Estimating the Burden of Tuberculosis in India: A Modelling Study

Sandip Mandal, Raghuram Rao[1], Rajendra Joshi[1]

Senior Advisor Data Analytics and Mathematical Modelling, John Snow Institute, [1]Central TB Division, Ministry of Health and Family Welfare, Govt. of India, New Delhi, India

Abstract

Background: Measurements TB incidence and mortality are crucial for monitoring progress towards SDG goals for TB. Until recently, WHO estimated TB burden in India with applied simple, transparent equilibrium models to data from Gujarat, an Indian state where the first state-level prevalence survey was conducted in 2011. However, since then there has been several interventions in India including national TB prevalence survey, infection survey, sub-national survey & certification which gives opportunity for national and sub-national estimates for TB incidence and mortality. **Methods:** We developed a model is a compartmental, deterministic framework, taking account of TB natural history, as well as India's healthcare system including health care seeking from public and private sector. To address changes in TB burden owing to COVID disruptions, we followed same model that used by WHO in the global TB Report 2022 with additional impact of delta wave in 2021. Major sources of data included National TB Prevalence survey, trends in caseloads in public and private sector including their contribution and mortality information. **Results:** We estimated total TB incidence of 2.77 million in the year 2022 as against 2.97 in the year 2015 and corresponding TB mortality of 0.32 and 0.36 million respectively. In terms of rate per 1,00,000 TB incidence in 2022 was 196 as compared to 225 in the year 2015 and mortality was 23 and 27 respectively. TB incidence estimates are similar to what was estimated by WHO, while mortality estimates appear different in our estimates due to different calibration targets depending on in-country published data. **Conclusion:** Even if TB burden is infeasible to measure directly, a range of data can nonetheless offer indirect evidence for its estimation: mathematical modelling can be a helpful tool for bringing together these diverse sources of evidence, and deriving estimates that are consistent with them all. While the RGI reported mortality is an important source of information, its quality and coverage for medically certified cause of deaths requires improvement in India.

Keywords: Tuberculosis, Incidence, mortality, estimation, India

INTRODUCTION

The END TB Strategy and SDG goals for TB call for 80% reduction in TB incidence rates and 90% reduction in TB deaths in 2030, compared with 2015.[1,2] However, there are substantial challenges in monitoring both of these measures. Direct and representative measurement of TB incidence requires an infeasibly large population to be followed up over a year or more, while in high burden countries, vital registration systems generally lack the coverage to provide reliable estimates of mortality. Therefore, incidence and mortality often must be estimated indirectly, using the available data.

The need for robust TB burden estimates is especially pressing in India, the country with the world's largest TB burden, which in 2021 accounted for over a quarter of global TB incidence.[3] Until recently, WHO estimates of TB burden in India applied simple, transparent equilibrium models to data from Gujarat,

an Indian state where the first state-level prevalence survey was conducted in 2011.[4] Although Gujarat accounts for around 4% of India's population, this was the largest survey of its time in India, and thus provided the best available information on TB burden in the country.

India subsequently embarked on a national prevalence survey, beginning in 2019.[5] Despite a temporary interruption during the first wave of COVID-19 in the country, the survey went on to cover 3,22,480 individuals in 25 different states, placing

Address for correspondence: Dr. Raghuram Rao, Joint Director, TB, Central TB Division, National TB Elimination Programme, Ministry of Health and Family Welfare, Govt. of India, Jeevan Vihar, Ground Floor, Sansad Marg, New Delhi, India.
E-mail: raor@rntcp.org

Access this article online

Quick Response Code:

Website:
www.ijcm.org.in

DOI:
10.4103/ijcm.ijcm_160_23

it amongst the largest prevalence surveys conducted so far. Findings from this survey have already yielded important insights into the distribution of TB burden in a country as vast as diverse as India. However, they also provide an invaluable opportunity to update estimates of TB incidence in India. In addition, recent years have seen a substantial increase in the availability of other, indirect sources of evidence for TB burden, such as systematic tracking of the sales of anti-TB drugs in the private healthcare sector[6,7] and a steady expansion of district-level initiatives to monitor TB control efforts.[8] In the present work, we brought together all of these data sources in a Bayesian evidence synthesis framework, to inform estimates for TB incidence and mortality in India over the last decade.

METHODS

Figure 1 shows an illustration of the model structure, with further details shown in the supporting information. Briefly, the model is a compartmental, deterministic framework, taking account of TB natural history, as well as India's healthcare system. It distinguishes TB care in the public and private sectors, in order to account separately for notifications from both of these sectors. For simplicity the model does not include HIV/TB coinfection or rifampicin resistance, which in 2019 accounted respectively for an estimated 2.3% and 4.2% of estimated TB incidence in the country.[9] Also for simplicity, the model does not incorporate age structure, and does not distinguish between pulmonary and extra pulmonary TB. In this analysis first we calibrated our model with the available data and estimated incidence in 2019. Then, to address changes in TB burden owing to COVID disruptions, we followed the same methodology as that used by WHO in the global TB Report 2022,[10] with one difference: WHO estimations assume a temporary reduction in TB transmission, during periods of lockdowns However in India, mobility data suggests that the

severe 'delta wave' from May to September 2021 showed stark reductions in mixing, as severe as during the national lockdown from April to July 2020, (see Figure S3, supporting information). Accordingly, we assumed a 50% reduction in TB transmission during the period of the delta wave as well as during the national lockdown, assuming (consistent with the WHO model) that TB transmission was at pre-pandemic levels at other times.

Programme data shows that the presumptive examination rate by the public sector has been increasing over time. Accordingly, we allowed the rate-of-presentation to care ($\gamma(t)$) to increase in a linear way from 2011 to 2020, with starting and ending values to be calibrated. We also estimated the probability p that a patient chooses the public (rather than the private) sector on each careseeking visit, again allowing this parameter to change in a linear way from 2011 to 2020.

Sources of data

- National prevalence survey: Results from India's prevalence survey suggested a prevalence of 312 [CI 286 - 337] per 100,000 population, when adjusted for age and other factors. In addition to this data, we also calibrated to the data for the proportion of prevalent TB that was on TB treatment (irrespective of bacteriological status) in the survey. A similar approach had been taken in earlier years for WHO incidence estimates based on Gujarat prevalence survey data, but assuming equilibrium conditions for the purpose of analytical tractability.[11] The role of this data is to inform the average duration of untreated TB (the higher this proportion, the shorter the duration). Thus, combined with prevalence, this data informs incidence estimates. In our current analysis, by virtue of using a dynamical transmission model, we were able to incorporate this data without need for equilibrium assumptions.
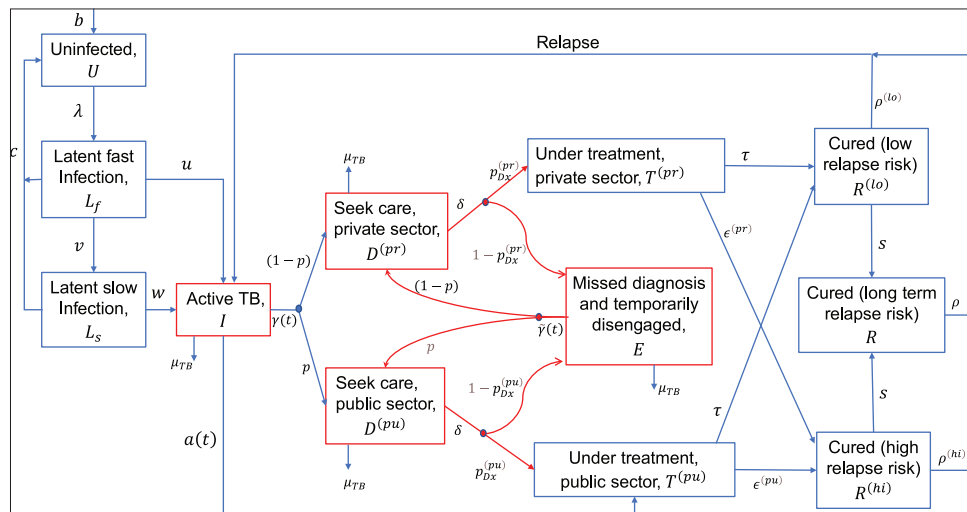
Figure 1: Schematic illustration of the model structure. Infectious compartments contributing the force-of-infection are shown in red. For clarity, the diagram omits certain rates incorporated in the model, including: self-cure; exogenous reinfection; and background mortality. Details of parameters and model equations are given in Table S1 of the supplementary document

- Trends in TB caseload in the private sector: The healthcare data company IQVIA collects comprehensive data on drugs sold through the private market in India, with data available from 2015. Tracking the volume of rifampicin-containing drugs gives a direct measure of the volume of anti-TB treatments being sold through the private sector each year. As described in previous work, it is challenging to translate directly from drug volumes to patient numbers,[12] because two factors necessary for this conversion are unknown: the average duration of TB treatment in the private sector, and the extent of overdiagnosis for TB in this sector. Accordingly, we did not attempt to use drug sales data directly to inform *numbers* of TB patients, but rather to inform *trends*. In particular, private sector sales of TB drugs have declined nationally by 44% from 2015 to 2019. After adjusting for private providers increasingly prescribing publicly supplied drugs in recent years (see supporting information), we calibrated the model to this proportionate decline in the number of patients initiating treatment in the private sector, between 2015 and 2019.

- Trends in TB caseload in the public sector: We calibrated model simulations for treatment initiations in the public sector against data for public notifications in 2011, and 2019. In 2011, patients were notified when initiating treatment and in 2019, when they were diagnosed. We therefore adjusted 2019 notifications downwards by the pre-treatment loss-to-follow up, in order to ensure comparability between these data. We also adjusted 2019 data for notifications that had been inaccurately attributed to the public sector, rather than the private sector.

- Private sector contribution to notifications: Since 2014, notifications from the private sector have steadily grown over time, as a result of successful efforts by the national programme to engage and coordinate with this sector. We did not aim to match these trends in the models, as they arose directly from programmatic efforts rather than any underlying epidemiological changes. Instead, we calibrated model-simulated treatment initiations in the private sector to notifications from this sector in 2019, adjusting for the proportion of privately treated patients that are notified (see supporting information).

- Under-reporting: In recent years, India's TB programme has implemented subnational certification (SNC), a district-level initiative to generate data for monitoring trends in TB burden over time. At the time of writing, this initiative covers more than 500 districts in the country, providing representative data for 21 states. An important part of this initiative is a sustained effort to find TB cases at the community level (see ref [8] for further details). Over the course of a year, all identified TB cases are compared against Nikshay, India's TB register, to quantify the proportion that had been notified. We used this data as a calibration target in the model, for the proportion of TB patients on treatment that were notified in 2019.

- Mortality: Since 2015, WHO estimates of TB mortality in India have been based on estimates by the Institute of Health Metrics and Evaluation (IHME), adjusted to align with WHO estimates for all-cause mortality.[11] Data updates for these estimates were last conducted by IHME in 2015.[12] Additionally, RGI data is another valuable source of mortality data in India.[13] However, across the country coverage of RGI data is incomplete. We therefore incorporated IHME mortality estimates but RGI reported moderated mortality rate for India. In this method, we assumed the central value is the average of the upper limit of RGI mortality and the lower limit of IHME mortality (~ 28 per 100,000 populations). And the extreme end of both these estimates (IHME and RGI) inform us the CI of this moderated mortality. (see supporting information).

- TB infection prevalence: During the prevalence survey, a subset of the population was tested using interferon gamma release assays (IGRA), to estimate the prevalence of TB infection. This prevalence was estimated at around 25%, notably consistent with estimates from a previous statistical study.[14] Although our model parameters for progression to active TB are based on a systematic review of studies from the pre-chemotherapy era, in India undernutrition and other factors are likely to play an important modifying role in these rates. Accordingly, we calibrated the amount by which rates should be increased, in order to match the measured prevalence of TB infection in 2019.

Calibration

To estimate uncertainty systematically from model inputs to model estimates, we performed calibration using Bayesian Markov Chain Monte Carlo (MCMC). We constructed the posterior density as follows: for each of the calibration targets described above, we constructed beta distributions to capture model proportions, and log-normal distributions to capture population rates, adjusting distribution parameters in order to match the central and uncertainty intervals of each calibration target. We also included wide prior uniform distributions for uncertain model parameters, such as the rates of treatment completion in the private sector.[15] We then took the posterior density to be proportional to the product of all likelihood and prior densities. For all practical purposes we calculated the log-posterior density, therefore taking a sum of the log-probability distributions of each of the individual likelihood components.

As an efficient way of sampling from the posterior distribution, we implemented adaptive MCMC,[16] which uses the covariance structure of already-drawn samples to inform the proposal distribution. We first generated 1,000 samples for model parameters using Latin hypercube sampling, then choosing the three parameter sets with the highest posterior density as starting conditions for independent MCMC chains. We ran each chain for 50,000 MCMC iterations. After discarding the burn-in and selecting every 50th sample, we drew 250 samples from the posterior density. For all model outputs, we took central estimates as the 50th percentile. We

quantified uncertainty using the 2.5th and 97.5th percentiles, denoting this range as the 95% Bayesian credible interval. We compared results from the three independent chains to ensure that they gave convergent estimates.

The model was calibrated to the following data:

RESULTS

The results of model calibration are shown in Figure 2. This shows the resulting comparisons between model outputs and indicator data.

Figure 3 shows model projection for incidence and mortality at national level from 2015 to 2022. Results suggest an incidence rate in 2022 of 196 (95% CrI 171 – 228) per 100,000 population and mortality rate of 23 (95% CrI 16 – 33) per 100,000 population. In terms of absolute numbers, these estimates correspond to an incidence and mortality, respectively, of

2.77 million (95% CrI 2.42 – 3.23) and 0.32 million (95% CrI 0.23 – 0.47). The impact of service disruption due to the COVID-19 pandemic is observed in the model estimates. Annual estimates for incidence and mortality from 2015 to 2022 are summarised in Table 1.

Sensitivity analysis

To assess the contribution of each of the model parameter, we performed a sensitivity analysis, using partial rank correlation coefficients. Figure 4 shows the result of that analysis that helps to identify the model inputs that are most influential for model outcomes. The model shows partial rank correlation coefficients of each of the model inputs against the model outcomes - estimated incidence and mortality rates in 2019 [Figure 4a and b]. Additionally, we performed a similar analysis to show how the input parameters correlates with the calibration outcome – public sector notifications [Figure 4c].

Model Calibration indicators

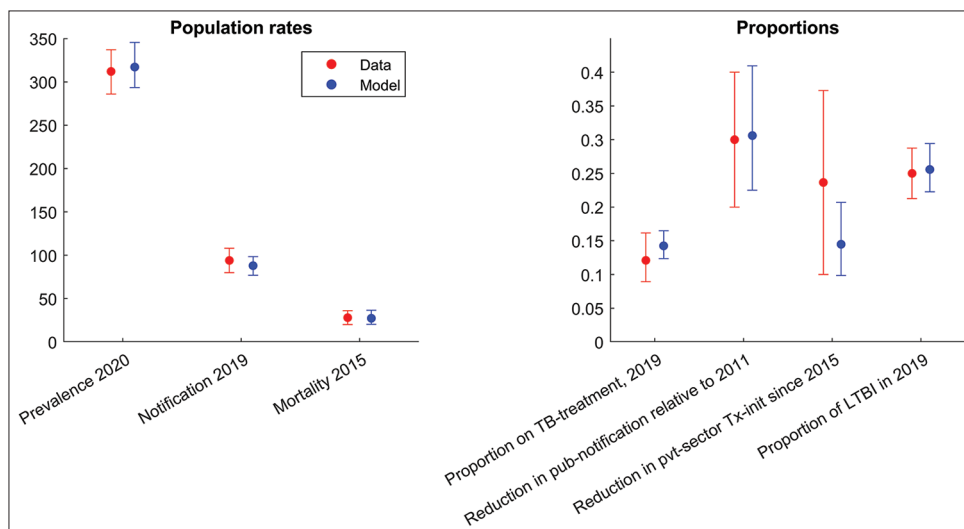| Indicator	Value			Comments and data source		
Prevalence per 100,000 population, 2020	312 (286 – 337)			National TB prevalence survey[v]		
Of prevalent TB, percent on treatment	12 (9.0 – 16)					
Population prevalence of LTBI (percent)	25 (21 – 29)					
Notifications per 100,000 population, 2019 (public sector)	94 (80 – 108)			Programmatic data allowing +/- 10% uncertainty. Because the model only addresses new and relapse cases initiating treatment, here we counted only new and relapse notifications, as well as accounting for 15% initial loss to follow up.		
Private sector treatment initiation in 2019 per 100,000	Greater than 42			Accounting for over-diagnosis in private sector, reducing private notifications by 15%. This target is set as a lower limit, because of uncertainty in the proportion of private providers that are notifying TB.		
Mortality per 100,000 population, 2015	28 (20 – 36)			By adjusting IHME estimates with the RGI data (see supporting information)		
Proportion of reduction of public sector notification in 2019 relative to 2011	0.3 [0.2 – 0.4]			Programmatic data		
Reduction in treatment initiation in private sector in 2019 relative to 2015	0.24 [0.19 – 0.37]			From private-sector drug sales data. Accounting for the increasing proportion of privately managed patients on publicly supplied drugs		
Notification through active case-finding	2017	2018	2019	2020	2021	Programmatic data
	26781	47307	62958	52273	73772	

Figure 2: Model calibration with the observed data. Dots show central estimates, while error bars show 95% uncertainty intervals

Table 1: Estimated incidence and mortality rate per 100,000 populations and in total numbers

Estimates	2015			2016			2017			2018			2019			2020			2021			2022		
	lo	mid	hi	lo	mid	hi	lo	mid	hi	lo	mid	hi	lo	mid	hi	lo	mid	hi	lo	mid	hi	lo	mid	hi
Incidence (per 1,00,000)	201	225	254	196	217	246	189	211	238	182	206	233	177	201	228	169	194	226	172	197	227	171	196	228
Incidence (numbers in million)	2.66	2.97	3.36	2.62	2.91	3.29	2.56	2.86	3.23	2.50	2.82	3.19	2.45	2.77	3.16	2.36	2.70	3.15	2.42	2.77	3.20	2.42	2.77	3.23
Mortality (per 1,00,000)	20	27	36	19	26	35	18	25	34	18	24	33	17	23	32	17	23	32	17	24	34	16	23	33
Mortality (numbers in million)	0.27	0.36	0.48	0.26	0.35	0.47	0.25	0.34	0.46	0.24	0.33	0.45	0.23	0.32	0.44	0.23	0.32	0.44	0.24	0.34	0.48	0.23	0.32	0.47

Figure Inputs are listed in order of decreasing sensitivity from top to bottom.

To compare against previous methods of burden estimation conducted by WHO, we also estimated incidence under a 'reduced' model, constructed to simulate equilibrium epidemic conditions: to do so, we assumed no temporal change in the rate of treatment uptake in the public or private sectors, and calibrated only to data from the prevalence survey, as well as notifications in 2019 and estimated TB mortality. This simplified model estimated incidence in 2020 is 194 [95% CrI 172 - 221] per 100,000 populations which is very close to our estimated incidence in 2020 [See Table 1].

DISCUSSION

Creating robust estimates of TB burden is an important need for public health, especially in light of India's ambitious commitment to end TB by 2025. Even if TB incidence is infeasible to measure directly, our work illustrates how cross-sectional data from prevalence surveys - together with other sources of evidence on TB burden - can provide valuable information. Our estimates for TB incidence in 2019 are comparable to earlier estimates by WHO, based on an equilibrium model, although our estimates of TB incidence decline in recent years are somewhat steeper though close to interim estimates published in recent Global TB report 2022.

In our model, the key sources of data informing trends over time are private sector drug sales, and public sector notifications - respectively, informing TB caseloads in the private and public sectors. As described above, both sources have limitations that mean they cannot directly be used to inform burden estimates: adjustments are needed, for example to extract treatment initiations from notification data, wherever the latter includes all diagnosed patients. Nonetheless, the Bayesian evidence synthesis approach that we have implemented allows us systematically to incorporate these adjustments in the analysis, together with their associated parametric uncertainties. Such approaches could be used in other settings where similar data is available: not only in other countries, but also for individual states within India. Indeed, extending the work shown here to state-level estimates in India is the subject of ongoing analysis, and will contribute to strategic planning for TB control priorities at the state level.

In general, the best approach for any given country will depend on the sources of data available. For example, a recent study used modelling approaches to estimate incidence and mortality in Brazil.[17] Importantly in Brazil, vital registration data provides a comparably robust base of evidence on which not just TB mortality, but also incidence estimates, can be developed. While such comprehensive VR data was not available in India at the time of these estimates, it is steadily developing. Currently available information on vital registration of deaths in India suggests increasing level of reporting, >90% of estimated deaths under vital registration system of Registrar General of India. However, coverage
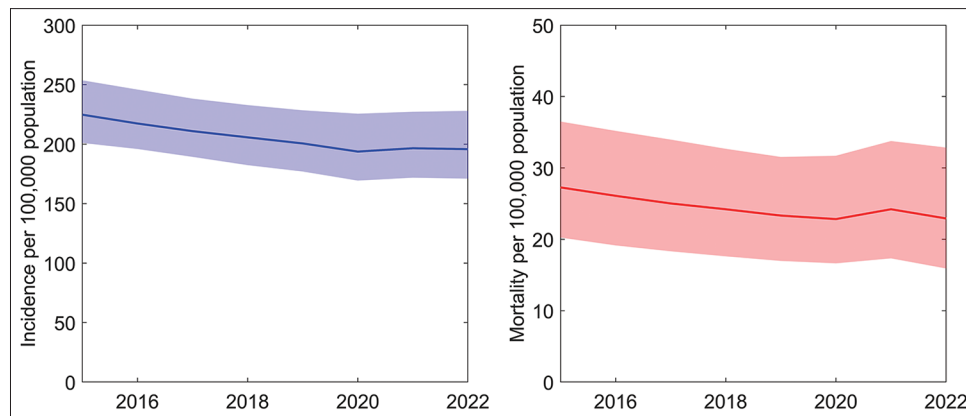
Figure 3: Estimated incidence and mortality rate from 2015 to 2022. Solid lines show the central estimate, while the shaded regions show the 95% credible intervals (CrI) of the estimates
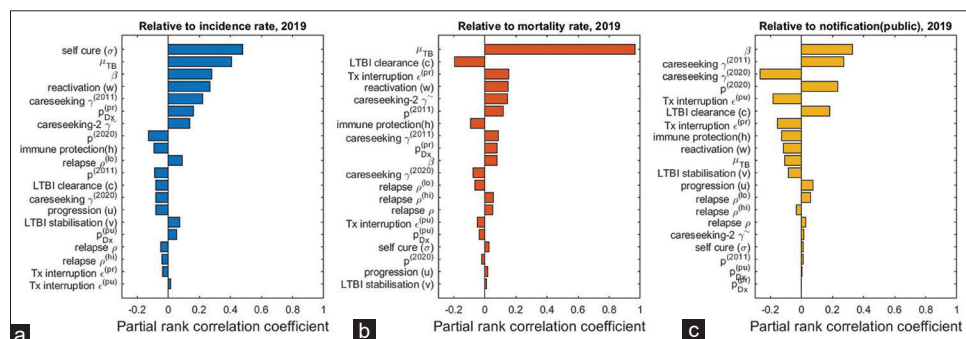
Figure 4: Sensitivity analysis. Partial rank correlation coefficients of each of the model inputs against incidence, mortality and notification rate in 2019

of Medically Certified Cause of Death (MCCD) among the registered deaths is still under 22%. Some of the smaller states have much better coverage such as Goa, which has 100% coverage of MCCD. While all States and Union Territories (UTs) in India are reporting MCCD by 2020, some of the larger states like Bihar & Uttar Pradesh still have poor MCCD coverage of less than 10%. Nonetheless in future, this data will be invaluable in improving our estimates of TB incidence and mortality in India.

As with any modelling analysis, our work has some limitations to note. Our estimates are undifferentiated by age, and by drug resistance or HIV status; while the latter two factors account for only a small proportion of TB incidence at the country level, at the subnational level they will nonetheless be more important for some states than others. All three factors are important areas for future work to address. Second, while our analysis has been supported by data on private sector drug sales, we have had to accommodate uncertainty in its implications for TB burden, most importantly in the proportion of private patients receiving publicly supplied drugs. The Bayesian calibration approach allows us systematically to incorporate these and other uncertainties into the estimation. Nonetheless, further data to narrow these uncertainties will be valuable for refining our estimates in future. Indeed, further evidence on the average duration of TB treatment in the private sector, and the extent of TB over-diagnosis in this sector, would be invaluable in

translating drug sales volumes directly to patient numbers. Finally, our calibration data for TB mortality come from IHME, which themselves are modelled estimates. While these estimates have been valuable measures for understanding TB burden in India, as mentioned above, it is hoped that they can be bolstered and further refined with increasing availability of vital registration data.

In conclusion, even if TB burden is infeasible to measure directly, a range of data can nonetheless offer indirect evidence for its estimation: mathematical modelling can be a helpful tool for bringing together these diverse sources of evidence, and deriving estimates that are consistent with them all.

group formed by TME will continue to review this model in future as well.

Financial support and sponsorship

Conflicts of interest

There are no conflicts of interest.

References

1. https://www.who.int/teams/global-tuberculosis-programme/the-end-tb-strategy

2. https://documents-dds-ny.un.org/doc/UNDOC/GEN/N17/207/63/PDF/N1720763.pdf?OpenElement

3. Global TB Report 2022, Geneva, World Health Organization 2022. License CC BY-NC-SA 3.0 IGO.

4. Population based survey for assessing prevalence of pulmonary tuberculosis cases in the state of Gujarat, India 2011-2012, Government of Gujarat, published by State TB Cell in 2013.

5. National TB Prevalence Survey in India (2019-2021), Summary Report, Indian Council of Medical Research, Ministry of Health & Family Welfare, Govt of India, Published in March 2022

6. Arinaminpathy N, Batra D, Maheshwari N, *et al*. Tuberculosis treatment in the private healthcare sector in India: an analysis of recent trends and volumes using drug sales data. *BMC Infect Dis* **19**, 539 (2019). https://doi.org/10.1186/s12879-019-4169-y

7. Nimalan Arinaminpathy, *et al*. The number of privately treated tuberculosis cases in India: an estimation from drug sales data. Lancet Infect Dis 2016; 16: 1255–60 http://dx.doi.org/10.1016/S1473-3099(16)30259-6

8. Jeyashree K, Thangaraj J, Rade K, *et al*. Estimation of tuberculosis incidence at subnational level using three methods to monitor progress towards ending TB in India, 2015–2020. BMJ Open 2022;12:e060197. doi:10.1136/ bmjopen-2021-060197

9. Global tuberculosis report 2020. Geneva: World Health Organization; 2020. Licence: CC BY-NC-SA 3.0 IG

10. https://www.who.int/publications/m/item/methods-used-by-who-to-estimate-the-global-burden-of-tb-disease-2022

11. Glaziou P, Sismanidis C, Pretorius C, Floyd K, Methods used by WHO to estimate the Global burden of TB disease https://arxiv.org/ftp/arxiv/papers/1603/1603.00278.pdf

12. https://vizhub.healthdata.org/gbd-compare/

13. Report on Medical certification of cause of death 2008, 2020, Registrar General of India https://crsorgi.gov.in/mccd-reports.html

14. Menzies NA, Wolf E, Connors D, Bellerose M, Sbarra AN, Cohen T, Hill AN, Yaesoubi R, Galer K, White PJ, Abubakar I, Salomon JA. Progression from latent infection to active disease in dynamic tuberculosis transmission models: a systematic review of the validity of modelling assumptions. Lancet Infect Dis. 2018 Aug;18(8):e228-e238. doi: 10.1016/S1473-3099(18)30134-8. Epub 2018 Apr 10. Erratum in: Lancet Infect Dis. 2018 Nov;18(11):1177. PMID: 29653698; PMCID: PMC6070419.

15. Nimalan Arinaminpathy, *et al*, Tuberculosis treatment in the private healthcare sector in India: an analysis of recent trends and volumes using drug sales data. BMC Infect Dis. 2019; 19: 539. doi: 10.1186/s12879-019-4169-y

16. Steve Brooks, Andrew Gelman, Galin Jones, and Xiao-Li Meng. Handbook of Markov Chain Monte Carlo. CRC Press, 2011.

17. Melanie H. Chitwood, A spatial-mechanistic model to estimate subnational tuberculosis burden with routinely collected data: An application in Brazilian municipalities. PloS Global Public Health. https://doi.org/10.1371/journal.pgph.0000725

Supplementary Information

Re-estimating Tuberculosis Incidence and Mortality in India during 2011-2022: A Modelling Study

TABLE OF CONTENTS

MODEL STRUCTURE

Figure S1 gives a schematic illustration of the overall model structure, with all model parameters listed in Table S1, and governing equations given below.

Parameters are defined below, and in Table 1. Infectious compartments contributing the force-of-infection are shown in red. For clarity, the diagram omits certain rates incorporated in the model, including: self-cure; exogenous reinfection; and background mortality.

EQUATIONS AND PARAMETER TABLES

Governing equations of the model are as follows. All state variables are written as proportions of the population (not as absolute numbers).

Uninfected (U):

$$\frac{dU}{dt} = b + c(L_f + L_s) - (\lambda + \mu)U$$

for a birth-rate b; force-of-infection λ; rate of clearance of LTBI c; and background mortality rate μ.

Latent, 'fast' infection (L_f):

$$\frac{dL_f}{dt} = \lambda U + (1-h)\lambda(R_{hi} + R_{lo} + R) - (\mu + u + v + c)L_f$$

for a progression rate u; a 'stabilisation' rate (to latent 'slow' status) v; and protection from reinfection h, amongst those previously infected.

Latent, 'slow' infection (L_s):

$$\frac{dL_s}{dt} = vL_f - (\mu + c + w)L_s$$

for a reactivation rate w.

Active TB (I):

$$\frac{dI}{dt} =$$
$$uL_f + wL_s + \rho_{hi}R_{hi} + \rho_{lo}R_{lo} + \rho R - (\gamma(t) + \mu_{TB} + \sigma + a(t))I$$

for relapse rates $\rho^{(hi)}$, $\rho^{(lo)}$, ρ; careseeking rate γ; TB mortality rate μ_{TB}; and self-cure rate σ.

Here, p and $(1 - p)$ are, respectively, the proportion-of-presentation to healthcare providers in the public and private sectors. The factor $\gamma(t)$ allows for an change in the notification over time, from 2011 onwards.

The term $a(t)$ represents the time-dependent effect of active case-finding.

Presented for diagnosis with provider type s ($D^{(s)}$):

$$\frac{dD^{(pu)}}{dt} = p\ (\gamma(t)I + \tilde{\gamma}(t)E) - (\mu_{TB} + \sigma + \delta)D^{(pu)}$$

$$\frac{dD^{(pr)}}{dt} = (1-p)\ (\gamma(t)I + \tilde{\gamma}(t)E) - (\mu_{TB} + \sigma + \delta)D^{(pr)}$$

for a rate-of-offering diagnosis δ. Here, $\tilde{\gamma}(t)$ is analogous to $\gamma(t)$, but attached to individuals who remain undiagnosed despite having previously sought care (i.e. compartment E below).
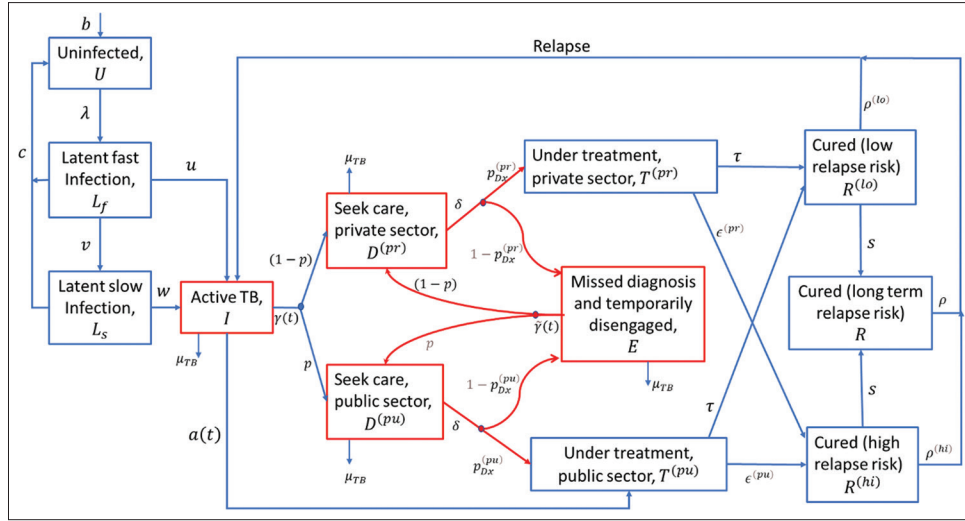
On TB treatment with provider type s ($T^{(s)}$):

$$\frac{dT^{(pu)}}{dt} = \delta p_{Dx}^{(pu)}D^{(pu)} + a(t)I - (\mu + \tau + \varepsilon^{(pu)})T^{(pu)}$$

$$\frac{dT^{(pr)}}{dt} = \delta p_{Dx}^{(pr)}D^{(pr)} - (\mu + \tau + \varepsilon^{(pr)})T^{(pr)}$$

for a treatment completion rate τ; and a treatment interruption rate $\epsilon^{(s)}$.

Missed diagnosis and temporarily disengaged from careseeking (E):

Figure S1: Schematic illustration of the model structure

$$\frac{dE}{dt} =$$

$$\delta\left(1 - p_{Dx}^{(pu)}\right)D^{(pu)} + \delta\left(1 - p_{Dx}^{(pr)}\right)D^{(pr)} - \left(\mu_{TB} + \sigma + \tilde{\gamma}(t)\right)E$$

Recovered with low relapse risk, following treatment completion ($R^{(lo)}$):

$$\frac{dR^{(lo)}}{dt} = \tau\left(T^{(pu)} + T^{(pr)}\right) - [(1-h)\lambda + \rho^{(lo)} + \mu + s]R^{(lo)}$$

for a rate of 'stabilisation' of relapse risk s.

Recovered with high relapse risk, following treatment completion ($R^{(hi)}$):

$$\frac{dR^{(hi)}}{dt} = \left(\varepsilon^{(pu)}T^{(pu)} + \varepsilon^{(pr)}T^{(pr)}\right) - [(1-h)\lambda + \rho^{(hi)} + \mu + s]R^{(hi)}$$

Long-term, 'stabilised' relapse risk (R):

$$\frac{dR}{dt} = s\left(R^{(lo)} + R^{(hi)}\right) - [(1-h)\lambda + \rho + \mu]R$$

Force-of-infection (λ):

$$\lambda = \beta(I + E + D^{(pr)} + D^{(pu)})$$

for a rate-of-transmission β.

Table of parameters
The notation $U[x,y]$ denotes a uniform probability distribution on the range $[x,y]$. Otherwise, parameter values in round brackets show 95% percentiles from the respective marginal posterior densities.

FURTHER INFORMATION ON DATA
Mortality rate estimation
For estimation of TB mortality in India, WHO currently relies on IHME estimations since 2015. However, the most recent data informing IHME estimates are from 2015. One important source of current data is vital registration collected through the *Registrar General of India* (RGI). In brief, this data includes verbal autopsies from trained physicians. The national coverage of this data is currently under 22%, and so it cannot be used as a stand-alone source of mortality estimates. However, it provides a valuable source of evidence to complement IHME estimates.

For comparison, the IHME estimate (WHO adjusted) for TB mortality in 2015 is 34 [32 – 36] per 100,000 population [WHO Global TB Report 2022]. RGI data suggests a rate of 22 [20 – 23] per 100,000 population.

Because of the incomplete coverage of RGI data, it is instructive also to compare the proportions of overall deaths attributed to TB. For IHME in 2015, this proportion 4.97 (4.57 - 5.38), whereas according to the RGI estimate this number is 3.17 (3.14 - 3.20), based on the Annual Reports on Medica Cause of Certified Deaths (MCCD) published by Registrar General of India (RGI) available at https://crsorgi.gov.in/mccd-reports.html.

Overall, therefore, the available evidence suggests that IHME mortality estimates should be adjusted downwards. As a simple but transparent approach, we assumed the central value is the average of the upper limit of RGI mortality and the lower limit of IHME mortality (~ 28 per 100,000 population). For uncertainty intervals, we took the upper limit of IHME estimates and the lower limit of RGI estimates, therefore adopting wide uncertainty. Thus, the mortality rate becomes 28 [20 - 36] per 100,000 population.

### Estimating declines in privately treated patients
As described in the main text, we used private sector drug sales data to inform trends in the number of patients being treated by private providers. In doing so, it was necessary to adjust for increasing uptake of publicly supplied drugs by private providers. Data from public-private-mix implementing agencies in India suggest that 20% of patients

Table S1: List of model parameters, values and sources

Parameter	Symbol	Value	Source/Notes
Natural history			
Infection rate (number of annual infections per case)	β	4.4 (3.4 – 6.2)	Model calibration, with priors U[0, 30]
Per-capita annual rate of progression from 'fast' latent infection	u	0.07 (0.01 – 0.13)	Calibration: Menzies (2018)[1] for central value, and U[0.1-20] on multiplying factor
Per-capita annual rate of stabilisation from 'fast' to 'slow' latent status	v	0.85 (0.78 – 0.95)	Menzies (2018)[1] for central value, and taking uniform priors of +/- 25%
Per-capita annual rate of reactivation from 'slow' latent infection	w	0.0035 (0.0004 – 0.0061)	Calibration: Menzies (2018)[1] for central value, and U[0.1-20] on multiplying factor
Per-capita annual rate of self-clearance of latent TB	c	0.031 (0.022 – 0.035)	Emery (2021)[2] for central value, and taking uniform priors of +/- 25%
Per-capita annual rate of TB mortality while untreated	μ_{TB}	0.07 (0.05 – 0.10)	Calibration: corresponding to case fatality rate (CFR) of undetected TB of 29% [20-40], during treatment in public sector 4% [3.7 – 4.4] and during treatment in private sector 1.2% [1 – 1.7] to meet the mortality target in 2015 i.e 28 [20-36] per 100,000 population.
Per-capita annual rate of TB mortality while in treatment in public sector	$\mu_{TB}(pu)$	0.103	
Per-capita annual rate of TB mortality while in treatment in private sector	$\mu_{TB}(pr)$	0.051	
Per-capita annual rate of TB self-cure	σ	0.17 (0.13 – 0.21)	Tiemersma *et al.*, 2011[3]
Protection from reinfection amongst those with prior infection	h	0.38 (0.13 – 0.73)	Andrews (2012)[4], assuming uniform priors of +/-25%
Per-capita annual rate of relapse in first two years after treatment completion	$\rho^{(lo)}$	0.083 (0.065 – 0.106)	Thomas A *et al* (2005)[5], Romanowski (2019)[6], Menzies (2009)[7] and Weis (1994)[8], with uniform prior using intervals of ±5%
Per-capita annual rate of relapse in first two years after self-cure or incomplete treatment	$\rho^{(hi)}$	0.14 (0.11 – 0.17)	
Per-capita annual rate of relapse>two years after last TB episode	ρ	0.0015 (0.0011 – 0.0018)	Most relapse occurs in first two years after recovery: Guerra-Assuncao (2015)[9]
Per-capita annual rate of 'stabilising' from high to low relapse risk	S	0.5	
TB services			
Rate-of-presentation to care, first careseeking visit — In 2011	$\gamma^{(2011)}$	0.30 (0.19 – 0.44)	Model calibration, with priors U[0.1, 10] for 2011; and for 2020 taking U[1, 10] as multiplying factor on $\gamma^{(2011)}$. As described in the main text, we allow for increasing rate over time motivated by increasing presumptive examination rates reported by the programme.
Rate-of-presentation to care, first careseeking visit — In 2020	$\gamma^{(2020)}$	0.55 (0.46 – 0.71)	
Rate-of-presentation to care, second and subsequent careseeking visits — Assuming same in 2011 and 2020	$\gamma\sim$	4.9 (0.8 – 12.4)	Model calibration, taking U[1, 40] for multiplying factor on $\gamma^{(2011)}$
Probability that a TB patient visits public provider, per careseeking attempt — In 2011	$p^{(2011)}$	0.50 (0 – 0.98)	Model calibration
Probability that a TB patient visits public provider, per careseeking attempt — In 2020	$p^{(2020)}$	0.6 (0.4 – 0.8)	
Per-capita rate of offering diagnosis	δ	52	Assumption, corresponding to 1 week
Probability of successful TB diagnosis and treatment initiation per careseeking visit — Public sector	$p_{Dx}^{(pu)}$	0.67 (0.34 – 0.89)	U[0.3, 0.9], motivated by Subbaraman (2016)[10]
Probability of successful TB diagnosis and treatment initiation per careseeking visit — Private sector	$p_{Dx}^{(pr)}$	0.64 (0.33 – 0.88)	U[0.3, 0.9], assumption
Per-capita annual rate of treatment completion	τ	2	Corresponds to average treatment duration of 6 months
Per-capita annual rate of treatment interruption — Public sector	$\in^{(pu)}$	0.52 (0.26 – 0.66)	Calculated using $\varepsilon^{(pu)} = \frac{1-P}{P}\tau$, for treatment completion rate P, and assuming U[0.75, 0.95] for P
Per-capita annual rate of treatment interruption — Private sector	$\in^{(pr)}$	2.2 (0.9 – 2.9)	As above, but assuming U[0.4, 0.8] for P
Demographics			
Per-capita annual rate of background mortality	μ	1/70	Corresponds to average lifespan of 70 years (World Bank 2021)[11]

Table S2: Estimation of trend of privately treated patients

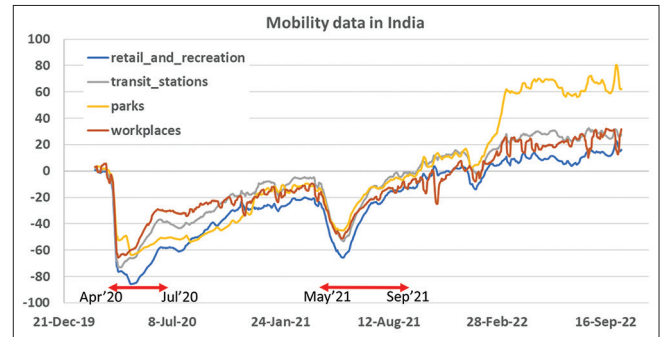| Average patient months	2015	2016	2017	2018	2019
Actual data	1466928	1245305	1056072	984014	815997
Adjusted patient months- assuming 30% treated with public drugs in 2019 (linearly changes from 2015 to 2019)	1466928	1346276	1242438	1269695	1165710
Adjusted patient months assuming 10% treated with public drugs in 2019 (linearly changes from 2015 to 2019)	1466928	1277236	1111655	1063799	906663
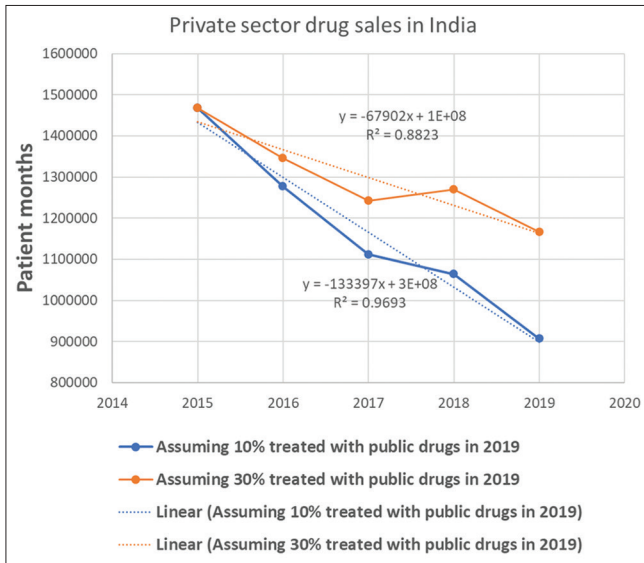
Figure S2: Linear fit with the adjusted patient months from 2015 to 2019 assuming 10% and 30% treated with public sector drugs in 2019

Figure S3: Mobility data in India during the pandemic period. The two red arrows show the period of national lockdown, and the delta wave in India, respectively

being managed by private providers in India were treated using public drugs. Accordingly, Table S1 below shows treatment volumes of all patients managed in the private sector under scenarios where 10% and 30% of patients receive public drugs. We fit these data -series using linear regression (see Figure S2) to estimate the average decline rate of patients in the private sector from 2015 onwards. Overall, this suggests the decline rate in private sector treatment initiation from 2015 to 2019 is 24% [19 - 37]. We used this decline rate as a target for model calibration.

Linear regression for public notifications and deflation of 2019 numbers

The total TB notification from the public sector notification in

2019 was 125 per 100,000 population. While this data reflects all diagnosed patients, the model only counts new/relapse cases who initiated treatment. We adjusted for both, including 15% initial loss-to-followup, ultimately yielding a rate of 94 [80 – 108] per 100,000 populations.
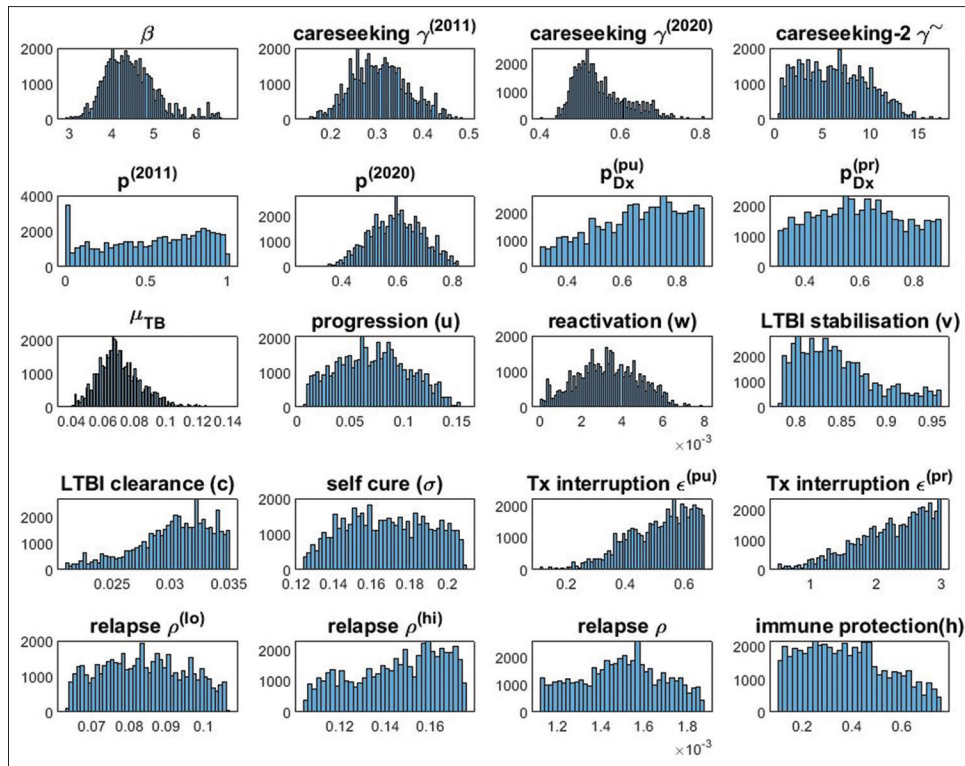
Mobility data[12]
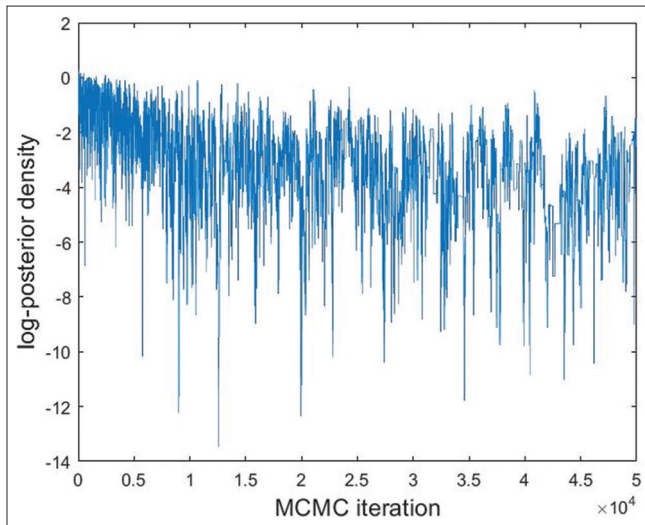Figure S3.

POSTERIOR DISTRIBUTIONS FOR MODEL PARAMETERS
Figure S4.

RESULTS OF THE MCMC CALIBRATION
We sampled from the posterior density using adaptive Bayesian Markov Chain Monte Carlo simulation[13] as described in the Calibration section of the main article. Figure S5 below shows the trace arising from the MCMC calibration.

Figure S4: Posterior distributions for model parameters

Figure S5: Trace plot arising from MCMC calibration, showing the log-posterior density over 50,000 iterations

REFERENCES

1. Menzies NA, Wolf E, Connors D, *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. 2018. doi:10.1016/S1473-3099(18)30134-8

2. Emery JC, Richards AS, Dale KD, *et al.* Self-clearance of Mycobacterium tuberculosis infection: implications for lifetime risk and population at-risk of tuberculosis disease. *Proc R Soc B Biol Sci* 2021;288:20201635. doi:10.1098/rspb.2020.1635

3. Tiemersma EW, van der Werf MJ, Borgdorff MW, *et al.* Natural History of Tuberculosis: Duration and Fatality of Untreated Pulmonary Tuberculosis in HIV Negative Patients: A Systematic Review. *PLoS One* 2011;6:e17601. doi:10.1371/journal.pone.0017601

4. Andrews JR, Noubary F, Walensky RP, *et al.* Risk of progression to active tuberculosis following reinfection with Mycobacterium tuberculosis. *Clin Infect Dis* 2012;54:784–91. doi:10.1093/cid/cir951

5. Thomas A, Gopi PG, Santha T, Chandrasekaran V, *et al.* Predictors of relapse among pulmonary tuberculosis patients treated in a DOTS programme in South India. Int J Tuberc Lung Dis. 2005 May;9(5):556-61. PMID: 15875929

6. Romanowski K, Balshaw RF, Benedetti A, *et al.* Predicting tuberculosis relapse in patients treated with the standard 6-month regimen: an individual patient data meta-analysis. *Thorax* 2019;74:291–7. doi:10.1136/thoraxjnl-2017-211120

7. Menzies D, Benedetti A, Paydar A, *et al.* Effect of duration and intermittency of rifampin on tuberculosis treatment outcomes: A systematic review and meta-analysis. PLoS Med. 2009;6. doi:10.1371/journal.pmed.1000146

8. Weis SE, Slocum PC, Blais FX, *et al.* The Effect of Directly Observed Therapy on the Rates of Drug Resistance and Relapse in Tuberculosis. *N Engl J Med* 1994;330:1179–84. doi:10.1056/NEJM199404283301702

9. Guerra-Assunção JA, Houben RMGJ, Crampin AC, *et al.* Recurrence due to relapse or reinfection with Mycobacterium tuberculosis: a whole-genome sequencing approach in a large, population-based cohort with a high HIV infection prevalence and active follow-up. *J Infect Dis* 2015;211:1154–63. doi:10.1093/infdis/jiu574

10. Subbaraman R, Nathavitharana RR, Satyanarayana S, *et al.* The Tuberculosis Cascade of Care in India's Public Sector: A Systematic Review and Meta-analysis. *PLOS Med* 2016;13:e1002149. doi:10.1371/journal.pmed.1002149

11. The World Bank. India demographic data. https://data.worldbank.org/country/india (accessed 14 Nov 2018).

12. Our world in Data, CIVID-19: Google Mobility Trends. https://ourworldindata.org/covid-google-mobility-trends (accessed on 10 Mar 2023).

13. Haario H, Saksman E, Tamminen J. An Adaptive Metropolis Algorithm. Bernoulli Published Online First: 2007. doi:10.2307/3318737