

## Research



**Cite this article:** Emery JC, Richards AS, Dale KD, McQuaid CF, White RG, Denholm JT, Houben RMGJ. 2021 Self-clearance of *Mycobacterium tuberculosis* infection: implications for lifetime risk and population at-risk of tuberculosis disease. *Proc. R. Soc. B* **288**: 20201635.

<https://doi.org/10.1098/rspb.2020.1635>

Received: 8 July 2020

Accepted: 30 November 2020

**Subject Category:**

Development and physiology

**Subject Areas:**

computational biology, health and disease and epidemiology, theoretical biology

**Keywords:**

*Mycobacterium tuberculosis*, infection, self-clearance, epidemiology, mathematical modelling

**Author for correspondence:**

Jon C. Emery

e-mail: [jon.emery@lshtm.ac.uk](mailto:jon.emery@lshtm.ac.uk)

Electronic supplementary material is available online at <https://doi.org/10.6084/m9.figshare.c.5238530>.

# Self-clearance of *Mycobacterium tuberculosis* infection: implications for lifetime risk and population at-risk of tuberculosis disease

Jon C. Emery<sup>1</sup>, Alexandra S. Richards<sup>1</sup>, Katie D. Dale<sup>2,3</sup>,  
C. Finn McQuaid<sup>1</sup>, Richard G. White<sup>1</sup>, Justin T. Denholm<sup>2</sup>  
and Rein M. G. J. Houben<sup>1</sup>

<sup>1</sup>TB Modelling Group, TB Centre and Centre for Mathematical Modelling of Infectious Diseases, Department of Infectious Disease Epidemiology, London School of Hygiene & Tropical Medicine, London WC1E 7HT, UK

<sup>2</sup>Victorian Tuberculosis Program, Melbourne Health, Melbourne, Victoria, Australia

<sup>3</sup>Department of Microbiology and Immunology, The University of Melbourne, Melbourne, Victoria, Australia

**id** JCE, 0000-0001-6644-7604; ASR, 0000-0001-9886-8104; KDD, 0000-0002-0655-1685; CFM, 0000-0001-6199-0931; RGW, 0000-0003-4410-6635; JTD, 0000-0002-9214-6431; RMGJH, 0000-0003-4132-7467

**Background:** it is widely assumed that individuals with *Mycobacterium tuberculosis* (*Mtb*) infection remain at lifelong risk of tuberculosis (TB) disease. However, there is substantial evidence that self-clearance of *Mtb* infection can occur. We infer a curve of self-clearance by time since infection and explore its implications for TB epidemiology. **Methods and findings:** data for self-clearance were inferred using post-mortem and tuberculin-skin-test reversion studies. A cohort model allowing for self-clearance was fitted in a Bayesian framework before estimating the lifetime risk of TB disease and the population infected with *Mtb* in India, China and Japan in 2019. We estimated that 24.4% (17.8–32.6%, 95% uncertainty interval (UI)) of individuals self-clear within 10 years of infection, and 73.1% (64.6–81.7%) over a lifetime. The lifetime risk of TB disease was 17.0% (10.9–22.5%), compared to 12.6% (10.1–15.0%) assuming lifelong infection. The population at risk of TB disease in India, China and Japan was 35–80% (95% UI) smaller in the self-clearance scenario. **Conclusions:** the population with a viable *Mtb* infection may be markedly smaller than generally assumed, with such individuals at greater risk of TB disease. The ability to identify these individuals could dramatically improve the targeting of preventive programmes and inform TB vaccine development, bringing TB elimination within reach of feasibility.

## 1. Introduction

Tuberculosis (TB) remains the largest cause of death by an infectious agent [1], with one-quarter of the global population estimated to have been infected with *Mycobacterium tuberculosis* (*Mtb*) [2]. It is commonly assumed that all such individuals retain a lifelong viable infection, defined here as being at risk of TB disease in the absence of treatment and reinfection [3,4]. This is unlikely to be true, however [5,6].

A range of evidence suggests that a proportion of an initially infected cohort may self-clear their infection, defined here as meaning that their risk of TB disease in the absence of treatment and reinfection becomes effectively zero [6–9]. Yet, there is also historical and recent evidence for a long-term rate of TB disease that persists for many years after infection [10–14].

The implications of these two observations for TB prevention are numerous. Tackling incident TB disease arising from the *Mtb* infected reservoir is necessary

to meet World Health Organisation End TB Strategy targets [2,15,16]. However, TB prevention policies are hampered by the size of the infected reservoir as estimated by current tests, which are sensitive to historical infection and not necessarily current, viable infections [5,17]. In combination with the relatively low individual risk of TB disease and the costs and potential side-effects of preventive therapy, the cost-benefit of mass testing and treatment for infection using current tests is often unacceptable at both the individual and population levels [18].

If, however, we consider that self-clearance of *Mtb* infection is possible, our estimates for the population at risk of TB disease in the absence of (re)infection could be significantly reduced, as many individuals initially infected with *Mtb* may no longer harbour a viable infection. Moreover, because the number of people progressing to TB disease would remain unchanged, the decreased population at-risk would lead to a greater lifetime risk of TB disease in those that retain a viable infection. This would, in turn, improve the cost-benefit threshold of preventive therapy should a test for viable infection become available.

To adequately capture the process and implications of self-clearance in a population over time, and in turn, motivate research for tests of viable infection, it is crucial to provide quantitative estimates of the extent of self-clearance by time since infection. While an estimate of the highest potential proportion of individuals that may have self-cleared their infection has been provided [6], an estimate of the proportion of individuals that self-clear by time since infection is currently missing.

To infer this metric from empirical studies, data need to include an estimate for the time of initial infection, or at least enable a reasonable estimate. Examples include tuberculin skin test (TST) reversions among cohorts of initial TST-converters [8,19] as well as an absence of viable bacilli in people with a history of, but died of causes other than, TB [7,20]. As this inference relies on subjective interpretation of indirect measures of self-clearance, we consider it important to make conservative assumptions to provide a strong lower bound for self-clearance over time.

In this paper, we use a modelling approach to provide this conservative estimate of the extent of self-clearance of *Mtb* infection and its impact on TB epidemiology. We review the potential evidence to inform a cohort model of TB natural history before estimating two key outcomes: the increase in lifetime risk of TB disease following infection for those who retain a viable infection; and the decrease in the population with a viable infection in three epidemiologically distinct settings of India, China and Japan.

## 2. Methods

### (a) Data for self-clearance of *Mycobacterium tuberculosis* infection

To find suitable evidence, we considered reviews of the spectrum of *Mtb* infection and TB disease published in the last decade [17,21–29] before extracting references seemingly pertinent to self-clearance of infection and references therein. We also considered specific references [6,30–36] known to the authors and references therein. We only considered studies in humans in which individuals received no chemotherapeutic treatment or Bacillus Calmette-Guérin (BCG)-vaccination (see the electronic supplementary material for further details of the literature review).

Following the review, two principal sources of evidence were found to be suitable for inferring self-clearance by time since

infection. Firstly, TST-reversion studies, where we interpret a transition from TST-positive to TST-negative as waning of the adaptive immune response as a result of clearance of *Mtb* in the host. Studies had to report a time of initial TST conversion (i.e. becoming positive on a TST) or be among initially TST-positive children to provide a reasonable prior for the age, and therefore, time of infection. Individuals then had to be retested after a stated time interval. We inferred the proportion of individuals that cleared their infection from the proportion that TST-reverted upon retest. In line with our aim, this would probably represent a lower bound for self-clearance, because it is possible that some individuals may retain an immune response in the absence of a viable infection [6].

Secondly, we used autopsy studies that attempted to identify viable *Mtb* bacilli in the lungs of individuals that had histopathology consistent with a historic active infection but died of causes other than TB. We inferred the proportion of individuals that cleared their infection from the proportion in which no viable bacilli could be identified via culture and/or guinea pig inoculation. Historical annual risk of infection (ARI) estimates were then used to provide a reasonable prior on the average age at first infection, and hence the average time since first infection, for each age group (see the model description in the following section for further details).

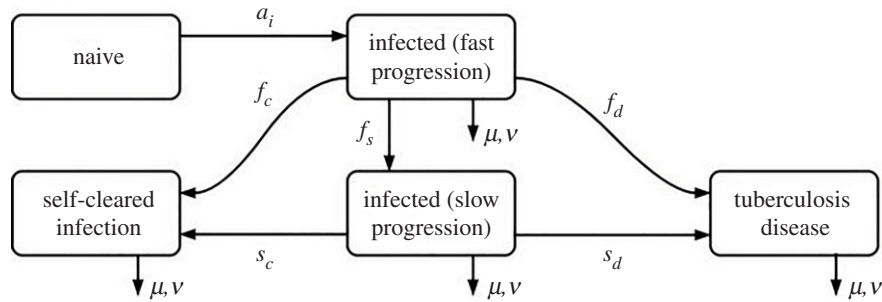
Data from eligible studies were reviewed for potential further biases that could lead to an overestimate of the extent of self-clearance and, where necessary, we made the more conservative selection. To mitigate against small, chance reversions caused by test instability [37], we selected only transitions from indurations of 10 mm or more to 5 mm or less. Similarly, to avoid biases caused by potential TST boosting or persistent non-converters [38], we opted for studies with fewest repeat TSTs. Finally, to provide more robust estimates, the age-group specific results were only extracted if the group consisted of at least 30 individuals.

Self-clearance can also be estimated under certain assumptions from the number of TST-positive individuals that do not develop TB disease following immunosuppression [6]. We did not include such data, however, because estimating the average time since infection is more problematic for contemporary immunosuppression studies, where transmission is highly heterogeneous both geographically and temporally [39]. Moreover, because most of these cohorts have comorbidities sufficient to require immunosuppression, individuals with a viable infection are more likely to have progressed to TB disease before being immunosuppressed. Overall, while such studies can, therefore, provide a useful upper bound for self-clearance of infection in the general population, they are at odds with our aim to provide a robust lower bound of self-clearance by time since infection.

### (b) Cohort model and lifetime risk of tuberculosis disease

A simple deterministic, compartmental model of TB natural history was used to represent a cohort of simultaneously infected people (see figure 1 and table 1). All individuals are modelled to be infected at a given age before either progressing rapidly to TB disease through the ‘fast progression’ infected compartment or more gradually through the ‘slow progression’ infected compartment, as is common in models of TB natural history [3,4]. Conversely, individuals may clear their infection, again through either of the two infected compartments. From here, they are no longer at risk of TB disease in the absence of reinfection. Background mortality is included to account for death from causes other than TB. For simplicity, and to retain conservative estimates, we did not include reinfection.

The model was simultaneously fitted to data on the proportion of individuals that self-cleared infection by time since birth or progressed to TB disease by time since infection. Data for self-clearance was inferred from the studies identified in the preceding



**Figure 1.** Cohort model of TB natural history allowing for self-clearance of *Mtb* infection. A description of the model is provided in the main text and the parameters are detailed in table 1. Note that  $a_i$  can represent either the age at infection of the TST-reversion cohort ( $a_{\text{TST}}$ ) or the autopsy cohort ( $a_{\text{AUT}}$ ). *Mtb*, *Mycobacterium tuberculosis*; TST, tuberculin skin test.

subsection [7,8] (see the results section and electronic supplementary material for further details). To account for potentially different ages of first infection between the studies, the model was simultaneously fitted to TST-reversion and autopsy data with the same natural history parameters but independent ages of infection. Data for progression to TB disease was taken from the placebo arm of the British Medical Research Council's BCG vaccine trials [10,40,41]. For greater than 10 years post-infection, we used a distal progression risk of 20 per 100 000 per year applied to the BCG vaccine placebo group as if infection remains lifelong (which is how distal progression risks are measured and presented) and is representative of the values found in the literature [42–44] (see the electronic supplementary material for further details). Background mortality was modelled using a gamma distribution, described by a mean life expectancy of 60 years with a standard deviation of 10 years. These values are broadly representative of current survival curves in low- and middle-income countries [45] and more likely to reflect the survival curves in the historical studies used. Table 1 summarizes the model parameters.

Model fitting was performed in a Bayesian framework using a flat prior over 0–7 years for the age at infection for the TST-reversion cohort, as detailed in the associated study [8]. The prior for the age at infection for the autopsy cohort was calculated by assuming a normal distribution for the contextual ARI of the study [7], informed by historical TST surveys among adolescents (see the electronic supplementary material for further details). A flat prior was used for all other varied model parameters.

A beta function was used to characterise the likelihood function comparing model outcomes with data on self-clearance and progression to TB disease. A delayed-rejection, adaptive-Metropolis Markov chain Monte Carlo (MCMC) algorithm was used to generate posterior estimates for the model parameters. Each fitting procedure consisted of a 500 iteration burn in, followed by a chain of 5000 further iterations. Results for the median and equal-tailed 95% uncertainty intervals for the posterior model parameters were then generated from these chains.

The model was then run 125 times, sampling from the posterior model parameters. The median and equal-tailed 95% uncertainty intervals for the proportion of the cohort that self-clear infection and the cumulative risk of TB disease among individuals that retain a viable infection, both over time since infection, were then calculated. To arrive at the latter, the risk of TB disease during each time step is calculated by dividing the number of individuals that progressed to TB disease during the time step by the number of individuals that had a viable infection at the beginning of the time step, before integrating over the desired timescale following infection.

For purposes of comparison, the above analysis was repeated assuming lifelong infection (i.e. the self-clearance rates  $f_c$  and  $s_c$  were set to 0 per year) and the model was fitted to progression to TB disease data only. To make appropriate comparisons, we use as an input the posterior results for the age at infection from the self-clearance scenario.

Finally, independent sensitivity analyses were performed to assess the impact of the assumed risk of distal progression to TB disease (using 5 and 35/100 000  $\text{yr}^{-1}$ ) and the mean life expectancy of the cohort (using 50 and 70 years).

### (c) Country-level model and population at risk of tuberculosis disease

To estimate the age-specific population with a viable *Mtb* infection in a given country for a particular year, we applied the cohort model to consecutive 5-year birth cohorts, analogous to the approach for estimating the global burden of latent TB infection in [2] (see the electronic supplementary material for further details). A time-dependent force of infection was applied to each cohort using ARI estimates, which were derived by re-performing the Gaussian process regression in [2]. We used only the median estimated ARI in each country, because the focus of this work is to illustrate the relative difference between the population with a viable infection in the self-clearance and lifelong infection scenarios, not estimates of absolute numbers. To parametrize the TB natural history components of the model, the posterior parameter values derived from the single cohort model were used. Uncertainty in the results was then solely owing to that of the natural history parameters.

We applied the model to three epidemiologically distinct settings: India, China and Japan. Japan, for example, has an older population and less recent transmission compared to China, and in turn compared to India. All three countries also have a sufficiently low prevalence of HIV infection as to not require modification of the TB natural history components of the model presented in figure 1.

For each setting, the model was run 125 times, sampling from the posterior model parameters. The median and equal-tailed 95% uncertainty intervals for the population with a viable *Mtb* infection in 2019 was calculated, by age as well as overall. For purposes of comparison, the above analysis was repeated assuming lifelong infection.

All analyses were conducted using R v. 3.5.0 [46]. Modelling and Bayesian fitting were performed using the deSolve [47] and FME packages [48], respectively. Plots were constructed using the ggplot2 [49] package. Replication data and analysis scripts are available on GitHub.

## 3. Results

### (a) Data for self-clearance of *Mycobacterium tuberculosis* infection

In total, four studies met our initial criteria for inferring self-clearance by time since infection, two of which were



**Table 1.** Cohort-model parameters, descriptions and estimated posterior values for the self-clearance and lifelong *Mtb* infection scenarios. (Values are the percentiles taken from the estimated marginal posterior parameter distributions. Parameter symbols are illustrated in the model diagram in figure 1.)

parameter	description (units)	self-clearance of <i>Mtb</i> infection			lifelong <i>Mtb</i> infection		
		50%	2.5%	97.5%	50%	2.5%	97.5%
$f_d$	rate of progression to TB disease from infected (fast progression) (per year)	0.0859	0.0696	0.111	0.0832	0.0676	0.116
$f_s$	rate of transition from infected (fast progression) to infected (slow progression) (per year)	0.919	0.717	1.25	0.870	0.674	1.26
$s_d$	rate of progression to TB disease from infected (slow progression) (per year)	0.00622	0.000135	0.00113	0.00269	0.000416	0.000525
$s_c$	rate of self-clearance from infected (slow progression) (per year)	0.0330	0.0254	0.0467	0.00 <sup>a</sup>	0.00 <sup>a</sup>	0.00 <sup>a</sup>
$f_c$	rate of self-clearance from infected (fast progression) (per year)	$1.40 \times 10^{-6}$	$2.90 \times 10^{-7}$	$2.61 \times 10^{-6}$	0.00 <sup>a</sup>	0.00 <sup>a</sup>	0.00 <sup>a</sup>
$a_{AUT}$	age at which autopsy cohort is first infected (years)	14.1	8.97	18.3	14.1 <sup>a</sup>	8.97 <sup>a</sup>	18.3 <sup>a</sup>
$a_{TST}$	age at which TST-reversion cohort is first infected (years)	4.92	0.971	6.88	4.92 <sup>a</sup>	0.971 <sup>a</sup>	6.88 <sup>a</sup>
$\mu$	mean background life expectancy (years)	60.0 <sup>a</sup>	60.0 <sup>a</sup>	60.0 <sup>a</sup>	60.0 <sup>a</sup>	60.0 <sup>a</sup>	60.0 <sup>a</sup>
$\nu$	standard deviation of background life expectancy (years)	10.0 <sup>a</sup>	10.0 <sup>a</sup>	10.0 <sup>a</sup>	10.0 <sup>a</sup>	10.0 <sup>a</sup>	10.0 <sup>a</sup>

<sup>a</sup>Denotes input values. TB, tuberculosis; *Mtb*, *Mycobacterium tuberculosis*; TST, tuberculin skin test.

TST reversion studies [8,19] and the other two autopsy studies [7,20].

Of the TST-reversion studies, Adams *et al.* [19] included a significant proportion of individuals that could have undergone at least five TSTs in under 2 years. Such a high rate of testing could result in bias owing to effects such as boosting [38]. To provide the more conservative estimate, we use only the results of Ferebee [8], in which initially TST-positive children were retested after a 10-year interval. Figure 2*a* shows this data presented over time since infection. Uncertainty in the horizontal axis reflects the 95% uncertainty interval of the estimate for the age of infection of the TST-reversion cohort, while the uncertainty in the vertical axis represents the equal-tailed 95% confidence interval of the data.

Of the two autopsy studies, Feldman & Baggenstoss [20] considered only the lesions found in the lung and their surrounding tissues, while Opie & Aronson [7] considered both the lesions and the bulk of the lung. This is the likely reason why Opie & Aronson identified viable *Mtb* bacilli in a markedly greater proportion of individuals, and to provide a more conservative estimate of self-clearance, we extracted data from this study only. Figure 2*a* shows this data presented over time since infection. The data have been aggregated into three age groups, where only groups with greater than 30 individuals have been included. The uncertainty in the horizontal axis reflects the 95% uncertainty interval of the estimate for the age of infection of the autopsy cohort, while the uncertainty in the vertical axis represents the equal-tailed 95% confidence interval of the data.

Owing to a paucity of data, we could not further disaggregate by sex, location (e.g. high versus low burden setting), or age at infection.

## (b) Cohort model and lifetime risk of tuberculosis disease

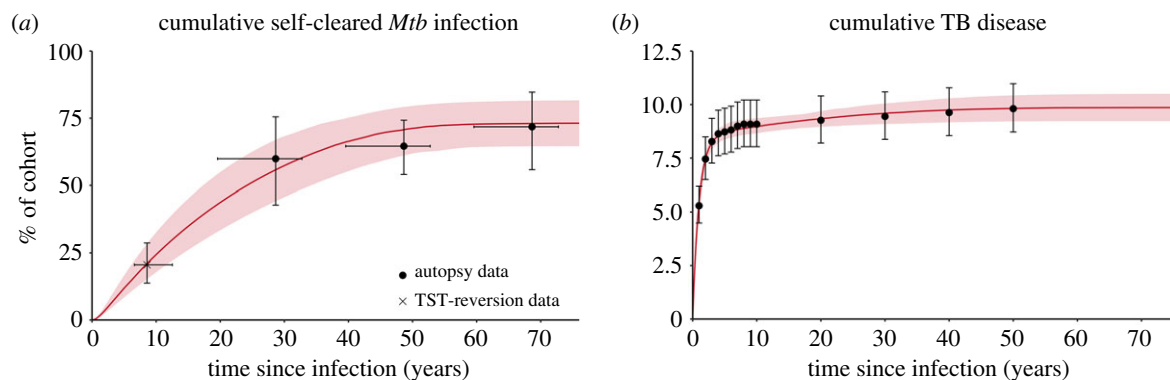
Table 1 shows the median and equal-tailed 95% uncertainty intervals for the posterior model parameters for both the case of self-clearance and lifelong infection. Prior versus posterior model parameters are presented in the electronic supplementary material, table S4.

Figure 2 shows the results of the cohort-model allowing for self-clearance of infection, fitted to the self-clearance and progression to disease data, including the median and equal-tailed 95% uncertainty intervals (see the electronic supplementary material, figure S4 for the model fitting results assuming lifelong infection). As such, we estimate 24.4% (17.8–32.6%, 95% uncertainty interval (UI)) of individuals self-clear within 10 years of infection and 73.1% (64.6–81.7%, 95% UI) over a lifetime.

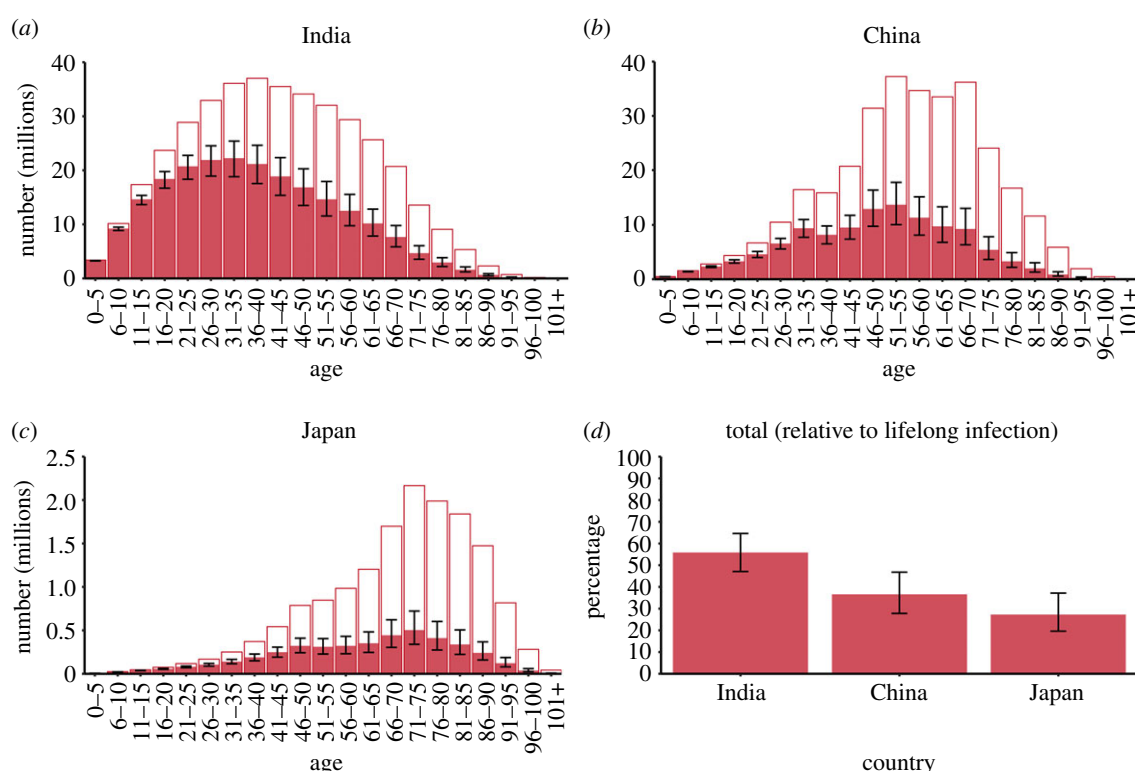
This translates into a lifetime risk of TB disease following infection, in those that retain a viable infection, of 17.0% (10.9–22.5%, 95% UI), compared to 12.6% (10.1–15.0%, 95% UI) in the standard scenario of lifelong infection.

## (c) Country-level models and population at risk of tuberculosis disease

Figure 3 shows the estimated number of people with a viable *Mtb* infection in India, China and Japan in 2019, disaggregated by age. Red outlines show the results assuming lifelong infection and red-filled show the results allowing for self-clearance of infection. Illustrated are the median



**Figure 2.** Cohort model results showing data for self-clearance of *Mtb* infection and progression to TB disease over time since infection. The figure shows the cohort model fitted to data on self-clearance of *Mtb* infection (a) and progression to TB disease (b) presented over time since infection. Data in (a) consist of TST studies (crosses) and autopsy studies (points). Uncertainty in the horizontal axis reflects the 95% uncertainty interval for the estimate for the age of infection. Uncertainty in the data in the vertical axes in both panels are the equal-tailed 95% confidence intervals. Model fits are median model output and equal-tailed 95% uncertainty intervals. *Mtb*, *Mycobacterium tuberculosis*; TB, tuberculosis; TST, tuberculin skin test. (Online version in colour.)



**Figure 3.** Country-level model results for the number of people with a viable *Mtb* infection in the lifelong infection and self-clearance scenarios. (a–c) The estimated population at risk of TB disease in 2019 disaggregated by age in three epidemiologically distinct settings for the cases of lifelong *Mtb* infection (red outlined) and self-clearance (red filled). (d) The total population with a viable infection in each setting, assuming self-clearance, expressed as a percentage of the total population with a viable infection assuming lifelong infection. Point values and error bars show median and equal-tailed 95% uncertainty intervals from the model outputs. Note that median estimates for the historical ARI have been used, such that uncertainty presented in the results is associated with uncertainty in the natural history model parameters only (table 1). For clarity, uncertainty in the lifelong infection results is not shown because they are much smaller than in the self-clearance scenario. *Mtb*, *Mycobacterium tuberculosis*; ARI, annual risk of infection. (Online version in colour.)

and equal-tailed 95% uncertainty intervals. The impact of self-clearance is most pronounced in older age groups.

Figure 3 also shows the total population with a viable *Mtb* infection in each setting allowing for self-clearance of infection, expressed as a percentage of the population with a viable infection assuming infection is lifelong. The impact is smallest in India (56%, 47–65% UI), followed by China (37%, 28–47% UI), with the greatest impact in Japan (27%, 20–37% UI).

See the electronic supplementary material, figures S5–8 and tables S5–12 for the results of sensitivity analyses performed using different mean life expectancies and risks of

distal progression to TB disease. Neither sensitivity makes a qualitative difference to our results, though the results are more sensitive to the assumed life expectancy than the risk of distal progression to TB disease.

## 4. Discussion

We have shown that self-clearance of *Mtb* infection is likely to have a significant impact on TB epidemiology. Our results provide a robust lower-bound estimate, with at least 24.4% (17.8–32.6%, 95% UI) of individuals self-clearing within 10 years of

infection, and 73.1% (64.6–81.7%, 95% UI) over a lifetime. Owing to the smaller population with a viable infection, the lifetime risk of TB disease among individuals retaining a viable infection is 17.0% (10.9–22.5%, 95% UI), a non-trivial increase compared to 12.6% (10.1–15.0%, 95% UI) using the prevailing assumption of lifelong infection. Moreover, the populations with a viable infection across the epidemiologically diverse settings of India, China and Japan are 35–80% (95% UI) smaller in the self-clearance scenario. Reductions are most pronounced in settings with an ageing population and less recent transmission (i.e. Japan), as more time has elapsed for people to self-clear remote infections and there are fewer recent infections.

With respect to related studies, Behr *et al.* recently estimated that over 90% of remotely infected individuals have cleared their infection, among cohorts of TST-positive individuals undergoing immunosuppression [6]. Our work, intended to provide a robust lower bound before exploring the resultant epidemiological implications, finds a lower, albeit still substantial, result for the extent of such self-clearance.

One immediate implication of our work is that a significant proportion of the 1.7 billion people currently estimated to be at risk of TB disease [2] are likely to have cleared their infection, such that the global number should be revised, or at least re-interpreted. Within that reduced number however, the increased risk of developing TB disease would improve the risk/benefit of preventive therapy programmes, particularly if used in conjunction with current methods for identifying high-risk groups [50]. More widely, self-clearance of infection, and whether those that have self-cleared have any protection from future reinfection, will have implications for mathematical modelling of TB natural history and interventions as a whole. Results of pre- versus post-infection novel vaccine candidate models, for example, will need to carefully consider the characteristics of those that have self-cleared their infection [51] as will analyses estimating the impact and cost-effectiveness of preventive therapy [52]. Importantly, however, any potential benefits of self-clearance will depend on developing and validating a test for viable *Mtb* infection that can outperform the positive predictive value of TSTs.

Our analysis has a number of potential limitations primarily owing to the absence of a validated test of viable infection, necessitating instead the indirect inference of self-clearance. With respect to TST-reversion studies, while it seems reasonable that a large reversion implies waning of the sensitization response as a result of clearance of infection, this is not always the case, as in Noguera-Julian *et al.* [53] whereby immunosuppression suppresses TST reactivity. Moreover, test instability can result in small chance reversions [37], which we partially mitigated against by requiring a reversion from an induration greater than 10 mm to one less than 5 mm. The possibility of false negative results remains, however. The lack of reversion in some participants after preventive therapy in a study by Houk *et al.* [54] should also be considered, although groups in the same study that started preventive therapy shortly after TST conversion did exhibit significant reversion thereafter. In either case, isoniazid is known to have limited ability to sterilize (i.e. fully clear) *Mtb* infection, at least in individuals living with human immunodeficiency virus [55], such that TST reversion may not necessarily be expected following Isoniazid preventive therapy alone. Finally, it is possible that some individuals were re-infected after initially self-clearing and reverting their TST during the 10-year follow up in [8], thus underestimating the extent of reversion and self-clearance.

Our use of autopsy studies to infer self-clearance also has certain limitations. Even though we chose the study with the most extensive exploration of the lung, it may be possible that viable *Mtb* bacilli could be found elsewhere. While *Mtb* DNA has been identified outside of the lung [56,57], identification of DNA does not equate to viable bacilli, and it remains unclear as to whether such reservoirs could seed future disease in the lung. Estimating the age of infection of the cohorts relied on ARI estimates for the late nineteenth and early twentieth centuries in the USA. While uncertain, it is accepted that the high ARI during this period was sufficient to assume all cohorts were first infected during adolescence. We also explicitly accounted for the uncertainty in the time since infection in our analysis.

Reinfection will have probably occurred in both data sources given the high background ARI in both contexts, which we did not consider in our modelling. Including reinfection would lead us to increase our estimate for the extent of self-clearance, either owing to secondary conversions among TST-reverters (TST reversion studies), or by estimating the time since last infection as opposed to the time since first infection (autopsy studies). As a consequence, the case is strengthened for our results providing a robust lower-bound estimate of self-clearance.

With respect to our cohort model, we explored parameter uncertainty that could materially alter our results. These were the assumed value for the rate of distal progression to TB disease and the life expectancy of the cohort. Independent sensitivity analyses found that none of these assumptions qualitatively altered our conclusions (see the electronic supplementary material).

Several open questions remain. While we have used TST-reversions to infer self-clearance of infection, in principle, they should also be factored into the ARI estimated from TST survey data among children, as has been considered in [37,58]. For our purposes, while this short-term self-clearance among children would increase the estimated ARI and in turn the number of people at risk of TB disease, the long-term self-clearance of infection in adults would probably outweigh this effect and still lead to a net reduction with respect to the assumption of lifelong infection. The wider issue of incorporating TST-reversions into the standard methodology for estimating ARIs has yet to be addressed, however [59]. Finally, it is unclear how self-clearance affects the protection afforded by previous *Mtb* infection [60], which will be a key input for transmission models intending to include self-clearance.

Our results highlight the urgent need for development and validation of a test for viable *Mtb* infection that can outperform the positive predictive value of TSTs. Such a tool would enable study of the immunology and epidemiology of self-clearance and in turn improve TB vaccine design and the targeting of preventive programmes. While research into correlates of risk to discern those most at risk of TB disease is ongoing [61,62], similar efforts to identify those least at risk would be similarly prudent [63]. Thereafter, in order to understand the impact of self-clearance on transmission and inform vaccine strategy development, it is important to better understand whether, and by how much, self-clearance of infection provides protection from future reinfection.

## 5. Conclusion

Owing to self-clearance of *Mtb* infection, the population with a viable infection may be markedly smaller than

generally assumed, with fewer individuals retaining a viable infection and yet each at greater risk of TB disease. Coupling these wide-ranging implications for TB epidemiology with the ability to identify individuals that have self-cleared could dramatically improve the targeting of preventive programmes and inform TB vaccine design, bringing TB elimination within reach of feasibility.

## References

- WHO. 2019 Global Tuberculosis Report. 2019 Geneva, Switzerland: World Health Organization; 2019. (13 December 2019).
- Houben RMGJ, Dodd PJ. 2016 The global burden of latent tuberculosis infection: a re-estimation using mathematical modelling. *PLoS Med.* **13**, e1002152. (doi:10.1371/journal.pmed.1002152)
- Ragonnet R *et al.* 2017 Optimally capturing latency dynamics in models of tuberculosis transmission. *Epidemics* **21**, 39–47. (doi:10.1016/j.epidem.2017.06.002)
- Menzies NA *et al.* 2018 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**, 228–238. (doi:10.1016/S1473-3099(18)30134-8)
- Mack U *et al.* 2009 LTBI: latent tuberculosis infection or lasting immune responses to *M. tuberculosis*? A TBNET consensus statement. *Eur. Respir. J.* **33**, 956–973. (doi:10.1183/09031936.00120908)
- Behr MA, Edelstein PH, Ramakrishnan L. 2019 Is *Mycobacterium tuberculosis* infection life long? *Br. Med. J.* **367**, 15770. (doi:10.1136/bmj.15770)
- Opie EL, Aronson JD. 1927 Tubercle bacilli in latent tuberculosis lesions and in lung tissue without tuberculosis lesions. *Arch. Pathol.* **4**, 1–21.
- Ferebee SH. 1970 Controlled chemoprophylaxis trials in tuberculosis. A general review. *Bibl. Tuberc.* **26**, 28–106.
- Lin PL *et al.* 2014 Sterilization of granulomas is common in active and latent tuberculosis despite within-host variability in bacterial killing. *Nat. Med.* **20**, 75–79. (doi:10.1038/nm.3412)
- Sutherland I. 1976 Recent studies in the epidemiology of tuberculosis, based on the risk of being infected with tubercle bacilli. *Adv. Tuberc. Res.* **19**, 1–63.
- Canetti G, Sutherland I, Svandova E. 1972 Endogenous reactivation and exogenous reinfection: their relative importance with regard to the development of non-primary tuberculosis. *Bull. Int. Union Tuberc.* **47**, 116–134.
- Vynnycky E, Fine PE. 1997 The natural history of tuberculosis: the implications of age-dependent risks of disease and the role of reinfection. *Epidemiol. Infect.* **119**, 183–201. (doi:10.1017/S0950268897007917)
- Horsburgh CR *et al.* 2010 Revisiting rates of reactivation tuberculosis: a population-based approach. *Am. J. Respir. Crit. Care Med.* **182**, 420–425. (doi:10.1164/rccm.200909-1355OC)
- Lillebaek T *et al.* 2002 Molecular evidence of endogenous reactivation of *Mycobacterium tuberculosis* after 33 years of latent infection. *J. Infect. Dis.* **185**, 401–404. (doi:10.1086/338342)
- Reid MJA *et al.* 2019 Building a tuberculosis-free world: the lancet commission on tuberculosis. *Lancet* **393**, 1331–1384. (doi:10.1016/S0140-6736(19)30024-8)
- Uplekar M *et al.* 2015 WHO's new end TB strategy. *Lancet* **385**, 1799–1801. (doi:10.1016/S0140-6736(15)60570-0)
- M. Pai *et al.*, 2016 Tuberculosis. *Nat. Rev. Dis. Primers* **2**, 16076. (doi:10.1038/nrdp.2016.76)
- Denholm J, Matteelli A, Reis A. 2015 Latent tuberculous infection: ethical considerations in formulating public health policy. *Int. J. Tuberc. Lung Dis.* **19**, 137–140. (doi:10.5588/ijtld.14.0543)
- Adams JM *et al.* 1959 Reversal of tuberculin reaction in early tuberculosis. *Dis. Chest* **35**, 348–356. (doi:10.1378/chest.35.4.348)
- Feldman WH, Baggenstoss AH. 1938 The residual infectivity of the primary complex of tuberculosis. *Am. J. Pathol.* **14**, 473–490.
- Barry CE *et al.* 2009 The spectrum of latent tuberculosis: rethinking the biology and intervention strategies. *Nat. Rev. Microbiol.* **7**, 845–855. (doi:10.1038/nrmicro2236)
- Young DB, Gideon HP, Wilkinson RJ. 2009 Eliminating latent tuberculosis. *Trends Microbiol.* **17**, 183–188. (doi:10.1016/j.tim.2009.02.005)
- Lin PL, Flynn JL. 2010 Understanding latent tuberculosis: a moving target. *J. Immunol.* **185**, 15–22. (doi:10.4049/jimmunol.0903856)
- Esmail H, Barry CE, Wilkinson RJ. 2012 Understanding latent tuberculosis: the key to improved diagnostic and novel treatment strategies. *Drug Discov. Today* **17**, 514–521. (doi:10.1016/j.drudis.2011.12.013)
- Esmail H, Barry CE, Young DB, Wilkinson RJ. 2014 The ongoing challenge of latent tuberculosis. *Phil. Trans. R. Soc. B* **369**, 20130437. (doi:10.1098/rstb.2013.0437)
- Dheda K, Barry CE, Maartens G. 2016 Tuberculosis. *Lancet* **387**, 1211–1226. (doi:10.1016/S0140-6736(15)00151-8)
- Behr MA, Edelstein PH, Ramakrishnan L. 2018 Revisiting the timetable of tuberculosis. *Br. Med. J.* **362**, k2738. (doi:10.1136/bmj.k2738)
- Drain PK *et al.* 2018 Incipient and subclinical tuberculosis: a clinical review of early stages and progression of infection. *Clin. Microbiol. Rev.* **31**, e00021–18. (doi:10.1128/CMR.00021-18)
- Lin PL, Flynn JL. 2018 The end of the binary era: revisiting the spectrum of tuberculosis. *J. Immunol.* **201**, 2541–2548. (doi:10.4049/jimmunol.1800993)
- Keller AE, Kampmeier RH. 1939 Tuberculin survey. Observations on medical students and undergraduate nurses at Vanderbilt University. *Am. Rev. Tuberc.* **39**, 657–665.
- Maclof A. 1946 The yearly persistence of cutaneous sensibility to tuberculin. *Rev. Path. Comparee et Hyg. Gen.* **46**, 241–257.
- Johnston RN, Ritchie RT, Murray IHF. 1963 Declining tuberculin sensitivity with advancing age. *Br. Med. J.* **2**, 720–724. (doi:10.1136/bmj.2.5359.720)
- Mysakowska H *et al.* 1966 An appraisal of the chemoprophylaxis of tuberculosis in the student population of Lublin, Poland, following thirty months' observation. *Am. Rev. Respir. Dis.* **93**, 628–629.
- Sepkowitz KA. 1996 Tuberculin skin testing and the health care worker: lessons of the Prophit Survey. *Tuber. Lung Dis.* **77**, 81–85. (doi:10.1016/S0962-8479(96)90081-7)
- Hawn TR *et al.* 2014 Tuberculosis vaccines and prevention of infection. *Microbiol. Mol. Biol. Rev.* **78**, 650–671. (doi:10.1128/MMBR.00021-14)
- Nduba V, Van't Hoog AH, Mitchell EMH, Borgdorff M, Laserson KF. 2018 Incidence of active tuberculosis and cohort retention among adolescents in Western Kenya. *Pediatr. Infect. Dis. J.* **37**, 10–15. (doi:10.1097/INF.0000000000001685)
- Fine PE *et al.* 1999 Tuberculin sensitivity: conversions and reversions in a rural African population. *Int. J. Tuberc. Lung Dis.* **3**, 962–975.
- Menzies D. 1999 Interpretation of repeated tuberculin tests. Boosting, conversion, and reversion. *Am. J. Respir. Crit. Care Med.* **159**, 15–21. (doi:10.1164/ajrcm.159.1.9801120)
- Public Health England. 2018 Tuberculosis in England – 2018 report. See <https://www.gov.uk/government/publications/tuberculosis-in-england-annual-report>.

**Data accessibility.** Replication data and analyses scripts are available on GitHub at [https://github.com/epidemery/mtb\\_self\\_clearance](https://github.com/epidemery/mtb_self_clearance).

**Authors' contributions.** R.M.G.J.H. conceived research; J.C.E. and R.M.G.J.H. designed the methodology; J.C.E. performed the analysis; J.C.E. and R.M.G.J.H. prepared the original draft of the paper; all authors provided detailed feedback on the methodology and analysis; all authors reviewed and edited drafts of the paper.

**Competing interests.** We declare we have no competing interests.

**Funding.** We received no funding for this study.



40. Sutherland I. 1968 *The ten-year incidence of clinical tuberculosis following 'conversion' in 2,550 individuals aged 14 to 19 at time of conversion*. The Hague, The Netherlands: Tuberculosis Surveillance Research Unit.
41. Sutherland I, Svandova E, Radhakrishna S. 1982 The development of clinical tuberculosis following infection with tubercle bacilli. *Tubercle* **63**, 255–268. (doi:10.1016/S0041-3879(82)80013-5)
42. Dale KD, Trauer JM, Dodd PJ, Houben RMGJ, Denholm JT. 2019 Estimating long-term tuberculosis reactivation rates in Australian migrants. *Clin. Infect. Dis.* **70**, 2111–2118. (doi:10.1093/cid/ciz569)
43. Ronald LA *et al.* 2018 Demographic predictors of active tuberculosis in people migrating to British Columbia, Canada: a retrospective cohort study. *Canadian Med. Assoc. J.* **190**, E209–E216. (doi:10.1503/cmaj.170817)
44. Styblo K. 1991 *Selected papers (Vol 24)*. The Hague, The Netherlands: KNCV.
45. United Nations, Department of Economic and Social Affairs, Population Division. 2019 *World Population Prospects 2019*, Online Edition. (December 18, 2019).
46. R Core Team. 2014 *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. See <http://www.R-project.org/>.
47. Soetaert K, Petzoldt T, Setzer RW. 2010 Solving Differential Equations in R: package deSolve. *J. Stat. Soft.* **33**, 1–28. (doi:10.18637/jss.v033.i03)
48. Soetaert K, Petzoldt T. 2010 Inverse modelling, sensitivity and Monte Carlo analysis in R using Package FME. *J. Stat. Soft.* **33**, 1–28. (doi:10.18637/jss.v033.i03)
49. Wickham H. 2016 *Ggplot2: elegant graphics for data analysis*. New York, NY: Springer.
50. WHO. 2018 *Latent tuberculosis infection: updated and consolidated guidelines for programmatic management*. Geneva, Switzerland: World Health Organization. (17 December 2019).
51. Harris RC *et al.* 2019 Age-targeted tuberculosis vaccination in China and implications for vaccine development: a modelling study. *Lancet Global Health* **7**, e209–e218. (doi:10.1016/S2214-109X(18)30452-2)
52. Johnson KT, Churchyard GJ, Sohn H, Dowdy DW. 2018 Cost-effectiveness of preventive therapy for tuberculosis with isoniazid and rifapentine versus isoniazid alone in high-burden settings. *Clin. Infect. Dis.* **67**, 1072–1078. (doi:10.1093/cid/ciy230)
53. Noguera-Julian A *et al.* 2019 Tuberculosis disease in children and adolescents on therapy with anti-tumor necrosis factor-alpha agents: a collaborative, multi-centre paediatric tuberculosis network European trials group (ptbnet) study. *Clin. Infect. Dis.* **71**, 2561–2569. (doi:10.1093/cid/ciz1138)
54. Houk VN, Kent DC, Sorensen K, Baker JH. 1968 The eradication of tuberculosis infection by isoniazid chemoprophylaxis. *Arch. Environ. Health* **16**, 46–50. (doi:10.1080/00039896.1968.10665013)
55. Houben RMGJ, Sumner T, Grant AD, White RG. 2014 Ability of preventive therapy to cure latent *Mycobacterium tuberculosis* infection in HIV-infected individuals in high-burden settings. *Proc. Natl Acad. Sci. USA* **111**, 5325–5330. (doi:10.1073/pnas.1317660111)
56. Tornack J *et al.* 2017 Human and mouse hematopoietic stem cells are a depot for dormant *Mycobacterium tuberculosis*. *PLoS ONE* **12**, e0169119. (doi:10.1371/journal.pone.0169119)
57. Neyrolles O *et al.* 2006 Is adipose tissue a place for *Mycobacterium tuberculosis* persistence? *PLoS ONE* **1**, e43. (doi:10.1371/journal.pone.0000043)
58. Sutherland T. 1971 The effect of tuberculin reversion upon the estimate of the annual risk of tuberculous infection. *Selected Pap. R. Netherlands Tuberc. Assoc.* **14**, 115–118.
59. Cauthe GM, Pio A, ten Dam HG. 2002 Annual risk of tuberculous infection. 1988. *Bull. World Health Organ.* **80**, 503–511; discussion 501–502.
60. Andrews JR *et al.* 2012 Risk of progression to active tuberculosis following reinfection with *Mycobacterium tuberculosis*. *Clin. Infect. Dis.* **54**, 784–791. (doi:10.1093/cid/cir951)
61. Petruccioli E *et al.* 2016 Correlates of tuberculosis risk: predictive biomarkers for progression to active tuberculosis. *Eur. Respir. J.* **48**, 1751–1763. (doi:10.1183/13993003.01012-2016)
62. Fiore-Gartland A *et al.* 2018 Considerations for biomarker-targeted intervention strategies for tuberculosis disease prevention. *Tuberculosis (Edinb.)* **109**, 61–68. (doi:10.1016/j.tube.2017.11.009)
63. Broderick C. 2019 Transcriptomic responses to preventive therapy for latent tuberculosis infection. In *Conf. TBSCIENCE 2019, Hyderabad, India, 29 October, 2019*.