varied widely across studies for each risk factor investigated. For population groups with no individual risk factors, annual incidence varied by several orders of magnitude, and 20-year cumulative incidence ranged from close to 0% to 100%. A substantial fraction of modelled results were inconsistent with empirical evidence—for 10-year cumulative incidence 40% of modelled results were more than double or less than half the empirical estimates. These results demonstrate substantial disagreement between modelling studies on a central feature of TB natural history. Greater attention to reproducing known features of TB epidemiology would strengthen future TB modelling studies, and readers of modelling studies are recommended to assess how well those studies demonstrate their validity.

Abstract word count: 218

Introduction

Latent infection is a defining feature of TB epidemiology. Upon infection with *Mycobacterium tuberculosis* (*M. tb*) approximately 5% of otherwise healthy adults will develop active disease within two years(1, 2) (so-called 'fast progressors'). Those who do not rapidly progress are classified as having 'slow progressing' latent TB infection (LTBI). With latent infection, individuals experience no adverse health effects and will not transmit *M. tb*, yet face an ongoing risk of developing active TB through reactivation. For individuals with long-established infection, the annual risk of active TB is low—empirical estimates are on the order of 10-20 per 100,000(3). However, due to high LTBI prevalence in many settings(4), reactivation TB can contribute a substantial fraction of incident TB cases, or even the majority in settings where transmission has been in sustained decline(5). The risk of active disease following infection also varies by individual characteristics, with infants(6), individuals with advanced HIV infection(7, 8), and other conditions affecting immune function(9-12) experiencing elevated progression risks.

As TB interventions can prevent transmission, they will generate benefits beyond the individuals receiving the intervention. In addition, the potential delay between infection and disease means that the consequences of improved TB control can be spread over many years. For these reasons, it is difficult for empirical TB policy evaluations to capture all impacts, and studies that forecast future disease trends or compare competing disease control policies commonly estimate results using dynamic transmission models. These models represent the mechanisms of transmission, natural history, and health system interactions that generate TB outcomes(13, 14). Despite over a century of epidemiological

TB research, concrete evidence on these underlying processes is imperfect(15), and studies have taken a variety of approaches for constructing and parameterizing transmission models. This variation can be consequential – in a modelling collaboration evaluating the post-2015 End TB Strategy(16), variation in epidemiological assumptions was identified as a cause of the wide range of estimates produced for the health impact(17) and cost-effectiveness(18) of expanded TB control. Several reviews have described standard TB modelling approaches(13, 14, 19), and methodological studies have examined specific modelling approaches(20-25). However, there has been little systematic investigation of assumptions made by published TB models. If these assumptions are not valid the results of these studies may be biased.

We undertook a systematic review of published studies employing dynamic TB transmission models, to assess the validity of assumptions about progression to active disease following initial infection (hereafter, "TB progression"). We describe how these studies modelled progression from initial infection to active disease, and the implications of these assumptions for predicted TB outcomes. We compare model predictions to empirical data(26-28) and discuss the implications for future modelling studies.

Methods

Inclusion/exclusion criteria

We included published studies using transmission dynamic models of tuberculosis in human populations to describe TB epidemiology or evaluate competing policy options. We excluded analyses where the force of infection was not modelled (ie were not transmission dynamic models). We excluded studies that provided insufficient information to describe (i) the model structure representing progression to active disease following initial infection, (ii) the

associated parameter values, and (iii) the population group(s) represented by the model, such that we could not reconstruct this part of the model. We excluded non-English language studies and unpublished reports. As one intent of this review was to describe the quality of assumptions made by modelling studies, we did not exclude studies based on quality criteria. The quality of studies was judged by their ability to reproduce empirical data, reported in the results. No additional quality assessment was undertaken. We followed PRISMA guidelines(29), and registered our protocol with PROSPERO (CRD42016030009).

Search strategy

We identified eligible studies by searching PubMed, Web of Science, Embase, Biosis, and Cochrane Library. We also searched a publication database compiled by the TB Modelling and Analysis Consortium(30), reference lists of eligible publications, several non-indexed journals, and the personal databases of the investigators to identify publications not included in the electronic search. Detailed search strategies are given in Table S1. We collected studies from the earliest available date (1962) to August 31st 2017.

Identification of studies

Titles and abstracts of collected studies were screened by one of two reviewers (EW, MB) to remove studies not meeting inclusion criteria, where this could be judged based on title and abstract alone (non-English language studies, non-transmission dynamic models). We retrieved the full texts for the remaining articles. Articles were assessed independently by two of five reviewers (AS, EW, DC, KG, and MB) to confirm they met inclusion criteria. Disagreements were resolved by discussion.

Extraction

For each study, we extracted bibliographic information as well as information on the study setting and how the model stratified the population by TB progression risk. For each of these model strata, we extracted data on model structure and parameter values describing TB progression. We also extracted the citations provided for parameter values. We did not extract information on TB progression risks following reinfection of previously exposed individuals, for whom risks of primary progressive TB are lower than for unexposed individuals(31).

We developed a typology of model structures and categorized models according to this typology (Figure 2). Where several different parameterizations were provided for the same population group, we used the values provided for the main analysis. Where a study provided a point estimate as well as upper and lower bounds we extracted the point estimate. Where a study only provided upper and lower bounds we took the arithmetic mean of these values. For each paper, extraction was undertaken independently by two of five reviewers (AS, EW, DC, KG, and MB). When extracted values differed between reviewers, the article was reviewed by an additional reviewer (NAM) and disagreements resolved through discussion.

[Figure 2]

Descriptive statistics

We calculated statistics to describe the distribution of studies according to publication year, setting, model structure, and population groups represented by model strata. We also identified the most commonly cited sources for model parameters.

Quantitative comparison of model predictions

We recreated the formulae of each model determining the risk of active TB for an individual initially infected with *M. tb*, matching the model structures shown in Figure 2. Using these formulae, and the parameter values extracted for each study and population group, we estimated the annual incidence of TB following initial infection in the absence of reinfection. For some studies this involved modifications to the original approach: (1) while some studies implemented their analyses by sampling progression parameters from a distribution, we used the point estimate (commonly the distribution mean) reported in the original paper. Even if the point estimate is equal to the mean of the parameter distribution this can produce small differences in simulation results, due to the non-linear relationship between parameters and modelled outcomes. (2) Some studies reported adjusting parameter values as part model calibration, yet did not report these adjusted values, and in this circumstance we used the original (unadjusted) values reported in the paper. (3) In some models individuals progress through multiple epidemiological or demographic processes simultaneously. If these processes influence TB progression or survival risks (e.g. aging, HIV progression), then accurately reproducing long-term cumulative incidence estimates is impossible without reconstructing all these different model components. As we only reconstructed the TB-specific parts of these models we do not report long-term cumulative incidence estimates in the presence of time-varying risk factors. (4) We did not allow for background mortality. While cumulative incidence estimates would be lower if background mortality were considered, this effect will be minor unless mortality rates are very high.

We stratified incidence predictions according to model structure, publication year, individual risk factors, study setting, and source of parameter assumptions. 'High-burden' settings

included countries on the WHO list of 30 high-TB burden countries(32), or, if a country was not specified, settings with incidence of 100 per 100,000 or higher. 'Low burden' settings included countries not on the WHO list of 30 high-TB burden countries or otherwise with incidence below 100 per 100,000. Studies with multiple HIV strata used a variety of approaches for describing HIV progression. 'Late HIV' was used for strata described as 'AIDS', 'WHO Stage 4 disease', 'advanced HIV', or with CD4 cell count $<200\mu L^{-1}$. 'Early HIV' was used for HIV strata not classified as 'Late HIV', in models with multiple HIV strata. We also distinguished model strata for HIV positive individuals receiving antiretroviral therapy ('HIV, on ART'). For age, we classified strata as 'Infant', if the midpoint of the age group fell in the range 0-2 years, and classified strata as 'Children (excl. infants)' if the midpoint of the age band fell in the range 2-10 years. We divided studies into those published before 2011 (the median publication year) and those published in 2011 and after, and according to whether the study cited any prior publications to justify parameter values for LTBI progression.

We plotted annual and cumulative incidence predictions to understand the behavior of each model, and summarized results as cumulative incidence at 2 and 20 years. These time points were chosen to represent rapid progression to active disease (primary progressive TB), and aggregate long-term risk respectively. For studies representing multiple population groups with different TB risk factors, we calculated risk ratios for TB incidence over the first two years, and for the twentieth year, to provide 'within study' comparisons of how risk factors were treated.

Comparison to empirical evidence

We reviewed the TB literature to identify studies reporting direct empirical evidence on TB progression risks following initial infection. To identify these studies we reviewed citations known to the investigators, studies cited in related reviews, and evidence cited in the studies included in the systematic review. As preventative treatment for LTBI reduces progression risks, the best evidence on TB natural history comes from historical studies conducted before preventive therapy became the standard of care for recently exposed individuals(33). Narrative reviews of these early studies have been compiled by Ferebee(1), Sutherland(2), and Styblo(34). From these reviews we extracted information on studies reporting quantitative estimates of annual risks of developing active TB following initial infection. Many of these studies had major limitations for estimating general population progression risks in the absence of reinfection, including small sample sizes, non-representative populations, settings that were likely to feature ongoing transmission, and non-specific TB diagnostics. For others the relevant features of study design, population and setting were not sufficiently described, and/or the original publication not available. Two studies provided precise estimates of TB progression risks in the years following initial infection, in both cases from the control arm of an intervention trial: (1) the British Medical Research Council's BCG trials(26, 27), which included 12,867 individuals in the unvaccinated study arm, and (2) the US Public Health Service's trials of isoniazid prophylaxis for TB household contacts(28), which included 12,594 individuals in the study control arm. Using summary data from these two studies we generated estimates of annual TB incidence for 10 years following TST conversion. We limited these comparisons to the first 10 years following infection to reduce the influence of attrition on the validity of empirical estimates. We compared these empirical estimates to model predictions for population groups with no individual risk factors affecting TB progression risk. All analyses were conducted in R version 3.3.2(35).

Replication data and analysis scripts available at https://dataverse.harvard.edu/dataverse/latent_tb_modelling_review.

The capacity of a model to fit the empirical estimates is determined by the model structure and the parameter values used. To separate these two factors we assessed whether each model structure was capable of reproducing the empirical results by adjusting the parameter values. To do so, we created a simple loss function using the results from the British Medical Research Council's BCG trials(26). This loss function represented the root mean squared error (RMSE) between model results and the empirical estimate for cumulative TB incidence over the first 10 years following infection. We used optimization algorithms (the "Nelder-Mead" and "BFGS" algorithms operationalized by the 'optim' function in R) to identify parameter values minimizing the loss function. We compared the predictions from these fitted models to the empirical estimates to understand the extent to which each model structure was capable of reproducing this evidence.

Results

Descriptive statistics on eligible studies

We identified 5,532 unique articles in the first stage of the review. We excluded 5,506 of these through title/abstract review, and a further 214 through full-text review. Three hundred and twelve studies met inclusion criteria and were included in the analysis (Figure 1). Table S2 lists included studies.

[Figure 1]

The earliest study included in the review was published in 1962, and 7% of studies were published before 2000. Of the 312 studies in the review, many included multiple strata to allow for differences in progression risk. A total of 680 observations were included in the analysis, where an observation represented an individual stratum within an included study. The majority of studies (62%) considered high burden settings, and 39% included model strata considering individual-level factors modifying TB progression. The most common risk factor considered by these studies was HIV (25%), followed by age (9%). Twelve different model structures were employed by these studies. Table 1 provides additional details about the features of included studies, and Figures S1 and S2 show the distribution of included studies by publication year, model structure, and subgroup.

[Table 1]

We identified the sources for TB progression parameters most commonly cited by the studies in the review. The three most commonly cited sources were Vynnycky and Fine 1997(36) (cited by 21% of all studies), Blower et al 1995(37) (12%), and Dye et al 1998(38) (10%). Each of these is a modelling paper included in our review. The top 15 most cited sources included a mix of modelling studies, empirical studies, and review articles (Figure S3). However, for 76 studies (24%) no citation was given for TB progression parameters.

Comparison of model predictions for population groups with no individual risk factors

We stratified model results by the population groups represented, study setting, model structure, and other study characteristics. Figure 3 presents model predictions of annual and cumulative TB incidence for model strata with no individual risk factors affecting TB

progression. This includes model strata for healthy adults, or for the overall population where models did not stratify by age or other risk factor.

[Figure 3]

We calculated the median prediction for annual and cumulative incidence for each year. Median annual incidence dropped from 77 cases per 1000 in the first year following infection to 1.7 per 1000 by year 20. Median cumulative incidence was 7.7% after the first year and 14.2% by the end of the 20th year. There was substantial variation between the predictions of individual models, with incidence rate predictions varying by several orders of magnitude. For the first year after infection the 90th percentile of incidence rate estimates was 52 times the 10th percentile (270 versus 5.2 per 1000). For the 20th year the same ratio was 786 (102 versus 0.13 per 1000). This variation is also evident in the cumulative incidence projections, with a ratio of 26 after 20 years (90% versus 3.5%).

Comparison of model predictions for different strata

Figure 4 presents the distribution of cumulative incidence predictions for various subsets of the model predictions after two years (commonly used to distinguish rapid progression from late reactivation) and twenty years. Cumulative incidence predictions were higher for strata dealing with any individual risk factor, and HIV in particular. Cumulative incidence predictions were higher for infants and lower for non-infant children. Distributions were approximately similar for studies conducted for high and low burden settings. Results for studies reporting no citations for TB progression parameters exhibited greater variation than those with at least one citation, particularly for 20-year results. Studies published after 2010 had greater variation in 20-year cumulative incidence than those published before that point.

Results for the different model structures were relatively similar except for Structure A, which exhibited greater variation in cumulative incidence at both 2 and 20 years, and substantially higher median incidence at 20 years, compared to other model structures. Figure S4 shows median annual and cumulative incidence projections stratified by model structure. While the trajectories of annual incidence differed by model structure, predictions produced using Structure A are markedly different to the majority of other structures, with no reduction in annual incidence over time, and steadily increasing cumulative incidence (this is also true for Structure J, though this approach was only used by one study)

[Figure 4]

We calculated incidence risk ratios associated with individual risk factors as compared to model strata from the same study without the risk factor (i.e. 'within study' comparisons). These results are shown in Figure S5, and corroborate the results shown in Figure 4, with greater TB progression risk modelled for all forms of HIV (particularly advanced HIV), and reduced risk associated with provision of antiretroviral therapy (ART) for HIV treatment and late childhood. There was no clear trend for the infant category, with some models suggesting increased risk and some suggesting reduced risk compared to adulthood, with the median risk ratio close to 1·0. Across all these comparisons there was wide variation between models, with the range of risk ratios for each comparison spanning several orders of magnitude.

Comparison of model predictions to empirical data

Figure 5 shows the distribution of incidence predictions for population groups with no individual risk factors (5^{th}, 25^{th}, 50^{th}, 75^{th} and 95^{th} percentiles) to empirical estimates for these

same quantities. Although the model results reproduce the general trend of the empirical estimates, with annual incidence rates declining over time, there is much greater variation in the modelling results, and median cumulative incidence after 10 years is 50-100% greater than both empirical estimates. For 10-year cumulative incidence, only 60% of modelling results were within a factor of 2 of either empirical point estimate, and only 77% were within a factor of 5. Ten-year cumulative incidence was greater than 50% for 15% of all modelling results, and less than 1% for 4·6% of results.

As a sensitivity analysis, we assessed the extent to which each model structure could reproduce the empirical results. When we fitted each model structure to the empirical estimates from the British Medical Research Council's BCG trials,(26) most structures were able to closely approximate the cumulative incidence estimates, with the exceptions being Structures A, D, and J, and to a lesser extent Structure E (Figure S6 and Table S3). When we reproduced the empirical comparison shown in Figure 5 excluding Structures A, D, and J, the variation was reduced but only modestly, with 71% of modelling results for 10-year cumulative incidence within a factor of 2 of the empirical point estimates, and 88% within a factor of 5. For results derived from Structures A, D, and J, 21% of modelling results for 10-year cumulative incidence were within a factor of 2 of the empirical point estimates, and 40% within a factor of 5.

Discussion

We conducted a systematic review of studies using dynamic TB transmission models, to understand how studies modelled progression to active TB following initial infection, and assess that validity of modelling assumptions by comparing model results to empirical

incidence estimates. We identified 312 studies that met our inclusion criteria, the majority of which were published after 2000.

We used the model structures and parameter values described by each study to reproduce the model predictions for TB incidence in the years following initial infection. These results demonstrated substantial disagreement between studies on a key feature of TB epidemiology—the rate at which infected individuals progress to active disease following initial infection. This variation was still apparent when we examined the subset of results that modelled the general population or population groups with no individual risk factors. When we compared the model results for groups with no individual risk factors to empirical evidence, it was evident that a substantial fraction of the modelled results were inconsistent with these data. For 10-year cumulative incidence 40% of all modelled results were either more than double or less than half the empirical point estimates.

One potential explanation for these findings is that the model structures adopted by some studies were inadequate, and when we tried to fit each model structure to the empirical data we found that three structures (A, D, and J) provided poor fit to the empirical evidence. Structure A assumes that infection with *M. tb* confers a constant rate of progression to active TB. This feature prevents these models from reproducing the declining time trend in TB progression risk shown in empirical data. By construction, these models will either underestimate near-term progression risks or overestimate long-term progression risks, or both. Structure D assumes immediate progression to active disease for all newly infected individuals. While this assumption is inconsistent with the natural history of TB in immunocompetent individuals, this structure was only used for individuals with advanced HIV who experience rapid disease progression, so this use may not be problematic.

Structure J produces progression risks that increase as a function of time since infection, which is inconsistent with the available empirical evidence.

While Structure E allowed for an immediate decline in progression risk following infection, the fit to empirical data was still crude. In a recent study examining different model structures, Ragonnet et al found that Structure E performed either worst or second worst of the six structures examined (depending on the fitting method)(39). In our analysis Structure E performed better than Structures A, D, and J, but the root mean squared error was still ten times worse than the other structures. This is notable, given that almost 50% of published models adopted this structure. Whether models with this structure will produce valid results will depend on the analysis, but it is unlikely to be appropriate for analyses that need to distinguish the elevated progression risks several years after infection from the much lower risks many years later. Apart from structures A, D, J, and potentially E, the other structures reported in the modelling literature appeared reasonable, based on their ability to reproduce empirical data when appropriate parameter values were used.

However, inadequate model structure provides only a partial explanation for the observed discrepancies. Even when we excluded Structures A, D, and J, almost 30% of all modelled results were either more than double or less than half the empirical point estimates for 10-year cumulative incidence. There are reasons to believe the epidemiology of TB progression will differ between populations – as some of the model strata we investigated pertained to the general population, each population will represent a different mix of factors such as nutrition, smoking, and diabetes that affect progression risks. As the distribution of these factors change between populations, so will TB progression rates. Recent evidence from other low-burden settings finds similar results to the empirical studies we used. In an

observational study of close contacts of TB cases in Australia, Trauer et al estimate a cumulative incidence of 5.4% over 4.5 years of follow-up for adults converting to TST or IGRA positivity(40). In a similar study in the Netherlands the 5-year cumulative incidence of active TB in adults was 6.7%(41). For high-burden settings, it is possible that part of this burden is explained through elevated progression rates. Estimating progression rates is difficult in settings with a high force of infection, given the need to distinguish reactivation from reinfection as a cause of incident disease, though some analyses have resolved this issue by studying individuals migrating from high- to low-burden settings(42-44). However, it is unlikely that differences in the distribution of factors determining progression risk would explain the magnitude of variation we observed in the modelling results. An alternative explanation is that a substantial fraction of these studies adopted assumptions that were incorrect, providing a poor representation of TB disease dynamics in their chosen population.

For population groups with individual factors modifying TB progression risks, model results were generally consistent with empirical evidence, with HIV associated with higher TB incidence compared to HIV-negative individuals, advanced HIV associated with higher incidence compared to early HIV(7, 8), and ART protective against TB for HIV-infected individuals(45). Although early infancy is empirically associated with rapid TB progression(6) this was not evident in the modelling results, potentially due to variation in the age ranges adopted by the models, and the fact that TB progression changes rapidly during this period (high in early infancy, lower in later childhood)(6). For later childhood, model results were consistent with the literature suggesting lower incidence compared to adulthood(6), though some recent studies have suggested faster progression during these ages(40, 41). While the

trends in the risk group results were generally consistent with empirical evidence, there was still substantial variation between models.

We found a range of evidence sources cited in support of the parameter values used in the studies we reviewed. These evidence sources included modelling studies, empirical studies, and review articles. Some of the evidence sources classified as modelling studies were rigorously calibrated to empirical evidence (most notably the Vynnycky and Fine(36) analysis cited by 21% of all reviewed studies), and so it should not be inferred that papers citing earlier modelling papers are necessarily less valid. However, it is possible that using earlier modelled studies as a source of parameter values plays a role in the heterogeneity of results we observed, as errors can be introduced in the process of extracting and repurposing these parameters. Even if the original model produced valid results, the same parameter values will have different implications when used in a model with different structure, or where the values of related parameters are different. Consequently, even when appropriate evidence is cited this does not necessarily imply the predictions produced by the model will be appropriate. For the 24% of studies that gave no citation for their parameter values, it is possible that these values were informed by empirical data collected as part of the study. However, it is unlikely that this explanation applies to more than a very small number of studies, if any. For the rest, the source of evidence is simply unknown.

There are several limitations to our analysis. Firstly, as we reproduced model predictions based on the content of published articles, it is possible that some of the extreme results represented typographical errors in how studies reported their approach, or that parameter values used in the analysis were modified from those reported in the paper. Although we performed double extraction, we did not contact original authors to confirm study

assumptions. Secondly, the way we programmed the models may have differed from the approach used in the original analysis. These differences could produce discrepancies between our results and those of the original analysis, though these discrepancies are likely to be minor. Thirdly, it is possible that some analyses were not attempting to reproduce TB epidemiology exactly, and that the disease was only used as a motivating example for investigating the properties of transmission models. While this may be true for some studies, we were not able to distinguish these studies in any way. For example, there was no clear difference between the predictions derived from analyses published in applied journals and those published in mathematical biology journals. Moreover, even if a particular study did not intend to fully capture TB epidemiology, it is still part of the TB modelling literature, and, as we did, readers might assume that the findings of these analyses pertain to real TB epidemiology even if this was not the intention. Finally, the empirical studies that we used as a point of comparison are not perfect. Not only do they represent particular populations, but the tests used to diagnose TB infection and active disease have imperfect sensitivity and specificity. Consequently, modelled results may not be expected to reproduce these results exactly.

Analyses that mischaracterize TB disease dynamics may produce biased estimates of descriptive epidemiology or the impact of policy change. For example, if model assumptions produce erroneously high incidence of active TB disease following initial infection, this could lead to overestimates of population-level incidence and prevalence, and overestimate the beneficial impact of interventions to reduce TB transmission. Similarly, if analyses do not allow for declines in incidence with time since infection then estimates of the impact of LTBI prophylaxis for individuals with distant infection will be biased upwards. Incorrect

assumptions about how risk factors modify TB incidence could harm the evaluation of interventions targeted at these risk factors. Moreover, as many modelling studies calibrate their transmission model to reproduce commonly reported TB outcomes, an incorrect assumption in one part of the analysis can lead to incorrect assumptions in other parts of the analysis. For example, for analyses calibrated to TB case notifications, if model assumptions produce erroneously high TB incidence following initial infection, this could lead to (amongst other things), a downward bias in estimated TB transmission, a downward bias in LTBI prevalence, and/or a downward bias in the fraction of TB cases detected. Each of these changes could introduce biases into the primary outcomes of an analysis. For example, underestimation of LTBI prevalence could lead to underestimation of the costs of a program to screen for and treat LTBI to avert active disease.

We evaluated a single characteristic of TB transmission models—the assumptions made about progression following initial infection. As we did not reproduce all features of all modelled analyses, we cannot draw conclusions about whether the discrepancies we described led to biased results in any given study. However, it is likely that these discrepancies led to biased results in some cases. While it may be impractical to reevaluate published results, our findings have clear implications for future work. This work is accelerating – there were 33 TB modelling publications in the first eight months of 2017, greater than the total for 2016, and greater than the sum of all papers published before 2000. For future studies that employ mathematical models to investigate TB epidemiology or compare policies, our results provide strong motivation to ensure structural assumptions are appropriate, and to check that analyses reproduce known features of TB epidemiology. For consumers of modelling studies, our results suggest that the findings of these studies should

not be accepted uncritically. Though there are major gaps in the evidence base for constructing and evaluating the validity of these models(15) it is still important (perhaps more important) to make the best use of the evidence that is available. Greater confidence may be placed in analyses where modelling approaches are clearly explained and justified with reference to the available evidence, and that can reproduce data relevant to the setting and population being modelled.

Acknowledgements

This study was funded by the U.S. Centers for Disease Control and Prevention, National Center for HIV, Viral Hepatitis, STD, and TB Prevention Epidemiologic and Economic Modeling Agreement #5U38PS004642. Peter White received funding from the United Kingdom National Institute for Health Research (NIHR) Health Protection Research Unit in Modelling Methodology at Imperial College London, in partnership with Public Health England (HPRU-2012-10080) and the Medical Research Council (MR/K010174/1). IA is funded by NIHR (SRF-2011-04-001; NF-SI-0616-10037), MRC and the Wellcome Trust.

Contributors

NAM, TC and JAS conceived the study, and ANH, RY, PJW, and IA helped refine the study approach. NAM, TC, JAS and EW developed the protocol for the systematic review. EW, DC, MB, AS, and KG identified relevant studies and extracted information. NAM conducted the analysis. NAM and EW developed the first draft of the manuscript. DC, MB, AS, TC, ANH, RY, KG, PJW, IA, and JAS edited the manuscript.

Conflicts of interest

PJW has received research funding from Otsuka SA for a retrospective study of multidrug-resistant tuberculosis treatment in several eastern European countries. There were no other conflicts of interest.

Citations

1. Ferebee SH. Controlled chemoprophylaxis trials in tuberculosis. A general review. Bibl Tuberc. 1970;26:28-106.

2. Sutherland I. Recent studies in the epidemiology of tuberculosis, based on the risk of being infected with tubercle bacilli. Advances in Tuberculosis Research. 1976;19:1-63.

3. Barnett G, Grzybowski S, Styblo K. Present risk of developing active tuberculosis in Saskatchewan according to previous tuberculin and X-ray status. Bull Int Union Tuberc. 1971;45:51-74.

4. Houben RM, Dodd PJ. The Global Burden of Latent Tuberculosis Infection: A Re-estimation Using Mathematical Modelling. PLoS Med. 2016;13(10):e1002152. doi: 10.1371/journal.pmed.. eCollection 2016 Oct.

5. Yuen CM, Kammerer JS, Marks K, Navin TR, France AM. Recent Transmission of Tuberculosis — United States, 2011–2014. PLOS ONE. 2016;11(4):e0153728.

6. Marais BJ, Gie RP, Schaaf HS, Hesseling AC, Obihara CC, Starke JJ, et al. The natural history of childhood intra-thoracic tuberculosis: a critical review of literature from the pre-chemotherapy era. Int J Tuberc Lung Dis. 2004;8(4):392-402.

7. Antonucci G, Girardi E, Raviglione MC, Ippolito G. Risk factors for tuberculosis in HIV-infected persons. A prospective cohort study. The Gruppo Italiano di Studio Tubercolosi e AIDS (GISTA). JAMA. 1995;274(2):143-8.

8. Selwyn PA, Hartel D, Lewis VA, Schoenbaum EE, Vermund SH, Klein RS, et al. A prospective study of the risk of tuberculosis among intravenous drug users with human immunodeficiency virus infection. New Engl J Med. 1989;320(9):545-50.

9. Bates MN, Khalakdina A, Pai M, Chang L, Lessa F, Smith KR. Risk of tuberculosis from exposure to tobacco smoke: a systematic review and meta-analysis. Arch Intern Med. 2007;167(4):335-42.

10. Chia S, Karim M, Elwood RK, FitzGerald JM. Risk of tuberculosis in dialysis patients: a population-based study. Int J Tuberc Lung Dis. 1998;2(12):989-91.

11. Lonnroth K, Williams BG, Cegielski P, Dye C. A consistent log-linear relationship between tuberculosis incidence and body mass index. Int J Epidemiol. 2010;39(1):149-55.

12. Jeon CY, Murray MB. Diabetes mellitus increases the risk of active tuberculosis: a systematic review of 13 observational studies. PLoS Med. 2008;5(7):e152.

13. Ozcaglar C, Shabbeer A, Vandenberg SL, Yener B, Bennett KP. Epidemiological models of Mycobacterium tuberculosis complex infections. Math Biosci. 2012;236(2):77-96.

14. White PJ, Garnett GP. Mathematical modelling of the epidemiology of tuberculosis. Advances in experimental medicine and biology. 2010;673:127-40.

15. Dowdy DW, Dye C, Cohen T. Data needs for evidence-based decisions: a tuberculosis modeler's 'wish list'. Int J Tuberc Lung Dis. 2013;17(7):866-77.

16. World Health Assembly. Post-2015 Global TB Strategy and Targets (A67/62). Geneva: World Health Assembly; 2014.

17. Houben RM, Menzies NA, Sumner T, Huynh GH, Arinaminpathy N, Goldhaber-Fiebert JD, et al. Feasibility of achieving the 2025 WHO global tuberculosis targets in South Africa, China, and India: a combined analysis of 11 mathematical models. Lancet Glob Health. 2016;4(11):e806-e15. doi: 10.1016/S2214-109X(16)30199-1. Epub 2016 Oct 6.

18. Menzies NA, Gomez GB, Bozzani F, Chatterjee S, Foster N, Baena IG, et al. Cost-effectiveness and resource implications of aggressive action on tuberculosis in China, India, and South Africa: a combined analysis of nine models. Lancet Glob Health. 2016;4(11):e816-e26. doi: 10.1016/S2214-109X(16)30265-0. Epub 2016 Oct 6.

19. Colijn C, Cohen T, Murray M. Mathematical models of tuberculosis: accomplishments and future challenges. BIOMAT. 2006:123-48.

20. Brooks-Pollock E, Cohen T, Murray M. The impact of realistic age structure in simple models of tuberculosis transmission. PLOS ONE. 2010;5(1):e8479-e.

21. Lipsitch M, Colijn C, Cohen T, Hanage WP, Fraser C. No coexistence for free: neutral null models for multistrain pathogens. Epidemics. 2009;1(1):2-13.

22. Cohen T, Colijn C, Finklea B, Murray M. Exogenous re-infection and the dynamics of tuberculosis epidemics: local effects in a network model of transmission. J Roy Soc Interface. 2007;4(14):523-31.

23. Wearing HJ, Rohani P, Keeling MJ. Appropriate models for the management of infectious diseases. PLoS Med. 2005;2(7):e174.

24. Feng Z, Huang W, Castillo-Chavez C. On the Role of Variable Latent Periods in Mathematical Models for Tuberculosis. J Dyn Differ Equ. 2001;13(2):425-52.

25. Colijn C, Cohen T, Murray M. Emergent heterogeneity in declining tuberculosis epidemics. J Theor Biol. 2007;247(4):765-74.

26. Sutherland I. The ten-year incidence of clinical tuberculosis following "conversion" in 2550 individuals aged 14 to 19 years. TSRU Progress Report. The Hague; 1968.

27. Medical Research Council. BCG and vole bacillus vaccines in the prevention of tuberculosis in adolescents; first (progress) report to the Medical Research Council by their Tuberculosis Vaccines Clinical Trials Committee. British medical journal. 1956;1(4964):413-27.

28. Ferebee SH, Mount FW. Tuberculosis morbidity in a controlled trial of the prophylactic use of isoniazid among household contacts. Am Rev Respir Dis. 1962;85:490-510.

29. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med. 2009;6(7):e1000100.

30. TB Modelling and Analysis Consortium. A systematic review of mathematical and economic TB modelling papers (http://tb-mac.org/Resources/Resource/4, last accessed July 26 2017) 2013 [

31. Andrews JR, Noubary F, Walensky RP, Cerda R, Losina E, Horsburgh CR. Risk of progression to active tuberculosis following reinfection with Mycobacterium tuberculosis. Clin Infect Dis. 2012;54(6):784-91.

32. WHO. Global TB Report 2016. Geneva: WHO; 2016.

33. American Thoracic Society, American Lung Association, U.S. Centers for Disease Control. Preventive therapy of tuberculosis infection. Am Rev Respir Dis. 1974;110:371-4.

34. Styblo K. Epidemiology of tuberculosis: selected papers. Vol. 24. The Hague, The Netherlands: Royal Netherlands Tuberculosis Association; 1991.

35. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2016.

36. Vynnycky E, Fine PE. The natural history of tuberculosis: the implications of age-dependent risks of disease and the role of reinfection. Epidemiol Infect. 1997;119(2):183-201.

37. Blower SM, Mclean AR, Porco TC, Small PM, Hopewell PC, Sanchez MA, et al. The intrinsic transmission dynamics of tuberculosis epidemics. Nat Med. 1995;1(8):815-21.

38. Dye C, Garnett GP, Sleeman K, Williams BG. Prospects for worldwide tuberculosis control under the WHO DOTS strategy. Directly observed short-course therapy. Lancet. 1998;352(9144):1886-91.

39. Ragonnet R, Trauer JM, Scott N, Meehan MT, Denholm JT, McBryde ES. Optimally capturing latency dynamics in models of tuberculosis transmission. Epidemics. 2017.

40. Trauer JM, Moyo N, Tay E-L, Dale D, Ragonnet R, McBryde ES, et al. Risk of Active Tuberculosis in the Five Years Following Infection . . . 15%? Chest. 2016;149(2):516–25.

41. Sloot R, Schim van der Loeff MF, Kouw PM, Borgdorff MW. Risk of tuberculosis after recent exposure. A 10-year follow-up study of contacts in Amsterdam. Am J Resp Crit Care. 2014;190(9):1044-52.

42. Aldridge RW, Zenner D, White PJ, Williamson EJ, Muzyamba MC, Dhavan P, et al. Tuberculosis in migrants moving from high-incidence to low-incidence countries: a population-based cohort study of 519 955 migrants screened before entry to England, Wales, and Northern Ireland. Lancet. 2016;388(10059):2510-8.

43. Ricks PM, Cain KP, Oeltmann JE, Kammerer JS, Moonan PK. Estimating the Burden of Tuberculosis among Foreign- Born Persons Acquired Prior to Entering the U.S., 2005– 2009. PLOS ONE. 2011;6(11):e27405-e.

44. Vos AM, Meima M, Verver S, Looman CWN, Bos V, Borgdorff MW, et al. High Incidence of Pulmonary Tuberculosis a Decade after Immigration, Netherlands. Emerg Infect Dis. 2004;10(4):736–9.

45. Suthar AB, Lawn SD, del Amo J, Getahun H, Dye C, Sculier D, et al. Antiretroviral therapy for prevention of tuberculosis in adults with HIV: a systematic review and meta-analysis. PLOS Med. 2012;9(7):e1001270-e.

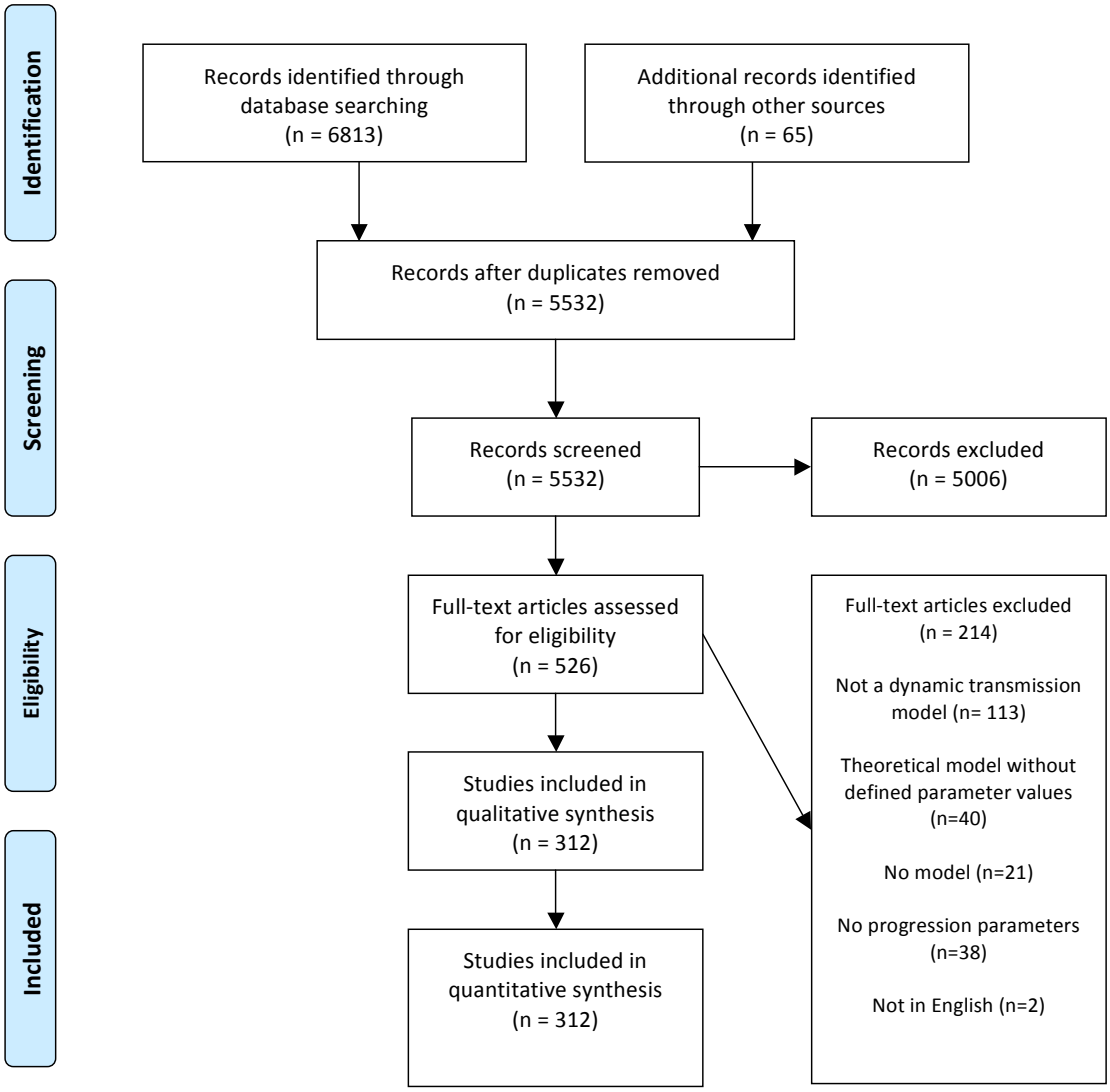

Figure 1: Flow diagram of studies assessed for the review*.

* Other sources included a database of modelling publications compiled by the TB Modelling and Analysis Consortium, the reference lists of eligible publications, a group of non-indexed journals, and the personal databases of the investigators to identify publications not included in the electronic search.

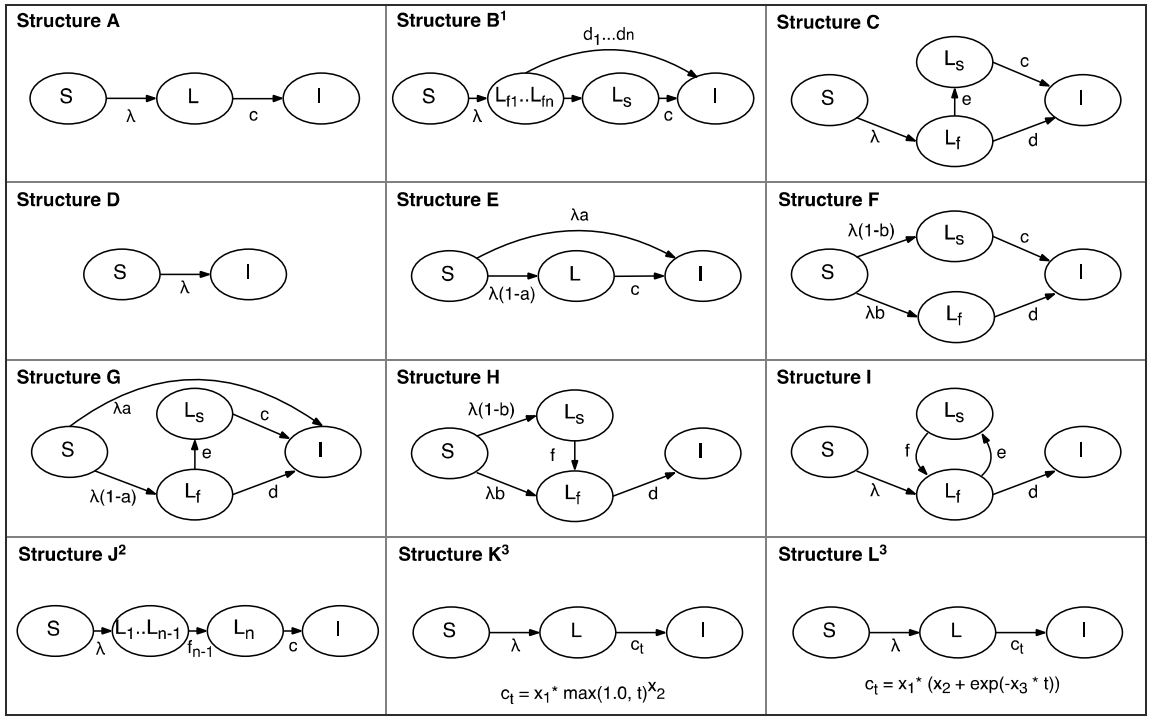

Figure 2: Classification of model types and transition probabilities*.

*Descriptions of compartments and transitions are provided below. Note that some model structures are special cases of other structures. For example, Structures A and C are special cases of Structures E and G respectively, with parameter 'a' set to zero.

S = Susceptible compartment (not infected with TB, not previously exposed)

L = Latent *M. tb* infection compartment

L_s = Slow Latent *M. tb* infection compartment

L_f = Fast Latent *M. tb* infection compartment

I = Active TB disease compartment

λ = Force of infection for *M. tb*

a = Probability of immediate progression to active TB compartment (I), for individuals in susceptible compartment (S) who are infected with *M.tb*

b = Probability of progression to fast latent compartment (L_f), for individuals in susceptible compartment (S) who are infected with *M.tb*

c = Rate of progression to active TB (I) for individuals in the Latent compartment (L) or slow Latent compartment (L_s)

d = Rate of progression to active TB (I) for individuals in the fast Latent (L_f) compartment

e = Rate of transition to the slow Latent compartment (L_s) for individuals in the fast Latent (L_f) compartment

f = Rate of transition to the fast Latent compartment (L_f) for individuals in the slow Latent (L_s) compartment

[1] Structure B involves a set of tunnel states for recent latent infection ($L_{f1}..L_{fn}$), whereby individuals not progressing to active TB transition deterministically to next tunnel state (n+1) each time step. Each of these compartments has a different progression risk ($d_1..d_n$).

[2] Structure J involves a sequence of latent compartments ($L_1..L_n$), with individuals only transitioning to the active TB compartment from the final compartment.

[3] Structures K and L involve a single latent compartment, with the rate of transition to active TB calculated as a function of time since infection. Both of these structures were implemented using individual based models, allowing time since infection to be tracked at the individual level.

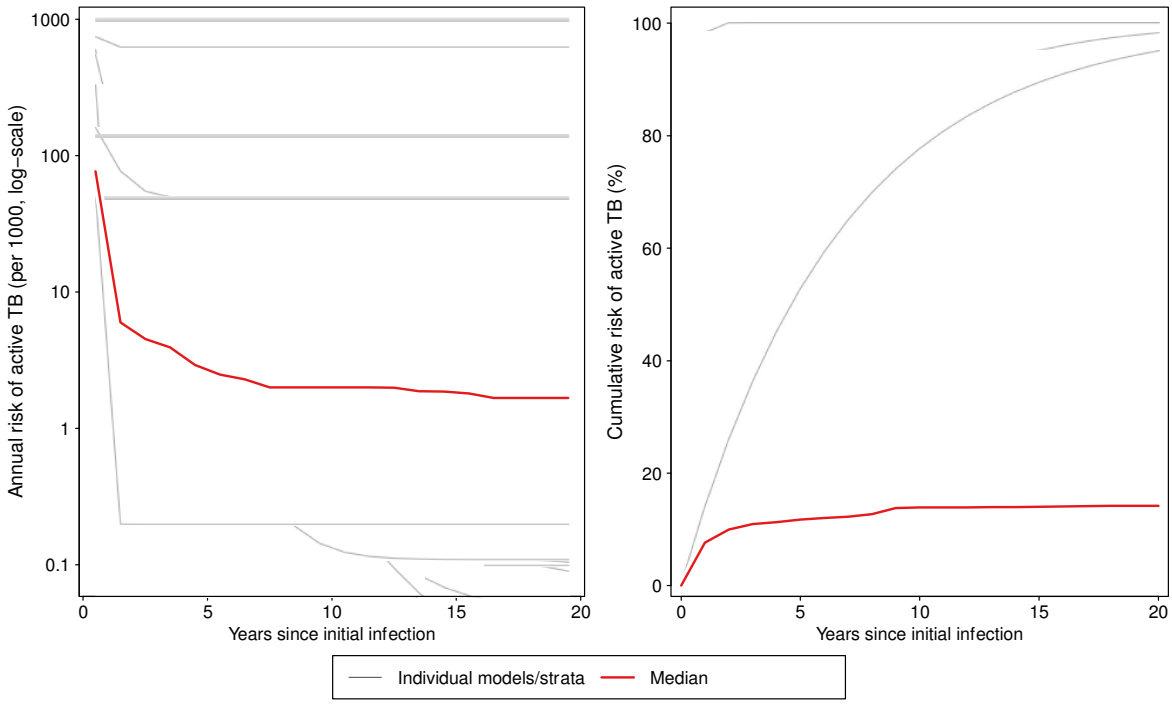

Figure 3: Model predictions for annual and cumulative incidence of active TB by years since *M. tb* infection, for population groups with no individual risk factors.

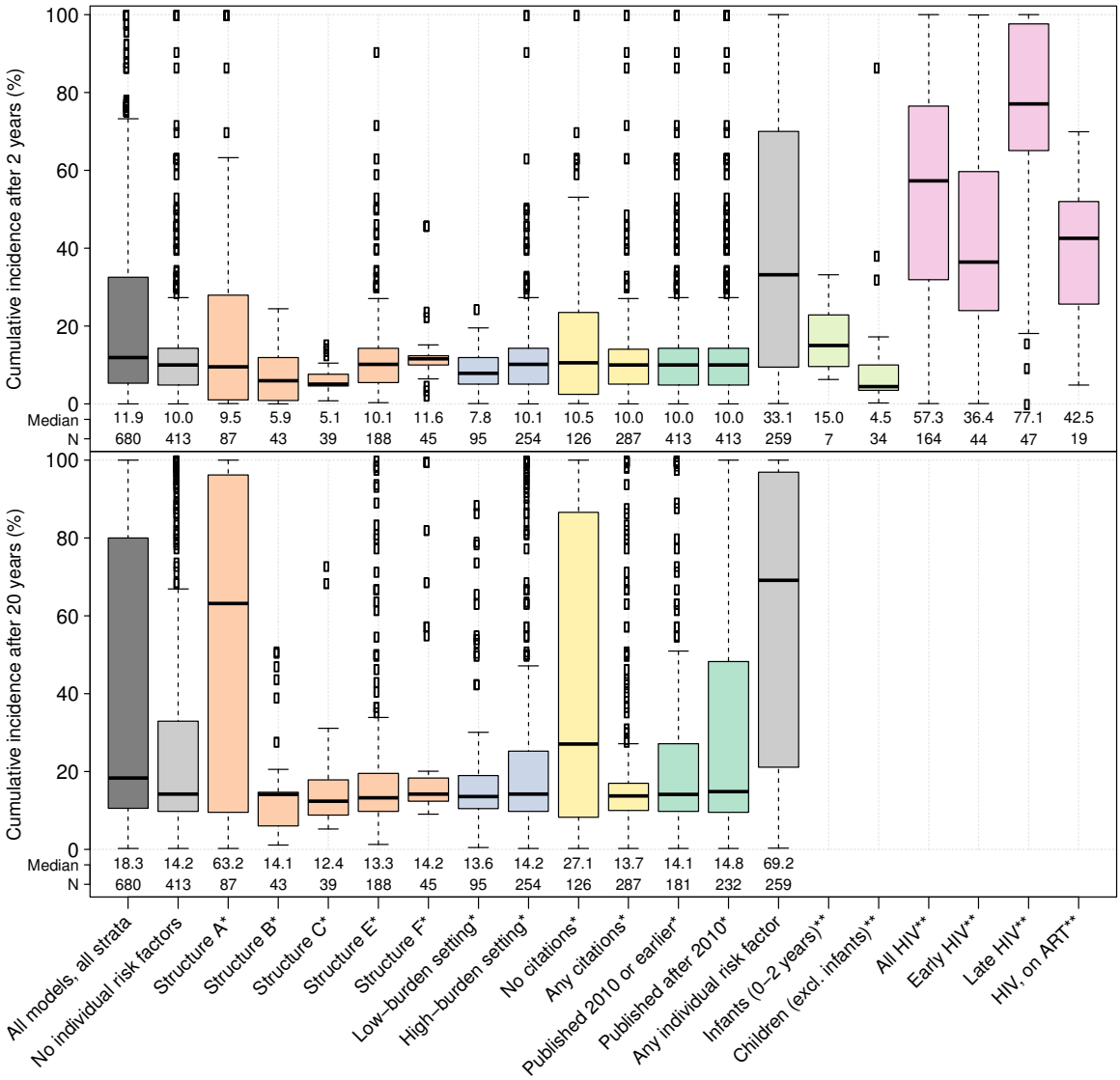

Figure 4: Distribution of model predictions for cumulative incidence of active TB at two and twenty years since *M. tb* infection, stratified by model structure, individual risk factors, and other study characteristics***.

* Only includes results for population groups with no individual factors modifying TB progression risks.

** Twenty-year cumulative incidence projections not shown for these groups due to potential for unmodelled changes in risk factors.

*** Individual results not shown for Structures D, G, H, I, J, and K, as <5 studies used these structures to model individuals with no other risk factors.

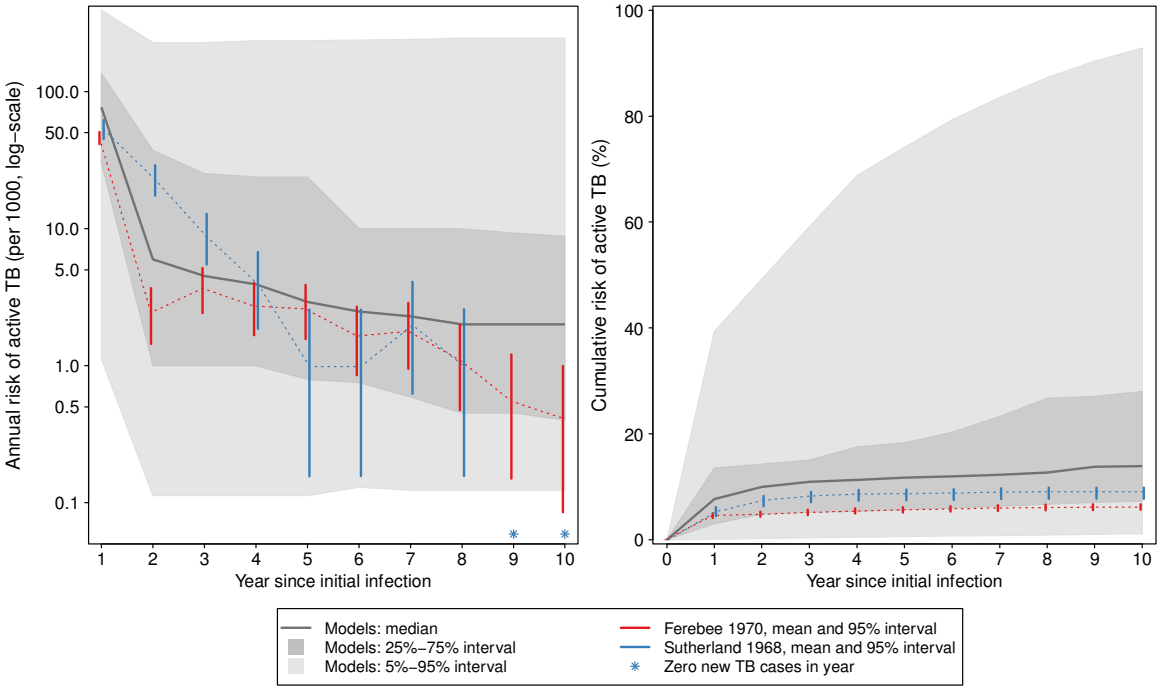

Figure 5: Comparison between model predictions and empirical evidence for annual and cumulative incidence of active TB by years since *M. tb* infection, for groups with no individual risk factors*.

* Empirical estimates based on the British Medical Research Council BCG trials (26), and the USPHS isoniazid trials (28).

Table 1: Descriptive statistics of included studies.

Category	No. publications (% of total)
Publication year	
1960-1969	4 (1.3%)
1970-1979	1 (0.3%)
1980-1989	1 (0.3%)
1990-1999	15 (4.8%)
2000-2009	95 (30.4%)
2010-2017	196 (62.8%)
Model Structure*	
A	60 (19.2%)
B	27 (8.7%)
C	33 (10.6%)
D	3 (1.0%)
E	153 (49.0%)
F	35 (11.2%)
G	1 (0.3%)
H	2 (0.6%)
I	2 (0.6%)
J	1 (0.3%)
K	1 (0.3%)
L	1 (0.3%)
Setting*	
High Burden	193 (61.9%)
Low Burden	72 (23.1%)
Not specified	72 (23.1%)
Risk strata*	
Age	29 (10.0%)
Drug resistance	10 (3.2%)
Foreign born	5 (1.6%)
Genetic susceptibility	4 (1.4%)
Poverty	1 (0.3%)
Rural/urban	1 (0.3%)
Sex	2 (0.7%)
Smoking	4 (1.4%)
Incarceration	2 (0.7%)
Diabetes	2 (0.7%)
Famine/nutrition	2 (0.7%)
HBV	1 (0.3%)
HIV	79 (27.1%)
Malaria	1 (0.3%)
Silicosis	2 (0.7%)
Any risk stratification	122 (39.1%)
Total	312 (100%)

* Categories sum to >100% as some papers are included in multiple categories (i.e. utilize multiple different structures, present results for multiple settings, or stratify progression risk along multiple dimensions).