Modeling the impact of case finding for tuberculosis: The role of infection dynamics

Authors: Theresa S Ryckman[1], Sourya Shrestha[2], Anthony T Fojo[3], Parastu Kasaie[2], David W Dowdy[2], Emily A Kendall[1]

[1]Department of Medicine, Division of Infectious Diseases, Johns Hopkins School of Medicine, Baltimore, Maryland 21287

[2]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland 21205

[3]Department of Medicine, Division of General Internal Medicine, Johns Hopkins School of Medicine, Baltimore, Maryland 21287

Abstract: Population-level case-finding is globally recommended as a strategy to reduce the burden of tuberculosis in high-burden settings, but mathematical models overwhelmingly project that case-finding will yield only modest reductions in incidence. We evaluated whether the common assumption of lifelong risk of progression after infection leads models to undervalue the impact of case-finding. Specifically, clearance of latent infection reduces the rate of long-term “reactivation”, which could result in more pronounced effects from interventions (such as case-finding) that reduce *M. tuberculosis* transmission. To investigate this possibility, we constructed two models - one assuming lifelong progression risk after infection (“Conventional”), and one assuming that some individuals clear their infections (“Clearance”). We calibrated these models to empirical data on tuberculosis burden in India and progression after infection. We used the calibrated models to project the impact of community-wide active case-finding campaigns covering 75% of the population every two years, with or without mass provision of tuberculosis preventive treatment (TPT). Relative to the Conventional model, the posterior distributions of Clearance model outputs were more similar to empirical estimates of disease incidence, infection prevalence, progression over time, and incidence trends, demonstrating better fit. The Clearance model also projected a greater impact of case-finding on incidence: 26% [95% uncertainty interval: 15-41%] vs. 11% [7-24%] reduction after 10 years. Given that clearance of some infections is biologically likely and results in better model fit, we conclude that the epidemiological impact of population-level case-finding for tuberculosis may be greater than that projected by models that exclude this effect.

Introduction

Reducing the substantial global burden of tuberculosis (TB; 10.8 million new cases and 1.3 million deaths in 2023) will require interventions to find and treat more individuals with undiagnosed TB (1). Efforts to actively screen populations at risk, find people with undiagnosed TB, and link them to treatment ("active case-finding") can be conducted at a community-wide scale, using combinations of diagnostic tools (e.g., chest X-ray, bacteriological testing) to detect even those who lack typical TB symptoms. Community-based active case-finding is being increasingly prioritized as a method for reducing the burden of TB in high-burden communities (2–4).

Empirical evidence on the impact of active case-finding has been mixed and is challenging to obtain, owing to the need for very large sample sizes to demonstrate effects at a population level (5). At least one case-finding study, the ACT3 randomized trial in Vietnam, detected a large and significant reduction in prevalence (6). However, given the limited empirical evidence and challenges in carrying out larger and longer-term trials, TB programs in high-burden settings may decide their strategies based on results from mathematical modeling, which can translate empirical data into long-term population-level projections. Several modeling studies project that the impact of mass case-finding on TB incidence will be limited and temporary, even with interventions intensive enough to achieve the short-term reductions in prevalence observed in studies like ACT3 (7–10). However, findings from models are limited by the simplifying assumptions underlying their structure and parameters, which can directly influence model projections.

Notably, most mathematical models assume that individuals infected with *Mycobacterium tuberculosis* (*Mtb*, the pathogen that causes TB) face a lifelong risk of progressing to disease (11). However, a more nuanced understanding of *Mtb* infection and immunoreactivity is emerging, and several lines of evidence suggest that some *Mtb* infections may clear absent intervention (11–14). Incorporating infection clearance into mathematical models could change key components of those models, such as whether incidence is driven more by recent or remote infection, that could affect the projected impact of interventions (such as case-finding) that seek to reduce *Mtb* transmission (15).

We therefore sought to evaluate whether incorporating infection clearance into a typical *Mtb* transmission model would alter the projected impact of active case-finding. To accomplish this, we used the same set of empirical data (burden data from India and historical longitudinal cohort data) to calibrate two otherwise similarly structured *Mtb* transmission models: a "Conventional" model with lifelong progression risk and a "Clearance" model in which some infections clear. We then used both models to simulate the impact of community-wide active case-finding every two years, and evaluated the impact on TB incidence over 10 years.

Results

Model calibration

We used a Bayesian method to calibrate the Clearance and Conventional model parameters to 11 targets grounded in empirical data. The Clearance model was able to simultaneously fit several of these multiple targets better than the Conventional model (Figure 1**,** Fig. S2). In particular, while both models yielded reasonable fits to data on annual TB incidence and cumulative progression over time since infection (Figure 1, panels A-C), the Conventional model often overestimated the prevalence of *Mtb* infection; the mean posterior estimate of *Mtb* infection prevalence was 56% [95% Uncertainty Interval 43-73%] under the Conventional model, as compared to 31% [20-44%] prevalence of interferon-gamma release assay (IGRA) positivity estimated from India's national TB prevalence survey and a posterior estimate of 34% [21-50%] from the Clearance model (Figure 1, panel D).

The Conventional model also tended to fit data on pre-intervention trends in incidence worse than the Clearance model; only 5% of posterior samples overlapped with target confidence intervals, compared to 53% of Clearance model posterior samples (Figure 1, panel E). The difference in the fit to pre-intervention

trends was similar regardless of whether reductions in incidence were modeled as resulting from higher treatment initiation (primary analysis) or declines in the transmission rate (Fig. S2).

Subsets of Conventional model parameter sets that fit some of these five targets did not perform as well for others (Table S5). Sensitivity analysis with different combinations of calibration targets further demonstrated tension under the Conventional model in simultaneously fitting these five targets – and that this tension was resolved by incorporating infection clearance (SI Text 2; Figures S3-S6). For example, even when the Conventional model was calibrated only to disease incidence, progression over time, and *Mtb* infection prevalence, infection prevalence was still overestimated (Fig. S3).

Compared to an annual risk of *Mtb* infection target of 2.0% [0.5-4.3%] based on tuberculin skin test (TST) positivity data in children, the Clearance model estimated a higher annual risk of infection than the Conventional model (mean 4.1% [2.2-6.7%] versus 1.4% [0.8-2.4%]; Figure 1, panel F). Fit to the remaining calibration targets was similar across both models (Figure 1, panels G-K). Posterior parameter and likelihood distributions are shown in Figures S7-S8.

Modeled infection dynamics

Recently acquired infection accounted for a greater share of total *Mtb* infection prevalence under the Clearance model (16% [8-28%]) compared to the Conventional model (4% [3-6%]; Figure 2, Fig. S9). As a result, the percentage of TB incidence attributable to progression from recent infection was also substantially greater in the Clearance model (63% [42-83%], versus 34% [20-66%] in the Conventional model). To illustrate the proportion of next-year incidence of disease attributable to transmission from the population with prevalent TB at a single point in time, we performed an “avertable fraction” analysis in which we moved the entire population with active TB to a recovered state (analogous to immediate diagnosis and treatment of all individuals with TB) and simulated incidence of disease over the following year. Under the Clearance model, incidence in the following year fell by 11% [7-14%] in this analysis, versus 6% [3-12%] under the Conventional model. SI Text 2 and Fig. S10 explain how these results are influenced by both the model calibration process and the inclusion of a clearance mechanism.

Interventions

Figure 3 shows the projected impact of community-wide active case-finding waves every two years, covering 75% of the population each time with a 65% sensitive screening algorithm, with or without mass TPT provision. With active case-finding alone, the predicted reduction in incidence at year 10 was greater in the Clearance model (26% [15-41%]) than the Conventional model (11% [7-24%]). However, for the combined intervention of both active case-finding and TPT, impact was similar across models (all 95% UIs within 75-83%). These findings were similar regardless of the mechanism assumed to underly pre-intervention trends in TB incidence (Figures S11-S12). Prevalence and mortality over time are shown in Figures S13-S14.

Comparison to empirical data

As an additional analysis, we evaluated the models’ performance in fitting evidence not used in calibration, from the ACT3 cluster randomized case-finding trial (6, 16). We simulated annual case-finding, using high levels of coverage to replicate the 44% decline in prevalence (relative to no intervention) observed in ACT3. Under these conditions, the Clearance model came closer (projected 22% [14-29%] incidence reduction) than the Conventional model (12% [7-27%]) to replicating the corresponding 38% [19-61%] reduction in incidence estimated from ACT3 (Table S6, SI Text 1).

Sensitivity analysis

The impact of case-finding overall, and as a share of the combined impact of case-finding plus TPT, remained higher with the Clearance model than the Conventional model in all sensitivity analyses performed, including those that varied intervention coverage and effectiveness (Figures S15-S18). The

impact of case-finding was greatest when more TB incidence was from recent infection, when pre-intervention incidence reductions were greater, and when the probability of treatment success was higher.

Discussion

In this study, we compared *Mtb* transmission models that differed only in whether they included the possibility of *Mtb* infection clearance. We evaluated models based on their fit to empirical data and their projections of the impact of TB case-finding on future TB incidence. Models that allowed clearance of some infections fit empirical data better and also projected a greater impact from case-finding interventions. This result suggests that case-finding could have greater epidemiological impact than most models have predicted.

Although evidence from longitudinal datasets (17) and population-based studies (18) support some persistence of reactivation risk, models representing this risk as lifelong have been based primarily on observational data from cohorts followed for only 5-10 years (11, 19). Extending this elevated reactivation risk into later decades of infection conflicts with several pieces of evidence about *Mtb* infection dynamics. First, in longitudinal cohort studies conducted in low-incidence settings, the vast majority of individuals who progress to TB disease do so in the first two years, with very few documented instances of progression beyond 5 years (11, 13). Second, TB immunoreactivity persists even after treatment for TB, and people with immunoreactivity do not universally progress to TB even after immunosuppression. For example, of people with HIV or solid organ transplant recipients with positive TST results who did not receive TPT, <20% progressed to TB over 2.5 to 5 years (14). These findings suggest that immunoreactivity sometimes indicates immunological memory rather than persistent infection with viable bacteria. Third, reversion of immunoreactivity is common (e.g., 37-58% among >45,000 children immigrating to the United States (20)), which may indicate clearance of infection prior to producing lasting immune memory (21).

Allowing for the possibility of clearance resulted in a model that better fit empirical evidence, including data on progression over time, disease burden, and declining incidence trends. In contrast with the Conventional model, by allowing some older infections to clear, the Clearance model could have enough recently acquired *Mtb* infections (which progress at a higher rate) to replicate empirical rates of TB incidence without exceeding data-consistent levels of *Mtb* infection. This finding is consistent with another modeling exercise that also required moderate to high rates of clearance from infection and early disease states to calibrate to empirical targets (22, 23). The higher proportion of recently acquired *Mtb* infections in the Clearance model also allowed this model to respond more rapidly to changes in *Mtb* transmission (e.g., via changes in the prevalence of infectious TB) and thus allowed both changing parameter values (in the pre-intervention period) and future case-finding interventions to achieve more rapid declines in TB incidence (see SI Text 2 and Fig. S10 for additional details).

Notably, we evaluated both transmission rate reductions and treatment rate increases as mechanisms to achieve pre-intervention incidence trends. In both variations, allowing for clearance improved consistency with empirical data and increased the impact of case-finding. We did not consider other potential mechanisms, such as demographic shifts or secular reductions in progression rates. Although the Clearance model tended to estimate a higher annual risk of infection than suggested by TST surveys of children, a growing body of evidence suggests that risks may be higher than such surveys would indicate (24–26). The Clearance model's predictions of incidence reductions after case-finding were also most consistent with those measured in the ACT3 trial - although both models underestimated those reductions (16).

Put together, these results suggest that models which include clearance of some infections may be more accurate representations of TB epidemiology at a population-level, and that case-finding could have a greater impact than previously projected by models that did not account for this dynamic. In particular, reaching 70% of the population with a case-finding intervention every two years was projected to reduce the incidence of TB by 26% after 10 years under the model with clearance, but only 11% without. In a

population of 10 million people, for instance, these results translate to differences on the order of tens of thousands of cases prevented, demonstrating the importance of structural choices – particularly when models are used to guide policy decisions.

Although we calibrated the models to epidemiologic data from India, these findings are expected to generalize across most high-burden settings with low HIV prevalence. Five key targets determined calibration success and influenced the balance of recent versus remote infection in driving epidemics (SI Text 2; Fig. S10), and the relationships between these five targets in most other high-TB-burden countries favor recent infection to a similar or even greater degree (Table S1). Most have either a lower ratio of *Mtb* infection prevalence to TB incidence than India, higher annual average reductions in incidence (where adequate data exist to inform trend estimates), or both (1, 27). In these settings, our results suggest that models that do not incorporate clearance are likely to fit empirical data even worse. Other unique epidemiologic features of India's TB epidemic, such as a high prevalence-to-notification ratio, were not major drivers of calibration fit or projected intervention impact.

On the other hand, these results may not generalize to settings with a high TB-HIV co-prevalence. People with HIV are at greater risk of TB progression, affecting the balance between infection prevalence and disease incidence and, potentially, the responsiveness of TB incidence to treatment scaleup. A setting-specific model that represents these dynamics would be needed to apply the structural questions considered here to a high HIV-prevalence setting.

We only explored one of several mechanisms through which models may underestimate the impact of mass case-finding. Other inherent limitations of compartmental models that may similarly dampen impact include homogeneity, memoryless-ness, and determinism. Models could be adapted to incorporate heterogeneity in transmissibility and susceptibility, to represent correlations in individual-level participation across successive intervention waves, or to allow greater dynamism of TB incidence by modeling stochasticity in the emergence and control of outbreaks, all of which may alter the impact of case-finding. The fact that none of our calibrated models were able to fully replicate ACT3 incidence reductions suggests a need for future work to assess the influence of these other mechanisms on case-finding impact.

Our model assumes immunoreactivity is equivalent to infection prevalence, and therefore that reversion of immunoreactivity is equivalent to infection clearance. A more nuanced representation would consider that some individuals may maintain immunological memory (remaining TST/IGRA positive) without being at risk for progression, while others who do not have a positive TST/IGRA (due to reversion or lack of conversion) may remain at elevated risk. Another limitation is that all model variations overestimated the proportion of people with TB who had begun to seek care, compared to self-report in India's national prevalence survey. However, this target was found to be relatively uncorrelated with intervention impact. Modeling the care-seeking process in more complexity, via different gradations of symptoms and care seeking, would likely improve fit without substantially affecting results related to case-finding impact.

From this study's results, we draw two main conclusions. First, in order to more accurately replicate population-level data, modelers should consider representing *Mtb* infection in a way that allows progression rates to decline beyond the first 5 to 10 years of infection (whether via clearance or some other mechanism). Second, by not allowing for infection clearance, most *Mtb* transmission models have likely underestimated the impact of active case-finding, and of other interventions that primarily reduce transmission or the burden of undiagnosed disease. As such, early empirical evidence suggesting that case-finding could result in large prevalence and incidence reductions may be a better reflection of impact than many existing model projections. Active case-finding remains a viable and potentially powerful strategy to reduce the burden of TB in high-burden settings. Finding ways to tailor case-finding interventions for enhanced efficiency and feasibility in high-burden settings should remain a priority.

Material and Methods

Overview

We adapted a previously published mathematical model of *Mtb* transmission to explore how assumptions about infection altered fits to empirical calibration targets and projections of the impact of interventions (7). In particular, we evaluated models that differed on whether all *Mtb* infections conferred lifelong progression risk. Both models were deterministic, dynamic compartmental models calibrated to a set of common empirical targets, representing a diverse set of epidemiological measures, including the prevalence of infection and disease, estimated rates of incidence and mortality, and observed progression over time after new *Mtb* infection.

Model Structure

The Conventional model followed a typical *Mtb* transmission model structure (Figure 4, SI Text 1, Fig. S1), in which a population transitions between 6 TB-related states. In this model, progression risk after *Mtb* infection is highest in the first 2 years but persists for life among all who have been infected. To model the potential for *Mtb* infection clearance, in the alternative Clearance model, we included a transition from the remote infection state to the recovered state (this transition rate was set to 0 in the Conventional model).

In both models, after progression to disease, TB starts in a non-care-seeking phase. From there, it can resolve to the recovered state or progress to symptom-driven care-seeking and, from there, either to cure through treatment or to death. Both models were further stratified into two age groups: children <15 years and adults ≥15 years, who face differential progression and mortality risks. Active case-finding and TPT interventions were simulated via movement to the recovered state from all TB disease states and all infection states, respectively.

Model Calibration

Model parameters were calibrated to prevalence-survey-based and WHO-reported disease burden data from India (selected as an illustrative but representative high-burden country; SI Text 1, Table S1), plus historical cohort data on progression over time after new *Mtb* infection (Table 1, SI Text 1, Tables S2-S3). India-specific targets included estimates from WHO of the incidence of TB, the rate of change in this incidence over time, and the case fatality ratio (28), as well as estimates from India's 2019-21 national TB prevalence survey of the prevalence of TB disease (in absolute terms and relative to annual notifications) and the prevalence of *Mtb* infection (based on immunoreactivity) (29). We also included an estimate of the annual risk of *Mtb* infection based on tuberculin skin test (TST) prevalence among children in India (30), which we adjusted to be population-representative by accounting for age variation in mixing and infectiousness (SI Text 1).

Additional targets were based on empirical estimates of time to progression to TB disease after infection, from two historical cohorts of recent TST converters (11, 31, 32). We extracted estimates of cumulative progression at 2 and 10 years (chosen to inform recent and remote progression rates). Since these studies were conducted in low-prevalence settings, we assumed that all TB in these cohorts resulted from progression rather than reinfection.

After a burn-in period to reach a steady state, we induced declines in incidence over 20 years to match data on incidence time trends. In the main analysis, we used an annual linear increase in the treatment initiation rate to generate these incidence declines, reflecting expansions in treatment that occurred over this period. In secondary analysis, we considered instead a declining transmission rate, reflecting development-related changes in opportunities for transmission or susceptibility to infection. Besides the target for annual reductions in TB incidence (which was compared to the average rate of change over this 20-year period), calibration targets were compared to model outputs after these 20 years of incidence decline.

The models were coded in R v4.2.2 and calibrated using Bayesian Incremental Mixture Importance Sampling with uninformed priors (SI Text 1), resulting in 50,000 posterior parameter set samples per model (33). The model code is available at www.github.com/rycktessman/tb-model-clearance.

Interventions

We used the calibrated Conventional and Clearance models to simulate the impact of community-wide active case-finding campaigns every 2 years, covering 70% of the population during each campaign with a 65% sensitive screening algorithm, with or without mass TPT provision for those without TB disease (SI Text 1; Table S4). We evaluated intervention impact on TB incidence over a 10-year horizon and compared the relative impacts of case-finding and TPT between models.

We also assessed key metrics of infection dynamics absent intervention, such as the proportion of prevalent *Mtb* infections that were acquired recently, the proportion of TB incidence arising from recently versus remotely acquired infections, and the proportion of the next year's TB incidence that could be attributed to the current population with prevalent TB. The latter was estimated by instantaneously moving all individuals in the active TB compartments to the recovered compartment and running the model for one year.

We conducted probabilistic sensitivity analysis to evaluate uncertainty across all model parameters. Main results are presented as mean values, with 95% uncertainty intervals (UI) representing the 2.5th and 97.5th percentiles across 50,000 parameter set samples. We also evaluated correlations between parameters and modeled outcomes, examined variation in projected intervention impact between the lowest and highest deciles of each parameter distribution (as a form of one-way sensitivity analysis), and conducted two-way sensitivity analysis on the percentages of active TB and *Mtb* infection cured via intervention.

Additionally, we assessed whether the models could replicate findings from the ACT3 cluster randomized case-finding trial, which is the only case-finding trial we are aware of that reported estimates of impact of incidence (6, 16). Specifically, we simulated three years of annual case-finding and tuned the proportion of the population with TB cured by the intervention to match the 44% reduction in TB prevalence observed in ACT3, relative to no intervention (6). We measured the resulting reduction in TB incidence in years 2 through 4 of the simulated intervention relative to no intervention, and we compared this to the analogous empirical reduction in TB incidence observed in ACT3, which we estimated as 38% [95% UI 19-61%] by adjusting the trend reported in the trial's intervention arm (16) for non-intervention-related declines in disease burden during the study (SI Text 1) (6).

Funding: National Institute of Allergy and Infectious Diseases (NIH K01AI182503 to TSR), National Heart Lung and Blood Institute (NIH R01HL153611 to EAK), Johns Hopkins University Tuberculosis Research Advancement Center (NIH P30AI168436 to PK and TSR).

References

1. World Health Organization, “Global Tuberculosis Report 2024” (2024).

2. O. Biermann, *et al.*, Active case-finding policy development, implementation and scale-up in high-burden countries: A mixed-methods survey with National Tuberculosis Programme managers and document review. *PLoS One* **15**, e0240696 (2020).

3. T. Pande, *et al.*, Finding the missing millions: lessons from 10 active case finding interventions in high tuberculosis burden countries. *BMJ Glob Health* **5** (2020).

4. T.-A. Nguyen, *et al.*, Population-wide active case finding as a strategy to end TB. *The Lancet Regional Health – Western Pacific* **46** (2024).

5. P. MacPherson, *et al.*, Community-based active-case finding for tuberculosis: navigating a complex minefield. *BMC Global and Public Health* **2**, 9 (2024).

6. G. B. Marks, *et al.*, Community-wide Screening for Tuberculosis in a High-Prevalence Setting. *New England Journal of Medicine* (2019). https://doi.org/10.1056/NEJMoa1902129.

7. S. Shrestha, *et al.*, Achieving a “step change” in the tuberculosis epidemic through comprehensive community-wide intervention: a model-based analysis. *BMC Medicine* **19**, 244 (2021).

8. R. Ragonnet, *et al.*, Estimating the long-term effects of mass screening for latent and active tuberculosis in the Marshall Islands. *Int J Epidemiol* **51**, 1433–1445 (2022).

9. R. M. G. J. Houben, *et al.*, Feasibility of achieving the 2025 WHO global tuberculosis targets in South Africa, China, and India: a combined analysis of 11 mathematical models. *The Lancet Global Health* **4**, e806–e815 (2016).

10. C. Dye, P. Glaziou, K. Floyd, M. Raviglione, Prospects for tuberculosis elimination. *Annu Rev Public Health* **34**, 271–286 (2013).

11. N. A. Menzies, *et al.*, Progression from latent infection to active disease in dynamic tuberculosis transmission models: a systematic review of the validity of modelling assumptions. *Lancet Infect Dis* **18**, e228–e238 (2018).

12. M. A. Behr, E. Kaufmann, J. Duffin, P. H. Edelstein, L. Ramakrishnan, Latent Tuberculosis: Two Centuries of Confusion. *Am J Respir Crit Care Med* **204**, 142–148 (2021).

13. M. A. Behr, P. H. Edelstein, L. Ramakrishnan, Revisiting the timetable of tuberculosis. *BMJ* **362**, k2738 (2018).

14. M. A. Behr, P. H. Edelstein, L. Ramakrishnan, Is Mycobacterium tuberculosis infection life long? *BMJ* **367**, l5770 (2019).

15. J. C. Emery, *et al.*, Self-clearance of Mycobacterium tuberculosis infection: implications for lifetime risk and population at-risk of tuberculosis disease. *Proceedings of the Royal Society B: Biological Sciences* **288**, 20201635 (2021).

16. G. B. Marks, *et al.*, A Direct Measure of Tuberculosis Incidence — Effect of Community Screening. *New England Journal of Medicine* **386**, 1380–1382 (2022).

17. G. W. COMSTOCK, V. T. LIVESAY, S. F. WOOLPERT, THE PROGNOSIS OF A POSITIVE TUBERCULIN REACTION IN CHILDHOOD AND ADOLESCENCE. *American Journal of Epidemiology* **99**, 131–138 (1974).

18. J. R. Campbell, N. Winters, D. Menzies, Absolute risk of tuberculosis among untreated populations with a positive tuberculin skin test or interferon-gamma release assay result: systematic review and meta-analysis. *BMJ* **368**, m549 (2020).

19. R. Ragonnet, *et al.*, Optimally capturing latency dynamics in models of tuberculosis transmission. *Epidemics* **21**, 39–47 (2017).

20. Z. Wang, *et al.*, US Postarrival Evaluation of Immigrant and Refugee Children with Latent Tuberculosis Infection Diagnosed Overseas, 2007-2019. *J Pediatr* **245**, 149-157.e1 (2022).

21. K. D. Dale, *et al.*, Overlooked, dismissed, and downplayed: reversion of Mycobacterium tuberculosis immunoreactivity. *Eur Respir Rev* **33**, 240007 (2024).

22. K. C. Horton, A. S. Richards, J. C. Emery, H. Esmail, R. M. G. J. Houben, Reevaluating progression and pathways following Mycobacterium tuberculosis infection within the spectrum of tuberculosis. *Proc Natl Acad Sci U S A* **120**, e2221186120.

23. A. S. Richards, *et al.*, Quantifying progression and regression across the spectrum of pulmonary tuberculosis: a data synthesis study. *The Lancet. Global Health* **11**, e684 (2023).

24. D. W. Dowdy, M. A. Behr, Are we underestimating the annual risk of infection with Mycobacterium tuberculosis in high-burden settings? *The Lancet Infectious Diseases* **0** (2022).

25. N. Kiwanuka, *et al.*, Incidence of tuberculous infection in an African city with endemic tuberculosis: A population-based study. *Int J Tuberc Lung Dis* **28**, 266–272 (2024).

26. R. Wood, *et al.*, Changing prevalence of tuberculosis infection with increasing age in high-burden townships in South Africa. *Int J Tuberc Lung Dis* **14**, 406–412 (2010).

27. R. M. G. J. Houben, P. J. Dodd, The Global Burden of Latent Tuberculosis Infection: A Re-estimation Using Mathematical Modelling. *PLOS Medicine* **13**, e1002152 (2016).

28. World Health Organization, Global Tuberculosis Report 2023. (2023).

29. Indian Council of Medical Research (ICMR), *et al.*, National TB Prevalence Survey in India (2019-2021). (2022).

30. V. K. Chadha, *et al.*, Trends in the annual risk of tuberculous infection in India. *Int J Tuberc Lung Dis* **17**, 312–319 (2013).

31. S. H. Ferebee, F. W. Mount, Tuberculosis Morbidity in a Controlled Trial of the Prophylactic Use of Isoniazid among Household Contacts1. *American Review of Respiratory Disease* **85** (1962).

32. I. Sutherland, The ten-year incidence of clinical tuberculosis following “conversion” in 2550 individuals aged 14 to 19 years. (1968).

33. A. E. Raftery, L. Bao, Estimating and Projecting Trends in HIV/AIDS Generalized Epidemics Using Incremental Mixture Importance Sampling. *Biometrics* **66**, 1162–1173 (2010).

Figure 1: Comparison of selected outputs from models of *M. tuberculosis* transmission against empirical calibration targets

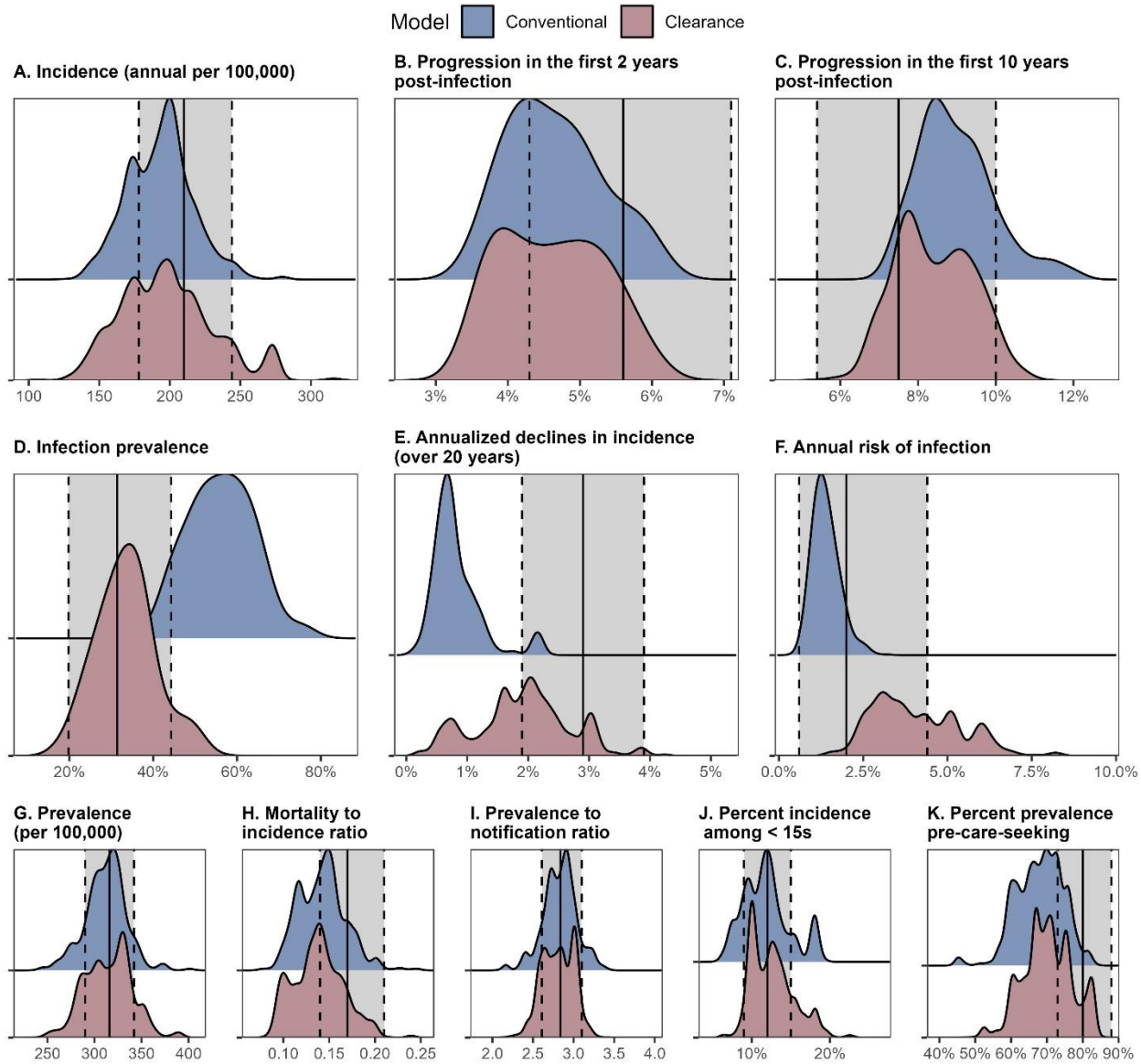
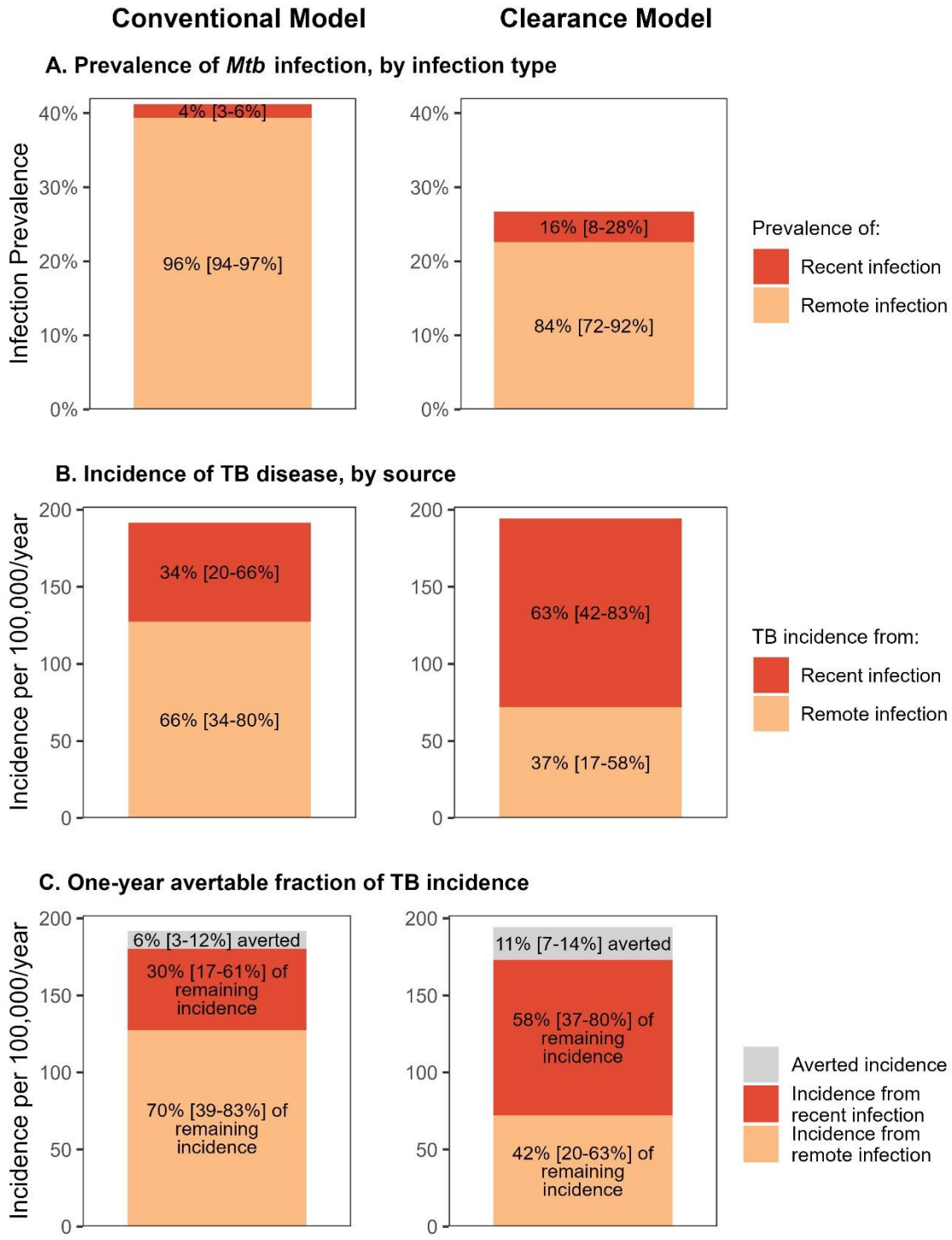


Figure shows the posterior distributions of 11 modeled outputs (shaded density plots) against corresponding calibration targets (mean values indicated by solid black vertical lines, 95% confidence intervals indicated by grey shading with dashed vertical lines). The 2 models differ by whether they include infection clearance (blue = Conventional model with no clearance; red = Clearance model allowing for such clearance). Panels A through F show the 6 targets which had the greatest influence on infection dynamics (see Fig. S10) and for which fit varied the most between the Conventional and Clearance models, to illustrate differences between the models. Panels G through K show the remaining 5 targets to which the models were calibrated, but for which posterior distributions varied less between the Conventional and Clearance models.

Figure 2 Effect of considering immune clearance on modeled TB infection dynamics



This figure illustrates how modeling *Mtb* infection clearance affects modeled infection and disease dynamics. Panel A demonstrates that, without a mechanism for immune clearance ("Conventional Model"), people who were infected remotely (i.e., more than two years ago) account for a larger absolute population, and also a larger fraction of those infected with *Mtb*. Because calibration data shared by both models results in similar progression rates after infection, a greater prevalence of recent *Mtb* infection

results in the Clearance model estimating that a larger proportion of incidence arises from recent (vs. remote) infection than the Conventional model (Panel B). Since interventions to diagnose and treat people with current TB disease will only affect the burden of recent infection, the Clearance model estimates a larger impact from such interventions (Panel C). In panel C, the “one-year avertable fraction” represents the fraction of new incident disease in a given year that would be averted if all people with prevalent TB at the start of that year were instantaneously diagnosed and treated at year-start. All results show means across 50,000 model simulations, with corresponding 95% uncertainty intervals in parentheses.

Figure 3: Projected impact of active case-finding and mass TPT provision

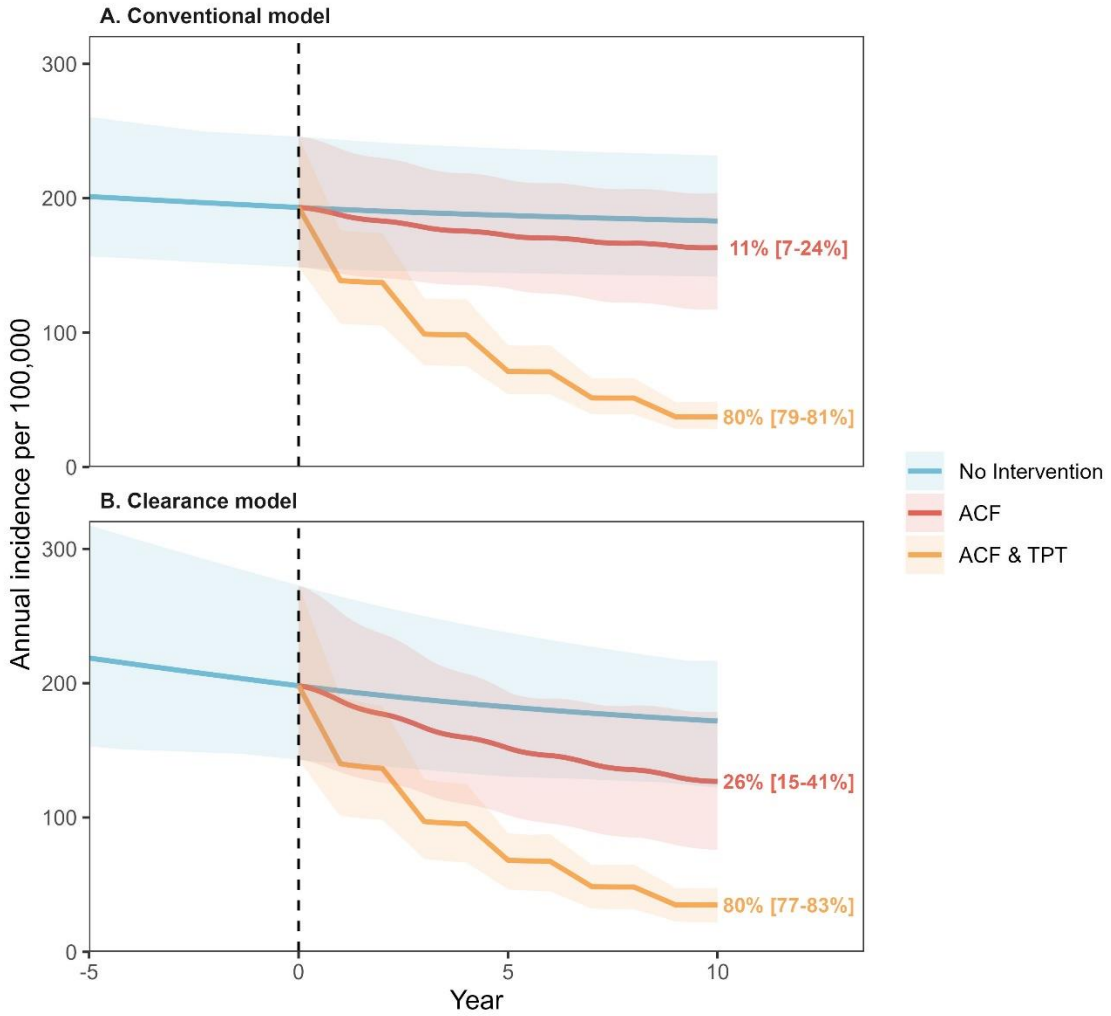
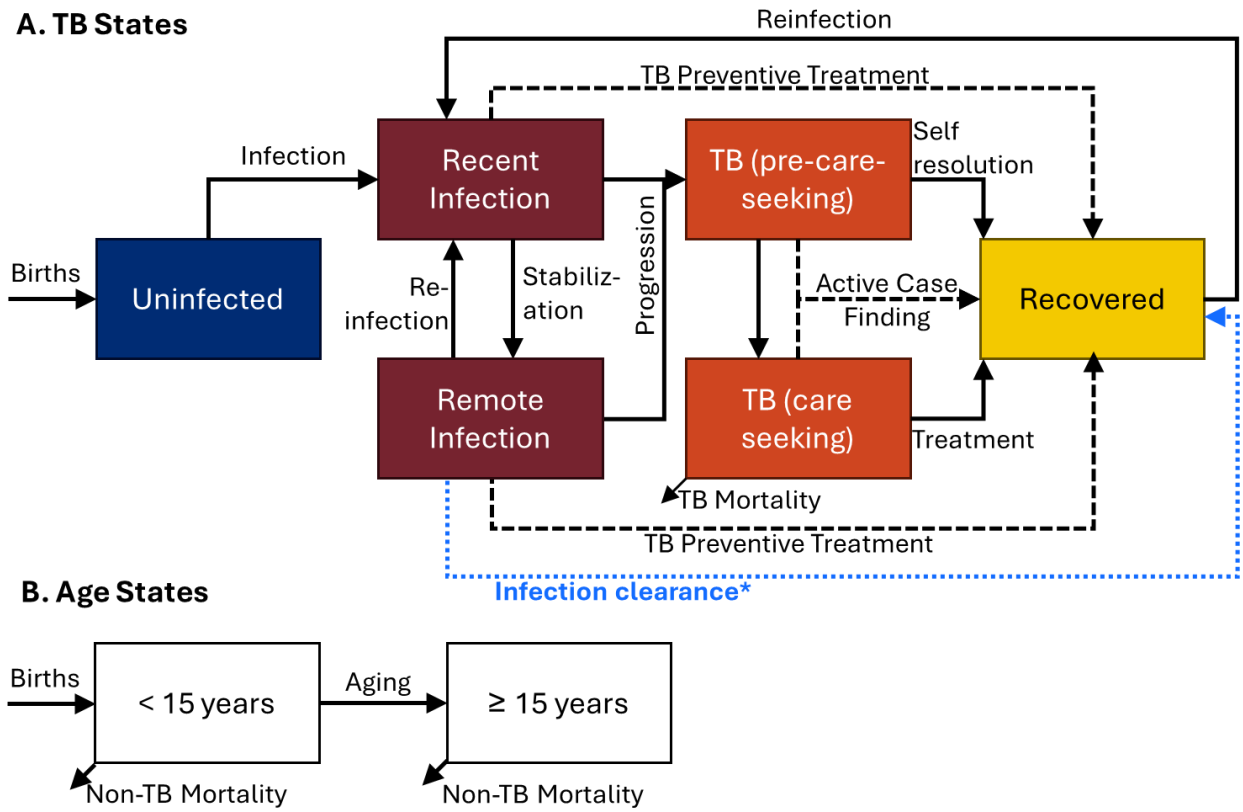


Figure shows the projected the mean annual TB incidence per 100,000 people (solid lines), with 2.5th and 97.5th percentiles (shaded ribbons), over ten years of simulated interventions (years 0 through 10), including: no intervention (blue), Active Case-Finding/ACF waves every 2 years (red-orange), adding TPT to the ACF waves (yellow). Despite overlapping uncertainty ribbons due to variation in baseline incidence across posterior parameter sets, all simulations projected a reduction in incidence from ACF and from ACF with TPT. Text labels indicate the percent reduction in incidence at year 10, relative to no intervention (text shows means, with 95% uncertainty intervals in brackets). Panels differ by whether the model structure includes infection clearance (panel A = conventional model without clearance; panel B = with clearance). Incidence trends in the 5 years preceding the intervention period are also shown in each panel (years -5 through 0).

Figure 4: Model structure



In this model diagram, boxes represent TB-related (panel A) and age-related (panel B) states, while arrows represent transitions between states. Intervention-related transitions (resulting from active case-finding or provision of TB preventive treatment) are shown via black dashed lines. All individuals enter the model uninfected and in the younger (< 15 years) age stratum. TB mortality occurs from the care-seeking TB state only, while non-TB mortality occurs from all states. The stabilization rate from recent to remote infection is 0.5 per year, implying that the average person with TB infection spends around 2 years with elevated progression risk before transition to remove infection, where progression risk is lower. Apart from the stabilization rate, birth and non-TB mortality rates, and intervention-related parameters, all parameters were calibrated (details in the text). Infection clearance (blue dashed arrow from Remote Infection to Recovered) was included only in the Clearance model, not the Conventional model. Additional details are available in SI Text 1 and Fig. S1.

Table 1: Calibration targets

Target	Value [95% CI]	Sources
Annual incidence per 100,00 (2021)	210 [178-244]	WHO Global TB Report (28)
Mortality to incidence ratio (2021)	0.17 [0.14-0.21]	
Percent of incidence among < 15s (2021)	12% [9-15%]	
Annual % reduction in incidence (2002-21)	2.9% [1.9-3.9%]	
Prevalence per 100,000, ≥15 years (2019-21)	316 [290-342]	India prevalence survey (29)*
Prevalence to notification ratio, ≥15 years (2019-21)	2.84 [2.61-3.10]	
Percent of prevalence pre-care-seeking, ≥ 15 years (2019-21)	80% [73-88%]	
Mtb infection prevalence, ≥ 15 years (2019-21)	31% [20-44%]	
Annual risk of infection (2009-10)	2% [0.5-4.3%]	Chadha et al. 2013 (30) (see SI Text 1)
Cumulative progression (2 years since infection)	5.6% [4.3-7.1%]	Ferebee & Mount 1962 and Sutherland 1968, reported in Menzies et al. 2018 (11, 31, 32)
Cumulative progression (10 years since infection)	7.5% [5.4-10.0%]	

Likelihood distributions for each calibration target are available in Table S2.

*The prevalence-to-notification ratio was reported in the India prevalence survey report, but was calculated from a combination of prevalence survey data and national notification data.