Data-driven model for the assessment of *Mycobacterium tuberculosis* transmission in evolving demographic structures

Sergio Arregui[a,b,1], María José Iglesias[c,d], Sofía Samper[d,e], Dessislava Marinova[c,d], Carlos Martin[c,d,f], Joaquín Sanz[g,h,2], and Yamir Moreno[a,b,i,1,2]

[a]Institute for Biocomputation and Physics of Complex Systems, University of Zaragoza, 50018 Zaragoza, Spain; [b]Department of Theoretical Physics, University of Zaragoza, 50009 Zaragoza, Spain; [c]Department of Microbiology, Faculty of Medicine, University of Zaragoza, 50009 Zaragoza, Spain; [d]Centro de Investigación Biomédica en red Enfermedades Respiratorias (CIBER), Carlos III Health Institute, 28029 Madrid, Spain; [e]Instituto Aragonés de Ciencias de la Salud, Instituto de Investigación Sanitaria (IIS) Aragon, 50009 Zaragoza, Spain; [f]Service of Microbiology, Miguel Servet Hospital, Instituto de Investigación Sanitaria (IIS) Aragon, 50009 Zaragoza, Spain; [g]Department of Genetics, Sainte-Justine Hospital Research Centre, Montreal, QC H3T1C5, Canada; [h]Department of Biochemistry, Faculty of Medicine, University of Montreal, Montreal, QC H3T1J4, Canada; and [i]Institute for Scientific Interchange, ISI Foundation, 10126 Turin, Italy

Edited by Barry R. Bloom, Harvard T. H. Chan School of Public Health, Boston, MA, and approved February 27, 2018 (received for review November 27, 2017)

In the case of tuberculosis (TB), the capabilities of epidemic models to produce quantitatively robust forecasts are limited by multiple hindrances. Among these, understanding the complex relationship between disease epidemiology and populations' age structure has been highlighted as one of the most relevant. TB dynamics depends on age in multiple ways, some of which are traditionally simplified in the literature. That is the case of the heterogeneities in contact intensity among different age strata that are common to all airborne diseases, but still typically neglected in the TB case. Furthermore, while demographic structures of many countries are rapidly aging, demographic dynamics are pervasively ignored when modeling TB spreading. In this work, we present a TB transmission model that incorporates country-specific demographic prospects and empirical contact data around a data-driven description of TB dynamics. Using our model, we find that the inclusion of demographic dynamics is followed by an increase in the burden levels predicted for the next decades in the areas of the world that are most hit by the disease today. Similarly, we show that considering realistic patterns of contacts among individuals in different age strata reshapes the transmission patterns reproduced by the models, a result with potential implications for the design of age-focused epidemiological interventions.

tuberculosis | infectious disease transmission | epidemiological models | dynamics population | global health

The control of tuberculosis (TB) has been one of the largest endeavors of public health authorities ever since the bacterium that causes it—*Mycobacterium tuberculosis*—was discovered (1). Recently, the development of global strategies for diagnosis and treatment optimization has led to TB burden decay worldwide (2), to the point that the End TB Strategy has allowed the scientific community to think that its eradication before 2050 is possible (3, 4). Nonetheless, such a goal is still far away, and TB remains a major public health problem (5–7), being responsible for 1.7 million deaths worldwide in 2016 (4). These dramatic data evidence the need for new epidemiological measures and pharmacological resources (8). In the task of forecasting the potential impacts of such new interventions, epidemiological models of TB transmission constitute a fundamental resource to assist decision making by public health agents (9).

Among the various limitations that TB modeling has to face in this context, achieving a proper description of the multiple ways whereby TB dynamics couple with populations' age structure has been identified as one of the most critical (5, 10). For example, patients' age is strongly correlated to the type of disease that they tend to develop more often, as well as to the probability of developing active TB immediately after infection [usually called "fast progression" (8)]. This way, while a larger fraction of children younger than 15 y of age develop noninfectious forms of extrapulmonary TB with respect to adults [25% vs. 10% (8, 11, 12)], the risk of fast progression is larger in infants (50% in the first year of life), then decays (20–30% for ages 1–2 y, 5% for 2–5 y, and 2% for 5–10 y), and increases again in adults (10–20% for individuals older than 10 y) (13). Additionally, transmission routes of TB, being a paradigmatic airborne disease, are expected to show significant variations in intensity across age (14, 15). The empirical characterization of these contact structures constitutes an intense focus of research in data-driven epidemiology of airborne diseases (16), and their influence on the transmission dynamics of diseases like influenza has been recently explored with relevant implications (17, 18).

Thus, if subjects' age modifies the disease-associated risks at the level of single individuals, it is likely that changes in the demographic age structure at the population level will impact TB

Significance

Even though tuberculosis (TB) is acknowledged as a strongly age-dependent disease, it remains unclear how TB epidemics would react, in the following decades, to the generalized aging that human populations are experiencing worldwide. This situation is partly caused by the limitations of current transmission models at describing the relationship between demography and TB transmission. Here, we present a data-driven epidemiological model that, unlike previous approaches, explicitly contemplates relevant aspects of the coupling between age structure and TB dynamics, such as demographic evolution and contact heterogeneities. Using our model, we identify substantial biases in epidemiological forecasts rooted in an inadequate description of these aspects, at the level of both aggregated incidence and mortality rates and their distribution across age strata.

Author contributions: C.M., J.S., and Y.M. designed research; S.A. performed research; S.A., J.S., and Y.M. contributed new reagents/analytic tools; S.A., M.J.I., S.S., D.M., C.M., J.S., and Y.M. analyzed data; S.A., M.J.I., S.S., D.M., C.M., J.S., and Y.M. wrote the paper; and S.A. and J.S. implemented the model.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: yamir.moreno@gmail.com or sergioarregui.sa@gmail.com.

[2]J.S. and Y.M. contributed equally to this work.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1720606115/-/DCSupplemental.

Published online March 21, 2018.

burden projections. This is mainly due to the slow dynamics that are characteristic of TB, which forces modelers to describe the evolution of the disease during long periods of time, typically spanning several decades. These timescales are rather incompatible with the assumption of constant demographic structures, at least presently, since worldwide human populations are presumed to age from the current median of 30 y old to 37 y old in 2050 (19). And yet, achieving a sensible description of TB transmission able to capture the effects of time-evolving demographic structures remains an elusive goal in TB modeling. Demographic dynamics are traditionally neglected in TB transmission models, the same way that contact structures are assumed to be homogeneous across age groups (8, 20, 21).

In this work, we incorporate empirical data on demographic dynamics and contact patterns into classical formulations of TB spreading models, thus unlocking less biased descriptions of the spreading dynamics of the disease. To this end, we present a TB spreading model (Fig. 1*A*) whereby we provide a data-driven description of TB transmission that presents two main differences with respect to previous approaches. First, our model incorporates demographic forecasts by the United Nations (UN) population division (19) (Fig. 1*B*) to describe the coupling between demographic evolution and TB dynamics. Second, the model integrates region-wise empirical data about age-dependent mixing patterns adapted from survey-based studies conducted in Africa and Asia (22–26) (Fig. 1*C*), instead of assuming that all of the individuals in a population interact homogeneously, as traditionally considered in the literature (8, 20, 21).

Upon model calibration in some of the countries most affected by the disease in 2015 and subsequent simulation of TB transmission dynamics up to 2050 (Fig. 1*D*), we scrutinize the implications derived from integrating these pieces of empirical data within our model and discuss their impact on the forecasts produced, at the level of both aggregated incidence and mortality rates and their distributions across age strata. Specifically, we quantify the effects of populations' aging on predicted incidence rates until 2050, as well as the impact on the age distribution of the disease burden that emanates from introducing empiric contact data into the models. Furthermore, we quantify the sensitivity of these effects to the different model inputs and assess their statistical significance and robustness under a series of alternative modeling scenarios.

Results

Baseline Forecasts of TB Incidence and Mortality. To illustrate the ability of our method to reproduce current epidemic trends in different scenarios, the model was applied to describe the TB epidemics in India, Indonesia, Nigeria, and Ethiopia (Fig. 2). These countries, which accumulated as much as ∼40% of the total TB burden worldwide in 2015, were selected because of their different temporal evolution trends, current and projected

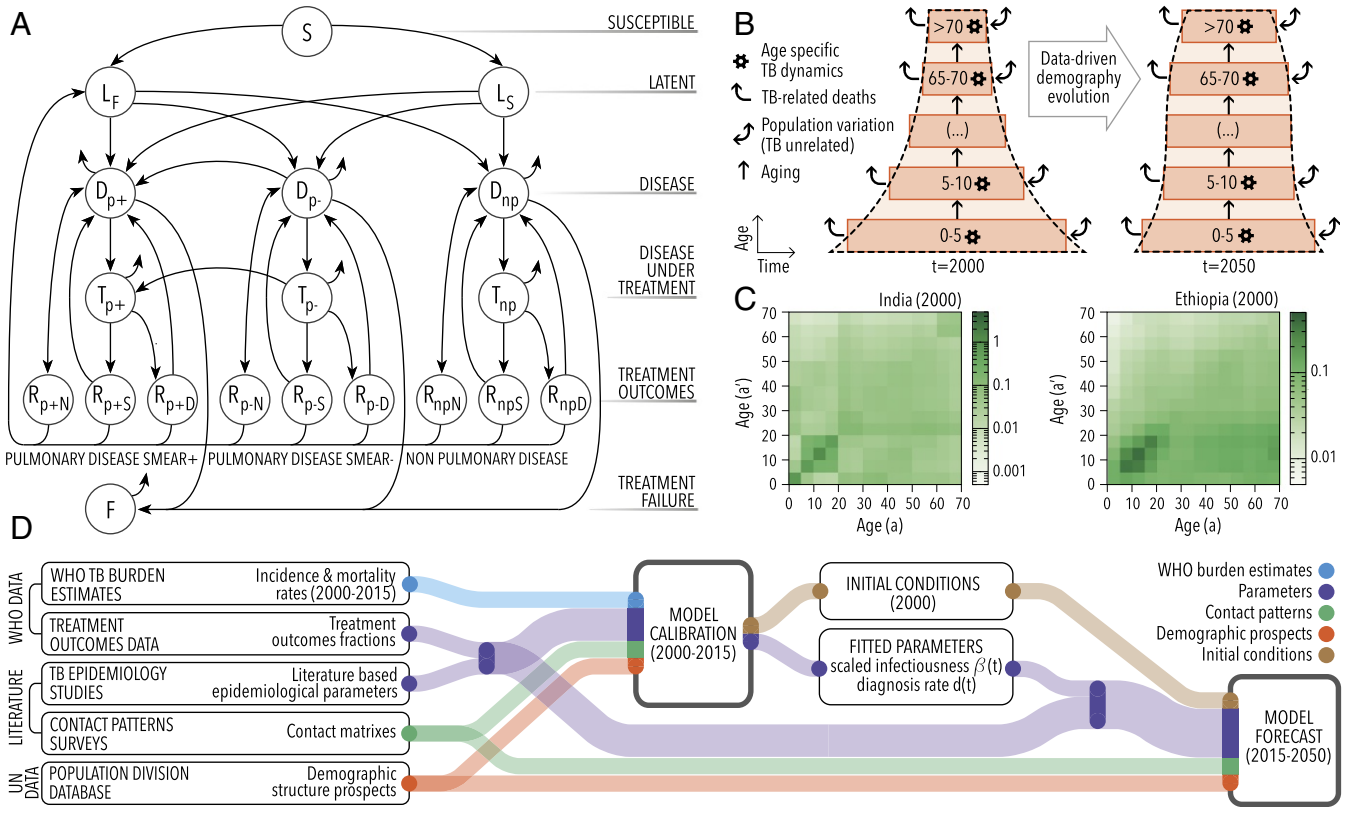

Fig. 1. Model description. (*A*) Natural history scheme of the TB spreading model. D, (untreated) disease; F, failed recovery; L, latent; R, recovered; S, susceptible; T, (treated) disease. Types of TB considered: np, nonpulmonary; p+, pulmonary smear positive; p–, pulmonary smear negative. Treatment outcomes: F, treatment failure; R_D, default (abandon of treatment); R_N, natural recovery; R_S, successful treatment. (*B*) Scheme of the coupling between TB dynamics and demographic evolution. The transmission model summarized in *A* describes the evolution of the disease in each age group, including the removal of individuals due to TB mortality (curved arrows). The evolution of the total volume of each age stratum is corrected (bidirectional arrows: TB-unrelated population variations) to make the demographic pyramid evolve according UN prospects. (*C*) Empirical contact patterns used for African and Asian countries. (*D*) Data flow scheme. Epidemiological parameters, contact matrices, and demographic prospects are used to calibrate the model, with the goal of reproducing observed TB incidence and mortality trends during the period 2000–2015. As a result of model calibration, scaled infectiousness, diagnosis rates, and initial conditions of the system in 2000 are inferred. These elements are then used (along with epidemiological parameters, contacts, and demographic data) to extend model forecasts up to 2050. For further details regarding model formulation and calibration, the reader is referred to *SI Appendix*.

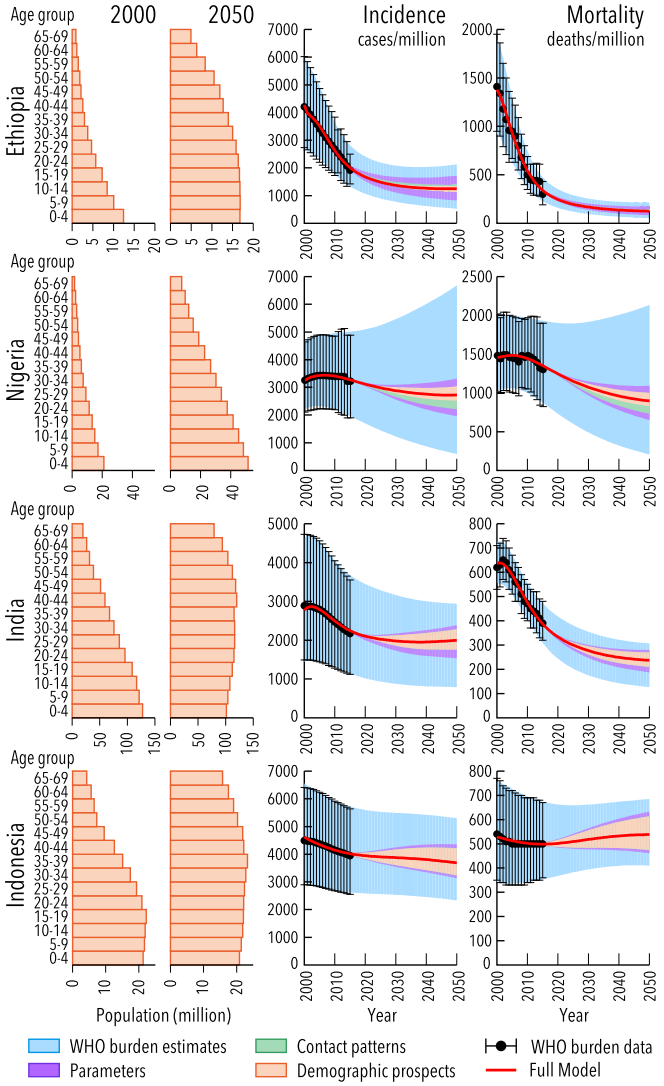

Fig. 2. Population structure at 2000 and 2050 (projection) and annual incidence and mortality rates predicted by our model in 2000–2050 for Ethiopia, Nigeria, India, and Indonesia. Colored areas represent 95% confidence intervals. The contribution to overall uncertainty that stems from each of the four types of input data is disclosed. (Contributions are cumulative.)

demographic profiles, and geographic locations. Remarkably, our model does not predict, in general, a sustained decrease in TB burden for the decades to come in these cases, whose incidence rates (per million habitants and year) range between 1,246 (524–2,124, 95% CI) (Ethiopia) and 3,669 (2,348–5,247, 95% CI) (Indonesia), in 2050. Additionally, we extended these analyses to the top 12 countries suffering from the highest absolute TB burden levels in 2015, producing satisfactory fits in all cases (*SI Appendix*, Fig. S1 and Table S1).

During simulations, our model produced TB detection ratios (TB cases diagnosed divided by new incident cases) that strongly correlate to the notification rates across countries reported by WHO (27) (*SI Appendix*, Fig. S2; Pearson correlation $r = 0.96$, $P = 4.3\text{E-}6$). Model-based case detection ratios are systematically larger than notification rates, which is an expected result, congruent with the fact that a fraction of all diagnosed cases is not reported to the WHO surveying system.

Regarding confidence intervals, colored areas in Fig. 1 quantify the contributions to global uncertainty that stem from the different types of input data processed by the model. These include epidemiological parameters (purple), demographic data (orange), contact patterns (green), and, most importantly, WHO-burden estimations (blue). Complementarily, in *SI Appendix*, Fig. S3, the individual contribution of each epidemiological parameter is further disclosed in an exhaustive sensitivity analysis. Of all of the different individual sources of uncertainty that could impact the model's forecasts, current WHO estimates for TB burden levels are the only ones that introduce more than a 15% deviation with respect to central estimates [the uncertainties in total number of TB cases projected in 2000–2050 that are propagated from WHO data span from 36% (Ethiopia, lower limit) to 92% (Nigeria, upper limit) with respect to central expectations].

Effects of Populations Aging on Aggregated TB Forecasts. As can be deduced from the demographic pyramids in *SI Appendix*, Fig. S1, all countries analyzed in this work are experiencing population aging to some extent, consistent with the overall trend that is forecasted for global human populations during the same period (19). The four countries selected in Fig. 2 lie at different points of the demographic transition by the beginning of the period under analysis (year 2000) and are expected to evolve at different paces into more or less aged populations by 2050.

To isolate the influence of populations aging on model outcomes, we compared our model with a simplified version where demographic evolution is neglected as done in previous approaches (8, 20, 21) (reduced model 1). In this reduced model the demographic structures are taken from their initial configuration in 2000 and remain static until 2050. Our results show that the demographic evolution leads to a systematic and significant increase in the predicted incidence rates, which is variable in size across countries [Fig. 3*A*: relative increase in incidence in 2050: full vs. reduced model 1: India, 39.6% (13.9–63.6, 95% CI); Indonesia, 23.4% (7.9–36.5, 95% CI); Ethiopia, 56.0% (29.2–62.1, 95% CI); Nigeria, 34.5% (9.1–42.9, 95% CI); see also *SI Appendix*, Figs. S1 and S4 and Table S2 for equivalent results in other countries]. Furthermore, the relative variation between incidence forecasts obtained from the full and the reduced model by 2050 significantly correlates with the intensity of the aging shift, as given by the change in the fraction of adults (age >15 y) in 2000–2050 (Fig. 3*B*, Pearson correlation $r = 0.66$, $P = 0.02$). This is indeed a natural consequence, since adults are burdened with higher incidence rates than children, and thus, populations' aging implies a relative increase of the demographic strata that is most affected by the disease (adults), in detriment of children, among whom TB incidence is lower (Fig. 3*C*).

Next, we built a series of synthetic demographic evolutions to simulate different scenarios (Fig. 3*D*). To this end, we used three pivotal examples extracted from actual cases of populations featuring young, triangular demographic pyramids (Fig. 3*D*, stage *i*, extracted from Ethiopia in 2000) and aged, inverted pyramids (stage *v*, extracted from China, 2050), as well as intermediate situations (stage *iii*, extracted from Indonesia, 2000, and stages *ii* and *iv*, built upon linear interpolation). Making use of these pivotal populations, we built synthetic transitions among them occurring in the period 2000–2050, which we then integrate in our TB model, in the four countries analyzed, instead of their own real demographic projections. As we can see in Fig. 3*D*, population aging appears associated with increased incidence rates, while eventual transitions toward younger populations would cause incidence forecasts to decline faster.

To further validate the general character of these results, we performed a series of robustness tests in scenarios that go beyond the assumptions made in our modeling framework. These include comparing full and reduced model 1 under the assumption of highly biased input data (i.e., burden data departing significantly from WHO uncertainty estimates; *SI Appendix*, Fig. S5), swapping contact structures across continents (*SI*

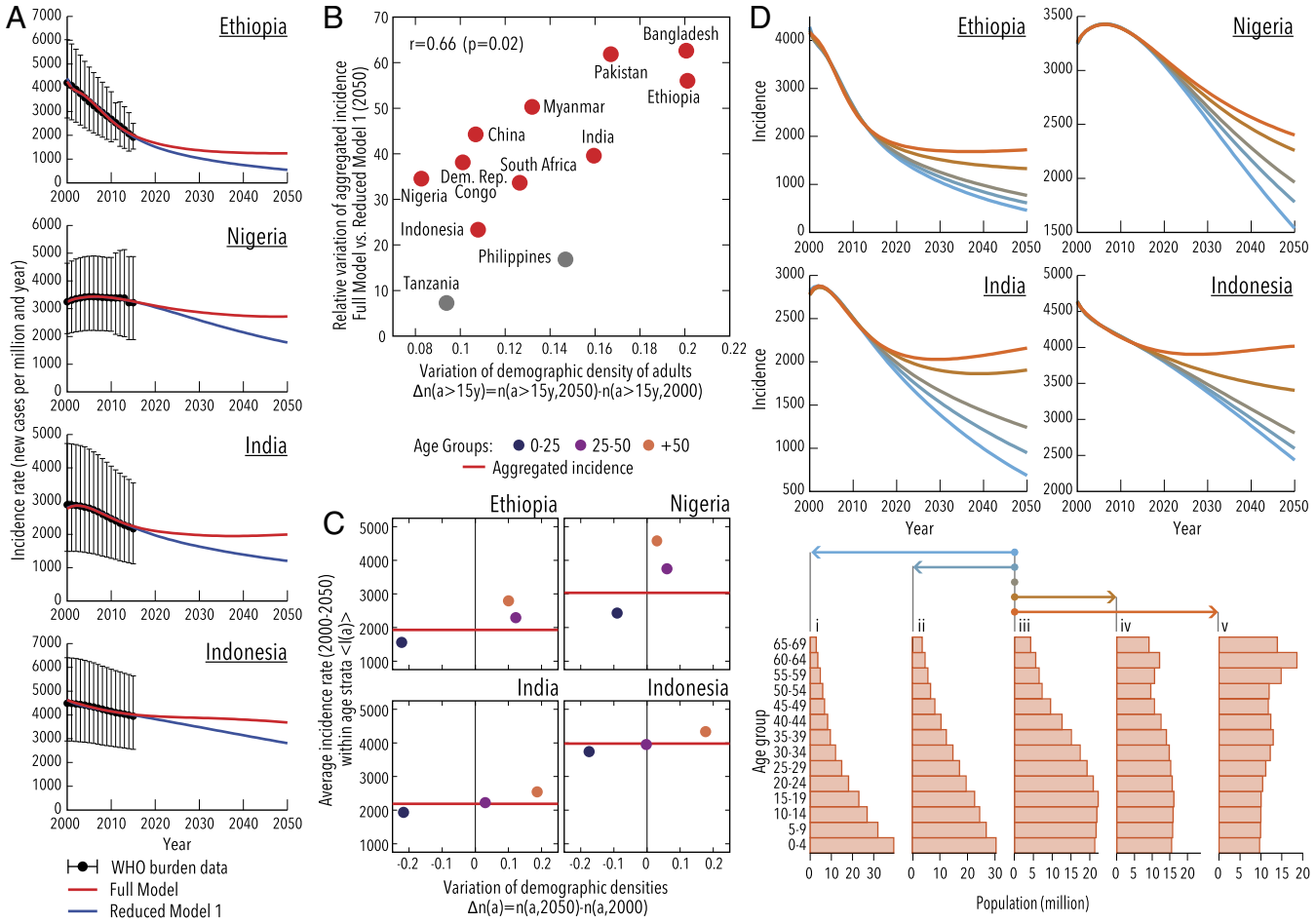

Fig. 3. Effects of demographic dynamics on model forecasts. (*A*) Incidence rates from 2000–2050 obtained from the full model (red) and reduced model 1 (constant demography, blue). Relative variation in incidence in 2050: full vs. reduced model 1: India, 39.6% (13.9–63.6, 95% CI); Indonesia, 23.4% (7.9–36.6, 95% CI); Ethiopia, 56.0% (29.2–62.1, 95% CI); Nigeria, 34.5% (9.1–42.9, 95% CI). (*B*) Relative variation of aggregated incidence at 2050 for the top 12 countries with highest absolute TB burden in 2015 vs. variation in the fraction of adults in the population during the period 2000–2050. In all countries but Tanzania and The Philippines, in gray, the variations in incidence are significant at a nominal $P = 0.05$. (*C*) Age-specific average incidence rate of TB vs. variation of age-strata population density in 2000–2050. Older individuals are, at the same time, those affected by higher TB incidence rates and those whose presence in the population is increasing as a result of populations' aging. (*D*) Incidence projections for synthetic scenarios of demographic evolution, including transition toward younger populations (*iii–i* and *iii–ii*), static populations (*iii* remaining constant), and realistic transitions representing populations' aging (from *iii* to *iv* and from *iii* to *v*). Pivotal demographic structures corresponding to stages *i*, *iii*, and *v* are taken from actual examples (Ethiopia in 2000, Indonesia in 2000, and China in 2050) and normalized to a common total population to rule out hypothetical system volume effects. Stages *ii* and *iv* are obtained upon linear interpolation. In each panel, the demographic evolution of each country is substituted by these synthetic scenarios: demographic transitions that go from stage *iii* in 2000 to different ending points in 2050. Colors in incidence series correspond to those in the arrows below, indicating the respective demographic transitions. Model calibration is repeated in each case.

Appendix, Fig. S6), and interrupting the time evolution of the fitted parameters after 2015 (*Materials and Methods* and *SI Appendix*, Fig. S7) as well as dispensing with the independent calibration of the reduced model to rule out the possibility of these differences arising from technical artifacts during model calibration (*SI Appendix*, Fig. S8). Remarkably, under all these alternative scenarios, the comparison between full and reduced model remains valid.

Collectively, our results show that ignoring the populations aging within TB spreading models generates forecasts of aggregated burden that are systematically and significantly lower than those obtained when this ingredient is taken into account.

Effects of Aging on Age-Specific Burden Levels. Next, we interrogated whether the effect of aging on TB burden estimates is only due to a relative increase of the age strata more hit by the disease (i.e., adults) or whether, in turn, significant increases in the incidence rates within age groups can be identified.

In Fig. 4, we show, for one example per continent—India and Ethiopia—the infection matrices between age groups described by each model and their difference. The entry (a, a') of these matrices represents the predicted number of infections (in 2050) from age-group a (infection source) to a' (infection target) per year per million people in group a'. For both countries, the differences between full and reduced model 1 point to a systematic underestimation of the number of infection events caused by adults as a consequence of ignoring demographic dynamics, as well as an overestimation—only appreciable in India—of infections caused by children during the period under analysis. Furthermore, once contagions are aggregated across infection sources within each target age group (Fig. 4, age-specific infection rates histograms, built as column-wise marginal sums of the infection matrices), significant differences between age-specific infection rates arise in both countries, mainly in adult age strata, where the full model predicts systematically larger incidence levels than the reduced model 1.

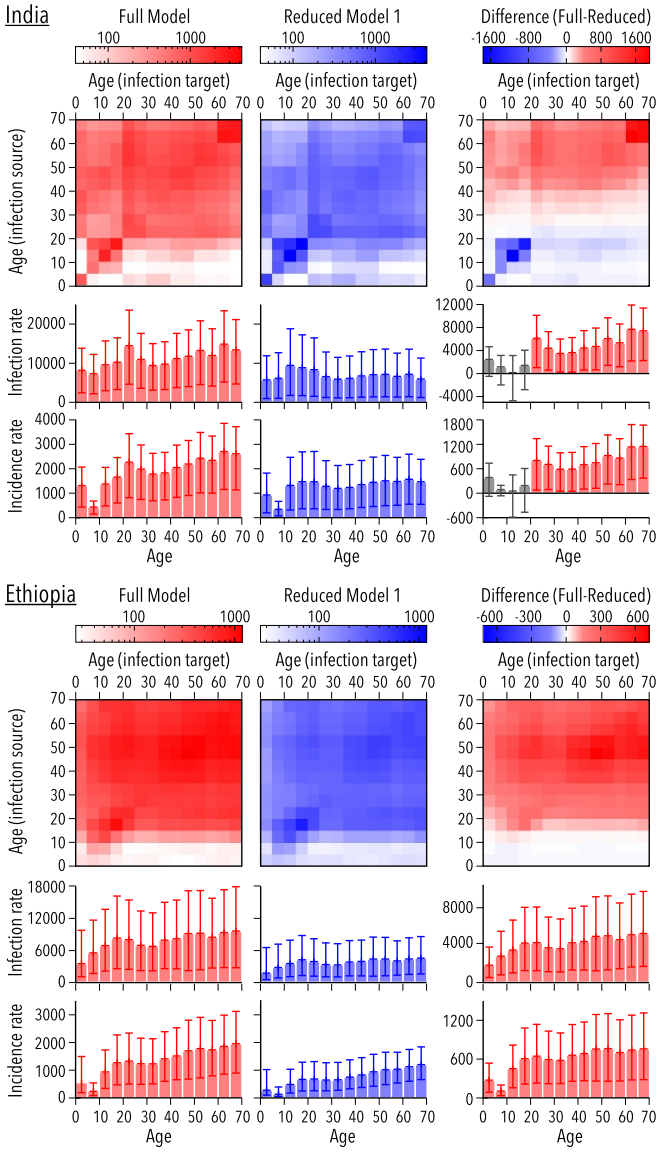

Fig. 4. Age-to-age infection rate matrices (number of infections from age group a to age group a' per year per million people in target age-group a') and age-specific infection and incidence rates forecasted in 2050 for India and Ethiopia (number of contagions or new active TB cases, respectively) per year and million individuals in a given age group. (*Left* column) The forecasts derive from the full model. (*Center* column) The forecasts derive from reduced model 1 (constant demography). (*Right* column) The difference (full model minus reduced model 1) of these three observables: infection matrices, age-specific infection rates, and age-specific incidence rates. Differences in incidence and infection rates are shown in gray when they are not statistically significant. Neglecting demographic dynamics appears associated to an underestimation of infections caused by adults in both countries and an overestimation of infections from children, mostly in India (infection matrices). At the level of infection/incidence rates (histograms), the full model produces larger age-specific infection and incidence rates than the reduced version, more intensely among adults.

This ultimately translates into an increase in age-specific incidence rates of active TB cases (Fig. 4, age-specific incidence histograms), which can be easily interpreted by attending to the larger probabilities of developing the most infectious forms of pulmonary TB that adults experience with respect to children (8). Adults, whose proportion increases in the system as a result of considering populations' aging, constitute not only the part of the demographic pyramid most hit by the disease, but also the one that contributes the most to overall spreading. Therefore, including populations' aging on model dynamics causes an increase not just in the aggregated burden levels across all age groups, but also within age strata.

Effect of Contact Pattern Heterogeneities. After discussing the impact that demographic dynamics have on model outcomes, we inspected what the effects are of including contact patterns in the TB model forecasts, either at the level of aggregated rates or within age-specific strata. To do so, we built a second reduced model where the empirical contact matrices estimated from survey studies conducted in Africa and Asia (22–26) are substituted by the classical hypothesis of contacts homogeneity (reduced model 2; see *Materials and Methods* and *SI Appendix*, section 2.3 for further details).

In Fig. 5, we represent the infection matrices that derive from the full and the reduced model 2 for India and Ethiopia in 2050. Clearly, empirical contact patterns reshape the distribution of contagions among age groups, giving a larger importance to assortative infections that take place among individuals of similar ages—specifically between adolescents and young adults—while penalizing infections from children to adults or vice versa. As a result, in this case, the infection and incidence rates of TB among children are higher in the reduced model, while the full model predicts more infection and disease burden among adults, with slight variations between the two countries that are due to the different contact data used in each case in the full model. In all of the countries analyzed, the opposite directions of the differences between the full and the reduced model that are found in children vs. adults tend to compensate each other. This results in similar global incidence rates produced by both models (*SI Appendix*, Fig. S9).

Once we showed that empirical contact patterns adapted from both African and Asian studies produce results that depart significantly from those obtained assuming homogeneous mixing, we interrogated whether the differences between the contact matrices used in both continents (Fig. 1*C*, for example) are significant enough to translate into differences in TB burden forecasts. To do this we conducted an additional test in one country—Ethiopia—in which we evaluated the differences in the TB burden distribution across ages that emanate from using contact data adapted from African, Asian, and, as a control, European studies. The results of this analysis are presented in *SI Appendix*, Fig. S10, and evidence that the different contact structures used in this work, derived from different empirical studies, introduce significant differences in the distribution of TB incidence. This emphasizes the importance of the estimation of high-quality, country-specific data about contact patterns for the production of robust epidemic forecasts in age-structured models.

Finally, we tested whether significant differences regarding age-specific distributions of incident cases can also be observed between the full and the reduced model 2 in a series of alternative modeling scenarios. The results of these tests (analogous to those presented in *SI Appendix*, Figs. S5–S8 for the effects of demographic dynamics) are shown in *SI Appendix*, Fig. S11, and indicate that the effects of empirical contact patterns on TB burden distributions are robustly significant under a wide spectrum of alternative situations.

Summarizing this part, and despite the reduced effect observed on aggregated rates, we showed that including empirical contact structures on TB model dynamics reshapes the transmission patterns among age groups and generates significant differences in age-specific infection and incidence rates. Additionally, we showed that considering different matrices estimated from studies conducted in different geographical areas significantly impacts the projected burden distributions.

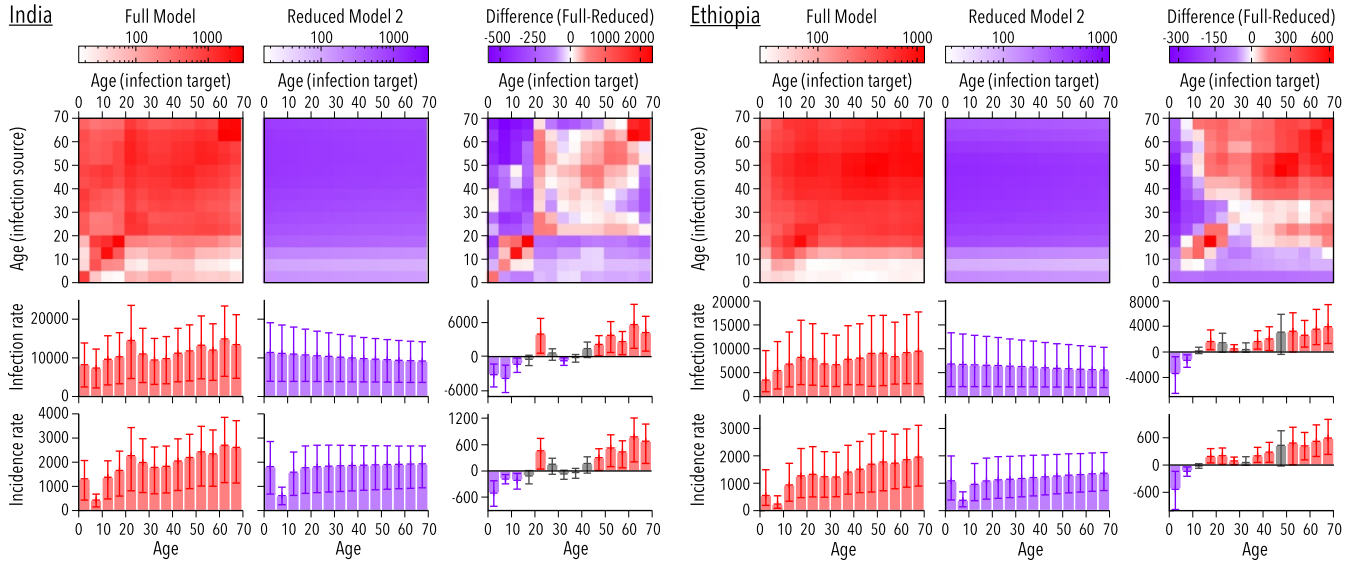

Fig. 5. Age-to-age infection rate matrices and age-specific infection and incidence rates forecasted in 2050 for India and Ethiopia, from the full model (*Left* column) and reduced model 2 (*Center* column), with the difference (full model minus reduced model 2) (*Right* column) shown in each case. Differences in incidence and infection rates are shown in gray when they are not statistically significant; otherwise they have the color associated to the model for which the rate is higher. When empirical contact patterns are introduced in the model, we observe an increased density of infections close to the diagonal (i.e., contagions taking place between individuals of similar age) and fewer infections taking place from children to adults or vice versa. At the level of infection/incidence rates, this translates into an underestimation of burden among adults (and an overestimation among children) associated to assuming contacts homogeneity in the reduced model.

Discussion

The model presented here was specifically designed to provide a suitable description of TB transmission dynamics in situations where demography is evolving at the same time that the epidemics unfold. Importantly, we showed that considering current populations' aging trends in TB transmission models is followed by a systematic increase in burden forecasts. This worrisome result can be understood in terms of the known mechanisms whereby age affects the transmission dynamics of the disease. In TB, adults are affected by higher age-specific burden levels than children, and, at the same time, they are more efficient spreaders, given their increased tendency to develop infectious forms of pulmonary TB (8). As a consequence, considering populations' aging translates into higher burden forecasts, simply by increasing the fraction of older individuals with respect to children.

These results suggest that the decay in TB burden levels that has been observed in most countries during recent decades might be harder to sustain than previously anticipated. Under this view, the socioeconomic and public health improvements that made possible the recent decline of TB worldwide would need to be intensified in many countries if the goal of TB eradication is to be pursued before 2050, at the same time that global aging of the human populations unfolds.

Furthermore, our model incorporates a data-driven description of the dependency of TB transmission routes with age, which, we have showed, exerts a significant influence on the forecasted age distribution of the disease burden, by reshaping infection routes. These results will impact the evaluation and comparison of novel epidemiological interventions, mostly if they are conceived to target specific age strata. That is the case of new preventive vaccines aimed at substituting or improving Bacillus Calmette–Guérin and their application through age-focused vaccination campaigns. In this context, previous works have concluded that a quick immunization of young adults through vaccination campaigns focused on adolescents is expected to produce a faster decline in TB incidence than an alternative strategy based on newborns' vaccination (21). Our results would further reinforce this hypothesis, to the extent that empirical contact patterns are followed by a relative increase of TB among adults with respect to children. It must be noted, though, that in the decision of what is the optimal age group to target in a hypothetical immunization campaign for a new vaccine, at least two additional aspects have to be considered, namely, whether the new vaccine is conceived to boost or substitute Bacillus Calmette–Guérin and whether previous exposure to environmental antigens is expected to compromise vaccine performance [via "blocking" (28, 29)]. Bacillus Calmette–Guérin substitutes and/or vaccines susceptible to lose immunogenicity due to exposure to mycobacterial antigens of individuals before vaccination might not be eligible for adolescent immunization anyway.

Despite all the improvements introduced in this work, our approach is not exempt from the strong limitations that affect all TB transmission models operating at this level of resolution. The outcomes of our model depend on a series of epidemiological parameters and initial burden estimates that are subject to strong sources of uncertainty. Even though we have registered these uncertainties and propagated them to the final model outcomes, future improvements and reassessment of these pieces of input data are generally expected to impact the quantitative outcomes of the model and to further delimitate the uncertainty ranges here reported. As a first example, International Health authorities insist on the importance of implementing systematic surveys of TB prevalence in many countries, as a means toward more accurate TB burden evaluations. Accordingly, they revise and update their burden estimates on a regular basis, as new data become available, which obviously impacts model calibration and results. Furthermore, the demand for epidemiological studies aimed at obtaining updated estimates of key epidemiological parameters in current epidemiological settings has been identified as a primary need for the development of more reliable TB models (30). Similarly, we have seen here the importance of obtaining data on contact patterns specific for each setting, by showing that different contact structures inferred from studies conducted in different parts of the world yield significantly different distributions of TB burden

across age (*SI Appendix*, Fig. S10). Importantly, the interpretation of these burden distributions of TB across age is hindered by the limited quality of the data available regarding TB distribution across age, which makes adventurous any comparison between model and data. For example, current WHO data structure splits TB incidence into only two major age groups (0–14 y vs. 15+ y), with alleged, heavy underreporting biases among children.

All these considerations, taken together, evidence the need of further studies, spanning from the implementation of systematic surveys that could unlock more accurate burden estimations (either aggregated or, very importantly, age specific) to the reestimation of key epidemiological parameters and contact patterns in specific epidemic settings.

Despite those limitations, in this work we have shown that abandoning the simplifications of constant demography and homogeneous contacts shared by previous models of TB transmission is not just technically feasible, but has significant effects on model outcomes. Remarkably, TB is not the only disease where long characteristic timescales and strong age dependencies concur (31, 32), which, despite the specific details of the transmission dynamics of each case, implies that similar corrections to what we have proposed here for the case of TB might be pertinent to correct bias of current epidemic models of other diseases too.

Materials and Methods

TB Natural History. The description of the natural history of the disease that we use in our model (Fig. 1*A*) is largely based on previous works by Dye and colleagues (8, 20), with fewer variations to make it compatible with the structure of data reported by WHO regarding disease type and treatment outcomes (27). Specifically, we deal with a compartmental, age-structured model based on ordinary differential equations, which was implemented in the programing language C through a fourth-order Runge–Kutta algorithm (time step = 1 d). The model presents two different latency paths to disease—fast and slow—and six different situations of disease, depending on its etiology (nonpulmonary, pulmonary smear negative, or pulmonary smear positive, characterized by an increasing infectiousness) and on treatment status. After disease, we explicitly consider the main treatment outcomes included in WHO data schemes: treatment completion, default, failure, and death, as well as natural recovery. The natural history model and transitions between the different states, including exogenous reinfections, endogenous reactivations, mother–child transmission, and smear progression (i.e., the transition from smear negative to positive during an episode of active TB) (7, 8, 13, 20, 27, 33, 34), are thoroughly detailed in *SI Appendix*, Fig. S12.

Age Structure and Demographic Evolution. The transmission dynamics defined by the natural history described above are executed in parallel in n = 15 age groups of a span of 5 y each, except for the last one, which contains all individuals older than 70 y (omitted from Figs. 3 and 4 to facilitate visual reading of the scales). The internal transitions between disease states within age groups are then complemented by transitions between age groups representing individuals' aging (Fig. 1*B*). That defines, per each age group, an a priori evolution term $\dot{N}_o(a,t)$ that describes the uncorrected time derivative of the population in age group a, at time t. Then, to make our demographic structures reproduce the curves reported by the UN Population Division, these empirical data series are fitted to smooth polynomials that are then derived to obtain $\dot{N}_{UN}(a,t)$. Finally, a correction term $\Delta_N(a,t) = \dot{N}_{UN}(a,t) - \dot{N}_o(a,t)$ is added to the uncorrected evolution in such a way that the final time derivative of the demographic structure, defined as $\dot{N}(a,t) = \dot{N}_o(a,t) + \Delta_N(a,t)$, verifies $\dot{N}(a,t) = \dot{N}_{UN}(a,t)$ by construction. This, along with the initialization of the population structures according to the UN data, ensures that the evolution of the demography reproduces the UN prospects for all countries and time points. The correction term $\Delta_N(a,t)$ represents the population variations that occur for causes foreign to TB: new births, introduced as susceptible individuals in the first age group, except for the fraction that undergoes perinatal infection (*SI Appendix*, section 2.1.10), as well as TB-unrelated deaths and migrations, which are distributed, for the rest of the age groups, among the different disease states proportionally to their respective volumes (see *SI Appendix*, section 2.5 for further details).

Countries Analyzed. The analyses presented in this paper were performed in India, Indonesia, Nigeria, and Ethiopia, four countries that were selected for their assorted geographic contexts, populations' aging prospects, and TB burden trends. In *SI Appendix*, section 1.1, we also analyzed 8 more countries—Pakistan, South Africa, Bangladesh, Democratic Republic of Congo, Myanmar, Tanzania, China, and The Philippines (*SI Appendix*, Fig. S1)—thus covering the 12 countries affected by the highest levels of TB in the world, measured in total numbers of incident cases.

Empirical Contact Patterns. Empirical data of age-dependent contact patterns have been adapted from statistical surveys conducted in different countries in Africa, Asia, and, as a control, Europe. In each case, contact matrices from studies conducted in different countries of the same continent [Kenya (22), Zimbabwe (23), and Uganda (24) in Africa; China (25) and Japan (26) in Asia; and Belgium, Germany, Finland, Great Britain, Italy, Luxembourg, The Netherlands, and Poland in Europe (14)] have been processed according to the following steps.

First, contact matrices from each study $\xi_s(a,a')$ are corrected to preserve symmetry (i.e., to make the total number of contacts between age-groups a and a' compatible with survey responses from both groups, conditioned by the demographic structure of the population of each study) and normalized to a common scale. Then, matrices corresponding to studies made on the same continent are averaged, weighted according to the number of participants in each study. As a result, we obtain one matrix per region $\xi_{reg}(a,a')$, which also guarantees that the reports of the contact frequency between a and a' are compatible, generating the same number of total contacts, given the demography of the region (i.e., the union of the countries being averaged at the time of the studies):

$$N_{reg}(a)\xi_{reg}(a,a') = N_{reg}(a')\xi_{reg}(a',a). \quad [1]$$

Second, to be able to use these averages in specific settings with different demography, we interpret the matrices $\xi_{reg}(a,a')$ as the product of two nuisance factors: the fraction of individuals in a' that exist in the population, $\frac{N_{reg}(a')}{N_{reg}}$, and an auxiliary matrix $\pi_{reg}(a,a')$:

$$\xi^{norm}_{reg} = \pi_{reg}(a,a')\frac{N_{reg}(a')}{N_{reg}}. \quad [2]$$

Under this interpretation, the auxiliary matrices $\pi_{reg}(a,a')$ capture the "intrinsic" intensity of contacts between groups a and a', once the effect of the demography has been removed, except for a common scale factor.

Next, the matrices $\pi_{reg}(a,a')$ of each region, as inferred from Eq. **2**, are adapted to the specific demography of the countries analyzed in this work. Contacts derived from studies conducted in Asia are applied in India and Indonesia, while contacts proceeding from the African studies are applied in Nigeria and Ethiopia. (European contacts are used only as a control in *SI Appendix*.) This yields the country-specific matrices

$$\tilde{\xi}_c(a,a',t) = \pi_{reg}(a,a')\frac{N_c(a',t)}{N_c(t)} \quad [3]$$

which allow us to incorporate the influence of the evolving demography on the contact structure of our model automatically. Finally, $\tilde{\xi}_c(a,a',t)$ is normalized dynamically at each time step to obtain the final contact matrices used in our model, denoted as $\xi_c(a,a',t)$. These matrices represent, at any time, the contact frequency that an individual of age a has with individuals of age a', relative to the overall frequency of contacts that any individual has with anyone else in the system (see *SI Appendix*, sections 2.2 and 2.3 and Fig. S13 for further details).

Data Flux and Model Calibration. The flux of data is summarized in Fig. 1*C*. The model makes use of four different types of inputs, including (*i*) each of the 19 literature-based epidemiological parameters (7, 8, 13, 20, 33, 34) (*SI Appendix*, Table S14); (*ii*) TB burden and treatment outcome proportions [reported at the WHO TB database (27), accessed on November 16, 2016 (*SI Appendix*, Table S15)]; (*iii*) contact patterns [estimated from different survey studies conducted in Africa (22–24) and Asia (25, 26)]; and (*iv*) demographic prospects reported in the UN Population Division database (accessed on November 16, 2016).

All these input data are integrated at the step of model calibration, whose goal is to reproduce the time series of aggregated incidence and mortality reported by the WHO for each country in the period 2000–2015. To achieve this goal, the initial conditions of the system and the values of

the only two parameters that do not proceed from bibliographic sources (the scaled infectiousness and the diagnosis rate) are estimated for each country. This procedure is completed using the Levemberg–Marquard optimization algorithm implemented in the C library levmar (*SI Appendix*, section 2.8 and Fig. S14). These two fitted parameters, which define a scale for the number of secondary infections caused by each infectious agent, as well as for the number of cases diagnosed per unit time in each country (*SI Appendix*, Fig. S15), are allowed to vary in time, as in previous works (20), to illustrate socioeconomic improvements that might impact the ability of public health systems to better control the disease and restrain its transmission. Finally, the estimates of the initial conditions and the fitted parameters are integrated with the rest of the inputs to produce model forecasts up to 2050.

Uncertainty and Sensitivity Analysis. The uncertainty of each independent source of input data was propagated to model forecasts. The contribution to overall uncertainty assigned to each type of input (epidemiological parameters, WHO estimates of TB burden and treatment outcomes, contact patterns, and demographic prospects) was calculated by repeating model calibration and forecast steps in a series of alternative scenarios where each uncertainty source is shifted sequentially from its expected value to its confidence interval limits. Finally, the deviations from the central estimate that correspond to these alternative scenarios are aggregated assuming mutual independence and linearly weighted to generate the final confidence intervals shown in Fig. 2 for aggregated burden projections and in Figs. 4 and 5 for incidence and infection rates within age group. In *SI Appendix*, section 1.3 and Fig. S3, the individual contribution of all single sources of uncertainty on aggregated incidence and mortality is disclosed, with red (blue) bars representing changes in burden rates associated to an increase (decrease) of the parameters/uncertainty sources from the central values. An important feature of this method is that it allows us to test how sensitive our forecasts are to inputs' uncertainty upon model recalibration, instead of testing the intrinsic sensitivity of the dynamics of the noncalibrated model to each input (see *SI Appendix*, section 4, for further details).

Further Specifications. For further model details, including definitions of model states, explicit enunciation of model differential equations, parameter values, and uncertainties, the reader is referred to *SI Appendix*, sections 2–4.

ACKNOWLEDGMENTS. We thank M. Gutierrez for assistance with figures. S.A. was supported by the Formación de Profesorado Universitario program of the Government of Aragón, Spain, and J.S. by the postdoctoral training program for nonresidents of Quebec from the Fonds de Recherche du Québec–Santé and by the Canadian Institutes of Health Research through a Banting fellowship. This work was partially supported by Gobierno de Aragón/Fondo Social Europeo', by Ministerio de Economía, Industria y Competitividad and Fondo Europeo de desarrollo regional funds through Grants FIS2014-55867-P and BIO2014-52580P, by Project FIS 15/0317 (to S.S. and M.J.I.), by Project TBVAC2020 (643381) funded by the European Commission H2020 (to C.M. and D.M.), and by the European Commission Proactive project Foundational Research on MULTIlevel comPLEX networks and systems Contract 317532 (to Y.M.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

1. Dormandy T (1999) *The White Death: A History of Tuberculosis* (Hambledon Press, London).
2. Borgdorff MW, Floyd K, Broekmans JF (2002) Interventions to reduce tuberculosis mortality and transmission in low-and middle-income countries. *Bull World Health Organ* 80:217–227.
3. Lienhardt C, et al. (2012) Global tuberculosis control: Lessons learnt and future prospects. *Nat Rev Microbiol* 10:407–416.
4. World Health Organization (2017) *Global Tuberculosis Report 2017* (World Health Organization, Geneva).
5. Dye C, Williams BG (2000) Criteria for the control of drug-resistant tuberculosis. *Proc Natl Acad Sci USA* 97:8180–8185.
6. Boily M, Lowndes C, Alary M (2002) The impact of HIV epidemic phases on the effectiveness of core group interventions: Insights from mathematical models. *Sex Transm Infect* 78:i78–i90.
7. Korenromp EL, Scano F, Williams BG, Dye C, Nunn P (2003) Effects of human immunodeficiency virus infection on recurrence of tuberculosis after rifampin-based treatment: An analytical review. *Clin Infect Dis* 37:101–112.
8. Abu-Raddad LJ, et al. (2009) Epidemiological benefits of more-effective tuberculosis vaccines, drugs, and diagnostics. *Proc Natl Acad Sci USA* 106:13980–13985.
9. Garnett GP, Cousens S, Hallett TB, Steketee R, Walker N (2011) Mathematical models in the evaluation of health programmes. *Lancet* 378:515–525.
10. Byng-Maddick R, Noursadeghi M (2016) Does tuberculosis threaten our ageing populations? *BMC Infect Dis* 16:119.
11. Lobato MN, Cummings K, Will D, Royce S (1998) Tuberculosis in children and adolescents: California, 1985 to 1995. *Pediatr Infect Dis J* 17:407–411.
12. Dye C, Williams BG (2008) Eliminating human tuberculosis in the twenty-first century. *J R Soc Interface* 5:653–662.
13. Marais B, et al. (2004) The natural history of childhood intra-thoracic tuberculosis: A critical review of literature from the pre-chemotherapy era [state of the art]. *Int J Tuber Lung Dis* 8:392–402.
14. Mossong J, et al. (2008) Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med* 5:e74.
15. Del Valle SY, Hyman JM, Hethcote HW, Eubank SG (2007) Mixing patterns between age groups in social networks. *Soc Networks* 29:539–554.
16. Melegaro A, Jit M, Gay N, Zagheni E, Edmunds WJ (2011) What types of contacts are important for the spread of infections? Using contact survey data to explore European mixing patterns. *Epidemics* 3:143–151.
17. Miller E, et al. (2010) Incidence of 2009 pandemic influenza A H1N1 infection in England: A cross-sectional serological study. *Lancet* 375:1100–1108.
18. Birrell PJ, et al. (2011) Bayesian modeling to unmask and predict influenza A/H1N1pdm dynamics in London. *Proc Natl Acad Sci USA* 108:18238–18243.
19. United Nations (2016) Population division database. Available at esa.un.org/unpd/wpp/index.htm. Accessed November 16, 2016.
20. Dye C, Garnett GP, Sleeman K, Williams BG (1998) Prospects for worldwide tuberculosis control under the WHO DOTS strategy. *Lancet* 352:1886–1891.
21. Knight GM, et al. (2014) Impact and cost-effectiveness of new tuberculosis vaccines in low-and middle-income countries. *Proc Natl Acad Sci USA* 111:15520–15525.
22. Kiti MC, et al. (2014) Quantifying age-related rates of social contact using diaries in a rural coastal population of Kenya. *PLoS One* 9:e104786.
23. Melegaro A, et al. (2017) Social contact structures and time use patterns in the Manicaland province of Zimbabwe. *PLoS One* 12:e0170459.
24. le Polain de Waroux O, et al. (2017) Characteristics of human encounters and social mixing patterns relevant to infectious diseases spread by close contact: A survey in southwest Uganda. bioRxiv:10.1101/121665.
25. Read JM, et al. (2014) Social mixing patterns in rural and urban areas of southern China. *Proc Biol Sci* 281:20140268.
26. Ibuka Y, et al. (2015) Social contacts, vaccination decisions and influenza in Japan. *J Epidemiol Community Health* 70:164–167.
27. World Health Organization (2016) Tuberculosis database. Available at www.who.int/tb/country/en/index.html. Accessed November 16, 2016.
28. Barreto ML, et al. (2014) Causes of variation in BCG vaccine efficacy: Examining evidence from the BCG REVAC cluster randomized trial to explore the masking and the blocking hypotheses. *Vaccine* 32:3759–3764.
29. Arregui S, Sanz J, Marinova D, Martín C, Moreno Y (2016) On the impact of masking and blocking hypotheses for measuring the efficacy of new tuberculosis vaccines. *Peer J* 4:e1513.
30. Dowdy DW, Dye C, Cohen T (2013) Data needs for evidence-based decisions: A tuberculosis modeler's 'wish list'. *Int J Tuberc Lung Dis* 17:866–877.
31. Hontelez JA, et al. (2011) Ageing with HIV in South Africa. *AIDS* 25:1665–1667.
32. Griffin JT, Ferguson NM, Ghani AC (2014) Estimates of the changing age-burden of plasmodium falciparum malaria disease in sub-Saharan Africa. *Nat Commun* 5:3136.
33. Picon PD, et al. (2007) Risk factors for recurrence of tuberculosis. *J Bras Pneum* 33:572–578.
34. Pillay T, Khan M, Moodley J, Adhikari M, Coovadia H (2004) Perinatal tuberculosis and HIV-1: Considerations for resource-limited settings. *Lancet Inf Dis* 4:155–165.

A data-driven model for the assessment of Tuberculosis transmission in evolving demographic structures. Supplementary Information

Sergio Arregui[a,b], María José Iglesias[c,d], Sofía Samper[d,e], Dessisslava Marinova[c,d], Carlos Martín[c,d,f], Joaquín Sanz[g,h,*], and Yamir Moreno[a,b,i,*]

[a]Institute for Biocomputation and Physics of Complex Systems (BIFI), University of Zaragoza, Spain
[b]Department of Theoretical Physics, University of Zaragoza, Spain
[c]Department of Microbiology, Faculty of Medicine, University of Zaragoza, Spain
[d]CIBER Enfermedades Respiratorias, Instituto de Salud Carlos III, Madrid, Spain
[e]Instituto Aragonés de Ciencias de la Salud. IIS Aragon
[f]Service of Microbiology, Miguel Servet Hospital, Aragón, Spain
[g]Saint-Justine Hospital Research Center, Montreal, Canada
[h]Department of Biochemistry,University of Montreal, Canada
[i]Complex Networks and Systems Lagrange Lab, Institute for Scientific Interchange, Turin, Italy
[*]These authors contributed equally to this work

Contents

1 Auxiliar results

1.1 Fit and forecast for the top 12 countries with highest absolute TB burden

In the main text we have performed the majority of our analyses in India, Indonesia, Nigeria and Ethiopia, which were selected for their different initial age-structures, TB burden trends and demographic shifts. In figure S1, we extend our analysis to the 12 countries affected by the highest absolute TB burden levels in 2015.

As seen in the figure, our model is able to reproduce satisfactorily the observed trends in 2000-2015 in all cases. In table S1 we show the different values of the fitting rest H for these countries. In all of them, except China and Philippines, the rest represents the residual sum of squares normalized by the input data uncertainty of each measure. In the last two examples, where confidence interval limits coincide with central estimates in some cases, we normalized the squared residuals by the average incidence and mortality rates, respectively (see section 2.8 for further details).

For a better understanding of the quality of the fit we also show the average relative difference between the data points and the outcomes obtained with our model once fitted during the fitting window. However, notice that this is not the magnitude that we optimize during our calibration process, it only gives an interpretation of the agreement with the data.

Country	H	**Relative Error (%)**
India	0.247	1.44
Indonesia	0.0241	0.86
Nigeria	0.0749	1.32
Pakistan	1.02	10.52
South Africa	0.790	4.95
Bangladesh	0.201	3.65
Dem. Rep. Congo	0.213	2.17
Ethiopia	0.671	3.93
Myanmar	1.14	5.87
Tanzania	2.13	7.62
China*	0.053	3.98
Philippines*	0.19	6.94

Table S1: Values of the rest H and relative error for the fit (averaged in the fitting window) for the countries fitted in this work. For the ten first countries, the rest is defined as the sum of the squared deviations between model and data, normalized by the uncertainty of input data in each case (see equation 52). In China and Philippines (*, see equation 55) the normalization factors are set up as the average incidence and mortality during the training period, thus their H values are not in the same scale of the rest of the countries. Further details are provided in section 2.8

From Figure S1 we see that, in general, incidence and mortality forecasts from the reduced model 1 are systematically lower than those obtained when the demography's dynamics is accounted for. These differences are explored in more detail in section 1.4

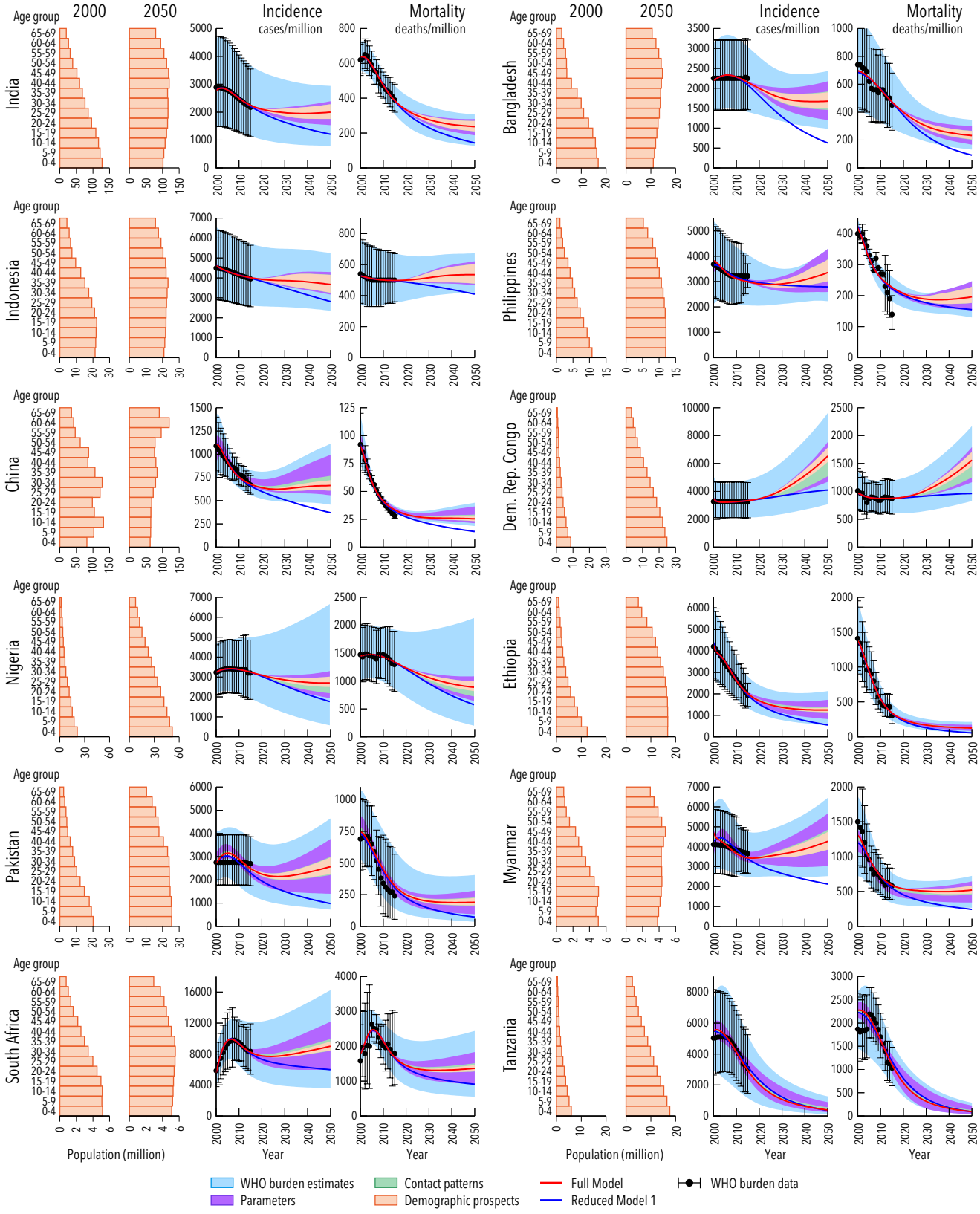

Figure S1: Demographic pyramids in 2000 and 2050, and incidence and mortality projections (2000-2050) for the 12 countries with the largest number of incident cases in 2015. Red lines correspond to the projections derived from our full model, and the different colored areas correspond to the uncertainties that stem from the different input data sources. Blue lines correspond to the predictions made when demography is considered to be constant (reduced model 1).

1.2 Case Notifications

The Case Notification Ratio (CNR), i.e., the fraction of all incident cases that a particular Health System detects and notifies to the WHO-surveillance systems each year, is a fundamental magnitude in the surveillance and control of TB. As a means of validating our model and its calibration procedure, we interrogated whether CNR values reported by the WHO in each country correlate to model-based Treatment coverage ratios, (defined as the fraction of incident cases per year that get diagnosed) despite the fact that these magnitudes are not considered or compared during the calibration step. In figure S2 we represent the results of this comparison in 2015, where we see that model-based treatment coverage fractions are strikingly correlated to CNR values reported by WHO across countries (Pearson correlation excluding Indonesia: $r = 0.96$, $p = 4.3e-6$), which reinforces the validity of our model.

Importantly, we find that treatment coverage is slightly higher than the CNR in every country, which can be interpreted in terms of under-reporting of diagnosed cases. The CNR and model-based Treatment Coverage ratios are closely related, but not fully equivalent, since a fraction of the total TB cases that a country detects and treats each year goes undetected to the WHO surveillance systems, despite TB notification being mandatory in most of the countries analyzed. An exception is found in Indonesia, not included in the figure, where our model predicts a Treatment coverage that is much higher than the CNR (Treatment coverage: 81.5% (CI: 78.3-83.7), CNR: 33% (CI: 23-50)). Precisely in Indonesia, previous studies have pointed out the presence of significant levels of TB under-reporting to the WHO surveillance systems [1], partly related to the fact that in this country the notification of TB cases is not mandatory.

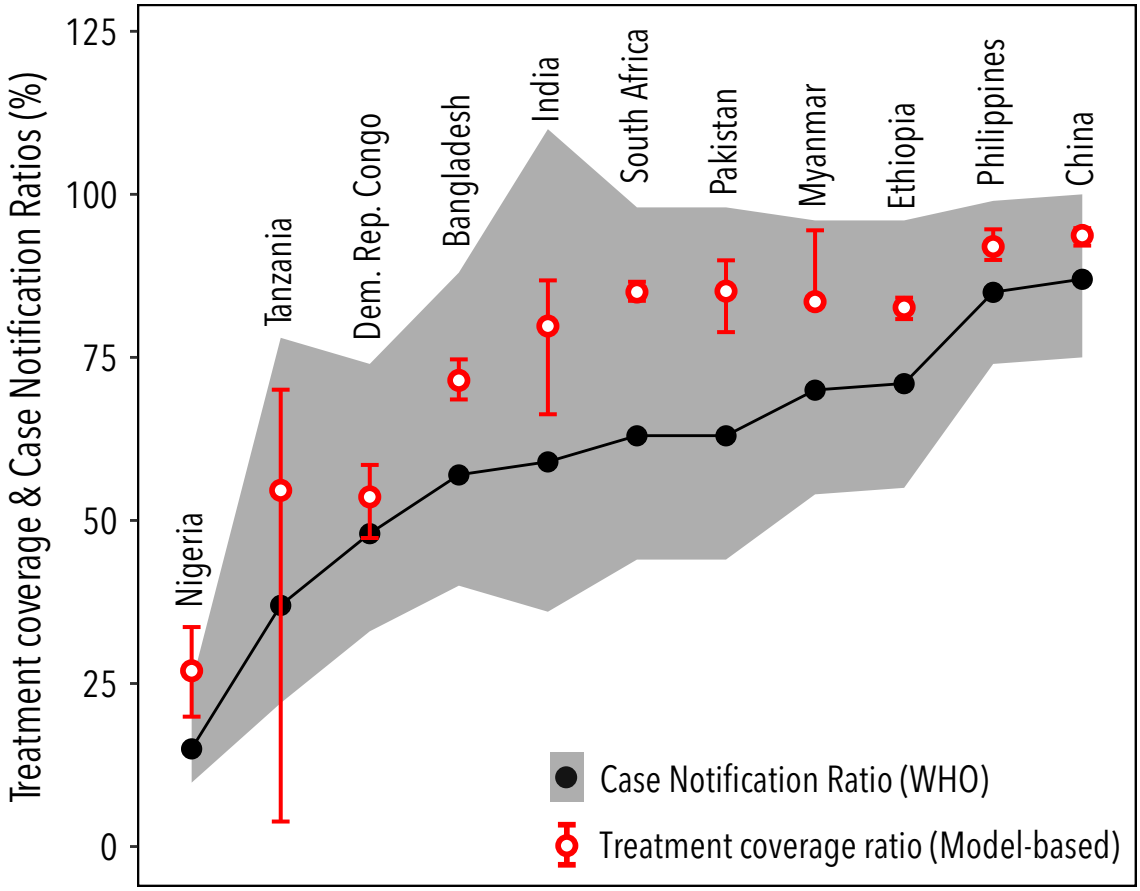

Figure S2: Treatment coverage and associated uncertainty obtained with our model (red symbols) and Case Notification Rates, with their correspondent Confidence Intervals, extracted from WHO data (grey) in 2015.

1.3 Sensitivity analysis

In figure 2 of the main text, as in figure S1, the contribution to overall uncertainty of each type of input data is disclosed: epidemiological parameters (purple), contact matrices (green), demographic prospects (orange) and WHO burden estimates (blue). In figure S3, the contribution of the uncertainty derived from each single epidemiological parameter is further shown.

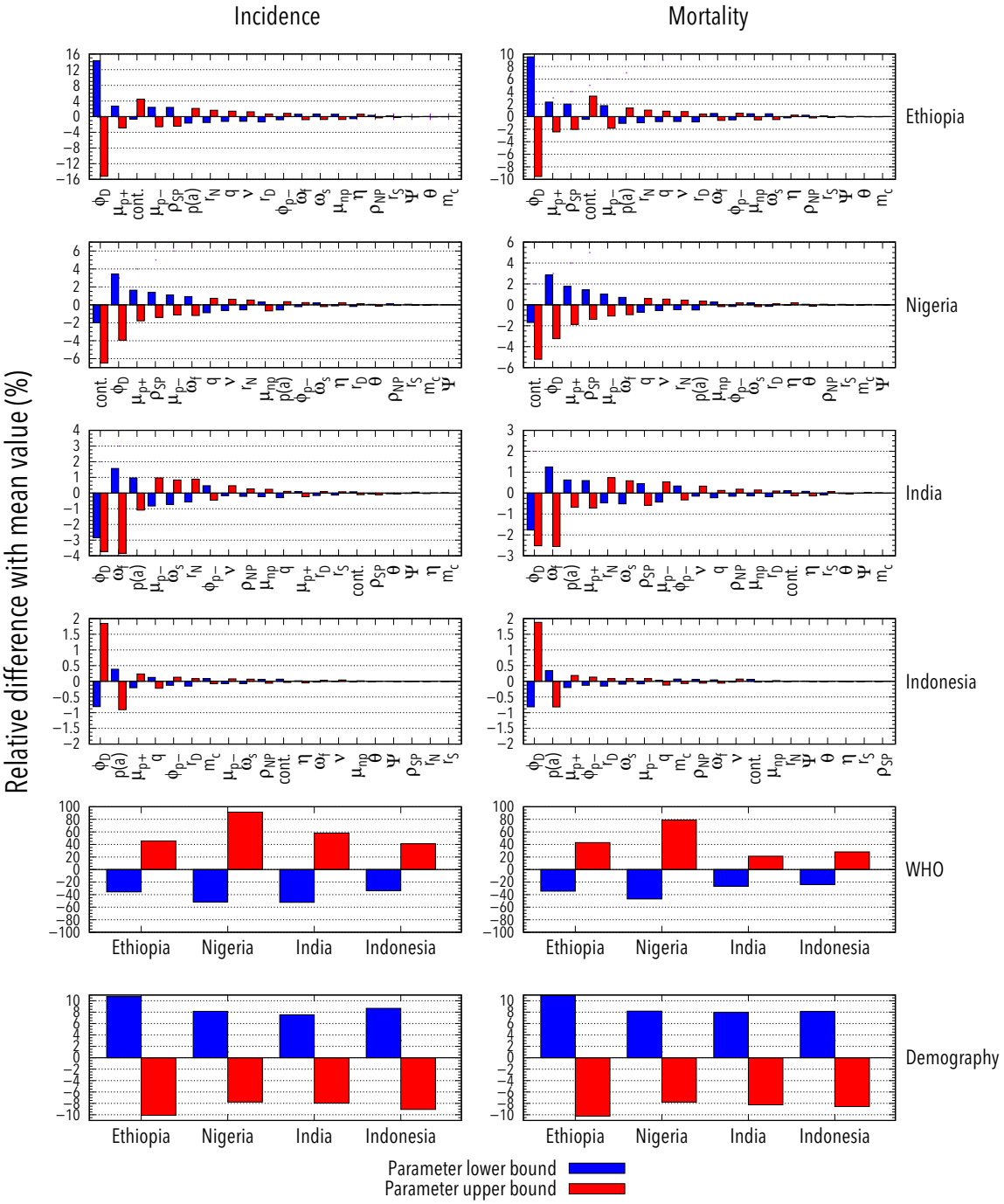

Figure S3: Model sensitivity analysis in India, Indonesia, Nigeria and Ethiopia. Red (blue) bars represent the variations in total number of TB cases/deaths that the model produces in 2000-2050 as a consequence of increasing (decreasing) the value of each uncertainty source to the upper (lower) limit of its respective confidence intervals, prior to model calibration. The bottom panels contain the sensitivities associated to the WHO burden estimates and the demographic projections of the four countries.

1.4 Relative Differences between Full Model and Reduced Model 1

In Figure S1 we have compared the incidence and mortality rates predicted by the full model, (i.e., considering demographic changes), and the reduced model 1 (for which the demographic pyramid is considered constant in time). It still remains pendant to stablish whether the difference between forecasts from the full and the reduced model are statistically significant or if, instead, their uncertainty may be larger than its magnitude. To shed light on this question, in figure S4 we represent, for the twelve countries analyzed, the time evolution of the relative differences between the incidence rates that each model produces. Along with the central estimates for these relative differences, we represent their correspondent uncertainty intervals, obtained as detailed in section 4.1.

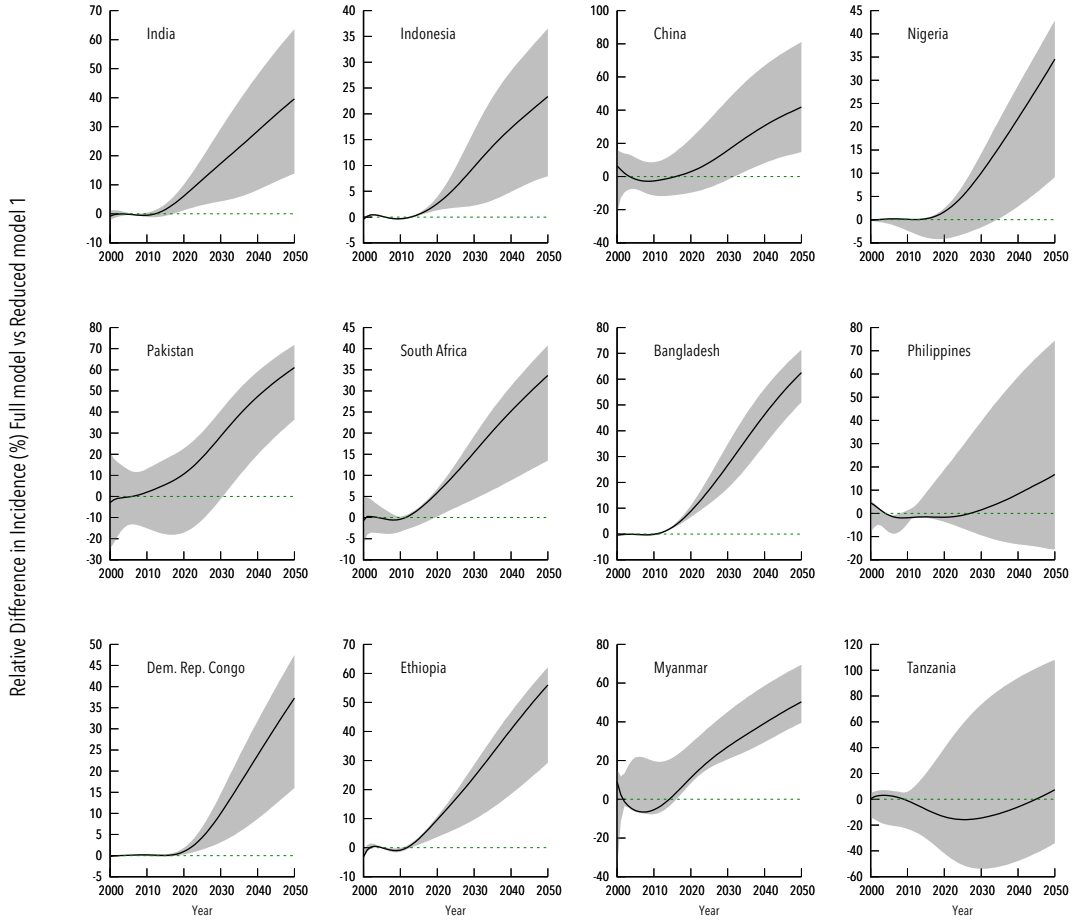

Figure S4: Relative differences of the incidence rate between the full model and the reduced model 1 in the 12 countries considered.

As we can see in Figure S4, differences between the full model and the reduced model 1 become significant in most countries, often right after the end of the fitting window (2015). In Table S2 we register the relative difference between the incidence rates in 2050 as derived from each model, alongside the corresponding 95% confidence intervals and the significance levels. By 2050, differences between models become statistically significant (at 99% significance level) for all countries except two: Philippines and Tanzania.

Country	Relative Difference (%)	Signifcance level
India	39.6 (13.9-63.6)	**
Indonesia	23.4 (7.9-36.5)	**
China	41.9 (14.7-81.1)	**
Nigeria	34.5 (9.1-42.9)	**
Pakistan	61.0 (36.4-71.8)	***
South Africa	33.7 (13.5-40.8)	***
Bangladesh	62.5 (51.1-71.4)	***
Philippines	16.8 (-15.5-74.4)	−
Dem. Rep. Congo	37.3 (16.0-47.5)	***
Ethiopia	56.0 (29.2-62.1)	***
Myanmar	50.3 (39.5-69.5)	***
Tanzania	7.4 (-34.1-108.0)	−

Table S2: Relative difference in the incidence rate in 2050 between full and reduced model 1 for the 12 countries with more TB cases in 2015. Significance levels: −: not significant, ∗:95%, ∗∗:99%, ∗ ∗ ∗:99.9%.

1.5 Robustness tests: effect of demographic evolution

During the next section we will study how the main result of the work (i.e., the higher TB burden predicted as a consequence of considering demography evolution in the model), is robust under a series of alternative modeling scenarios.

Different burden levels

All the forecasts produced in this work are based on estimates of incidence and mortality reported by the World Health Organization. These estimates are based on a combination of epidemiological observations and empiric criteria that are known to be affected by high levels of uncertainty, as reflected by the large confidence intervals that these data present in many countries, which has been taken into account and incorporated to our model forecasts. However, the broad nature of these estimates often leads to re-evaluation of methods and values reported by the WHO in their periodic TB reports, with the result that, in some cases, the burden estimates vary beyond the range of uncertainty initially assumed as a result of these methodological updates. For example, the burden estimates that we use for India turned out to be underestimated in previous publications [2].

Taking it into account, the question of whether the effects of demographic evolution are robust under wide variations in the input burden data constitutes a valid concern. To show that those effects are indeed robust under a wide range of initial burden levels, in figure S5 we have repeated the simulations for the full and reduced model 1 under alternative scenarios where initial burden levels have been doubled/halved. Checking the relative difference between models in the incidence rate at 2050 (table S3), we see that it remains statistically significant ($p<0.05$) in every case. Thus, even if the data of incidence and mortality are not extremely reliable, and could be proven to be biased in the future, the need to improve current models and incorporate the evolution of demography hold as a general conclusion valid for a wide range of initial burden levels.

Country	Bias	Relative Difference (%)	Significance level
Ethiopia	Double	52.6 (25.6-59.2)	***
	Half	57.9 (31.3-63.8)	***
Nigeria	Double	31.1 (9.9-39.1)	**
	Half	37.3 (12.8-63.4)	**
India	Double	39.0 (13.4-61.2)	**
	Half	40.2 (6.5-61.0)	*
Indonesia	Double	21.1 (6.6-33.8)	**
	Half	27.9 (11.2-41.8)	***

Table S3: Relative differences for incidence rates in 2050 between full and reduced model 1, as obtained from biased TB burden estimations (WHO estimates doubled/halved with respect to reported values). Significance levels: −: not significant, ∗:95%, ∗∗:99%, ∗ ∗ ∗:99.9%.

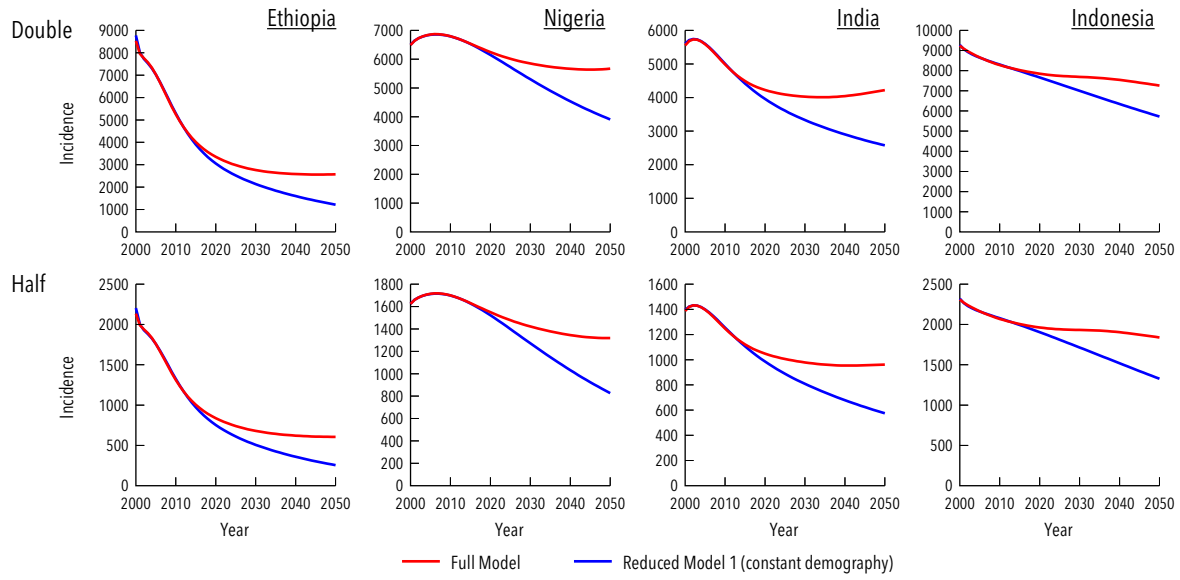

Figure S5: Incidence rate forecasts from full (red) and reduced model 1 (blue) under biased input burden estimates (WHO estimates doubled/halved with respect to reported values).

Different Contact Patterns

During this work, we have implemented two different contact matrices: one adapted from statistical surveys conducted in Africa, and another one from Asian studies. This supposes an arguable improvement with respect to the assumption of contacts homogeneity, and even with respect to using the same contact structure in all countries. However, the available bibliography on empiric contact patterns is still reduced, and we are far away from an ideal situation where empiric data of comparable quality standards is at hand at a country-specific level. In this sense, the question of whether the effects of demography evolution might be dependent or not of the specific contact structure that we use in each country is pertinent, and should be explicitly addressed.

To this end, in figure S6, we show epidemic forecasts derived from the full and the reduced model 1, under scenarios where the contact structure of each country has been substituted by the other matrices considered across the paper, including a contact matrix built from the European Polymod study.

Once again, we have obtained the same result: considering the evolution of demography leads to higher burden prospects, independently of the contact pattern used in our simulations. In table S4 we show the relative differences between models in the incidence rate in 2050 and the associated significance levels.

Contacts	Country	Relative difference (%)	Significance level
African	Ethiopia	56 (29-62)	$***$
	Nigeria	35 (9-43)	$**$
	India	26 (3-53)	$*$
	Indonesia	15 (5-24)	$**$
Asian	Ethiopia	65 (56-76)	$***$
	Nigeria	47 (33-64)	$***$
	India	40 (14-64)	$**$
	Indonesia	23 (8-37)	$**$
European	Ethiopia	67 (57-77)	$**$
	Nigeria	57 (41-75)	$***$
	India	42 (14-61)	$**$
	Indonesia	28 (12-47)	$***$

Table S4: Relative differences for incidence rates in 2050 between full and reduced model 1, evaluated using different contact matrices. Significance levels: $-$: not significant, $*$:95%, $**$:99%, $***$:99.9%.

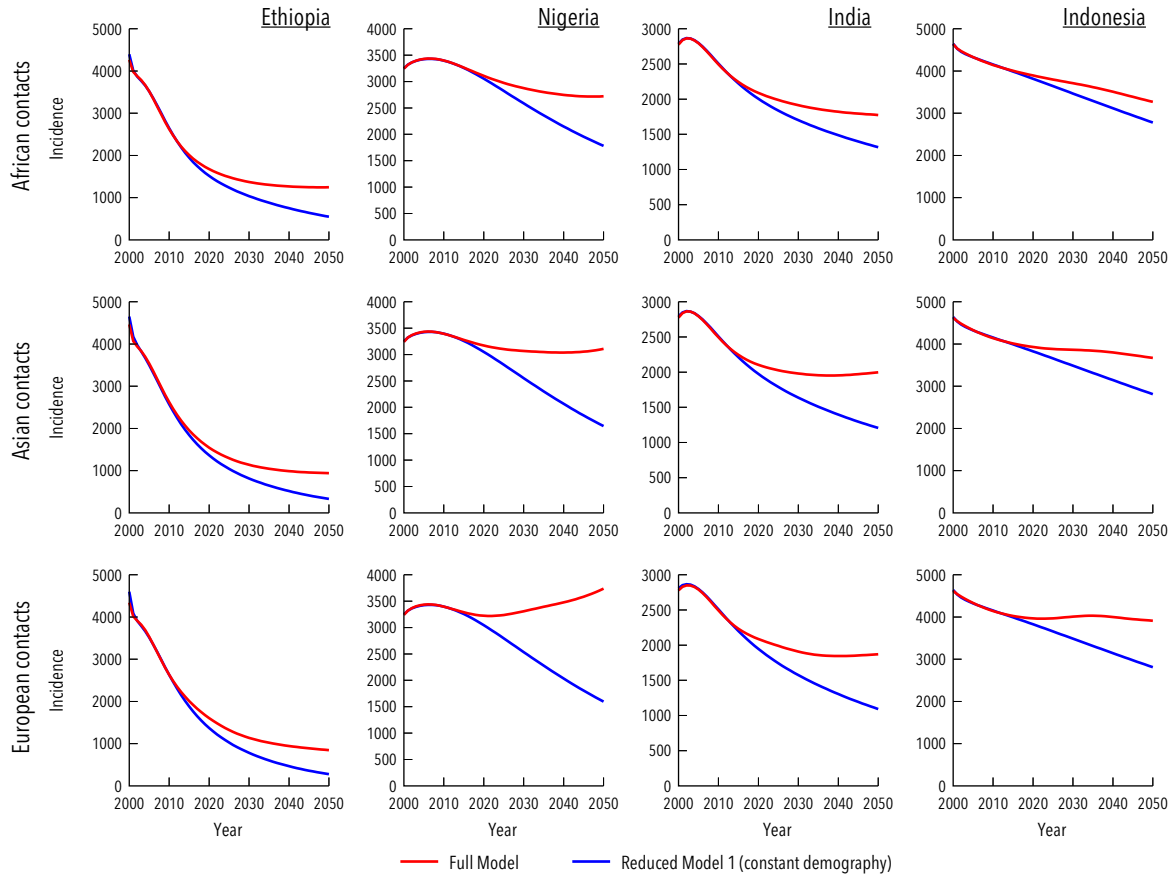

Figure S6: Incidence Rates forecasted from full (red) vs reduced model 1 (blue), for three different contact patterns (African, Asian and European).

Different evolution of fitted parameters

The model forecasts presented in this work are produced under the hypothesis that, after the training period (2000-2015), the time evolution of scaled infectiousness and diagnoses rates will still be governed by the sigmoid curves described by equations 50 and 51. However, if the pace of variation of these parameters slows down in future years from our expected trends, the TB burden rates will increase from the forecasts reported. To explore the behavior of the model in that situation, and re-evaluate the difference between full and reduced model 1, we have repeated the comparison in alternative scenarios where the pace of variation in the fitted parameters is slowed down from 2015 a 50% and a 100% from its expected trend.

Country	Variation rate reduction	Relative difference (%)	Significance level
Ethiopia	50%	56.3 (28.9-62.6)	$***$
	100%	57.1 (28.1-64.2)	$***$
Nigeria	50%	32.5 (10.0-41.1)	$**$
	100%	30.0 (10.6-38.7)	$**$
India	50%	45.5 (27.5-62.7)	$***$
	100%	51.7 (37.4-64.1)	$***$
Indonesia	50%	31.3 (16.8-43.6)	$***$
	100%	37.3 (23.6-49.0)	$***$

Table S5: Relative differences for incidence rates in 2050 between full and reduced model 1, evaluated applying different reductions on the variation rate of the fitted parameters after 2015. Significance levels: $-$: not significant, $*$:95%, $**$:99%, $***$:99.9%.

The underestimation of TB burden that stems from ignoring demographic evolution is also robust against

variations in the time-evolution of fitted parameters after 2015. In table S5 we show the relative difference of the incidence rates in 2050, and we check again that these differences are significant for all alternative scenarios.

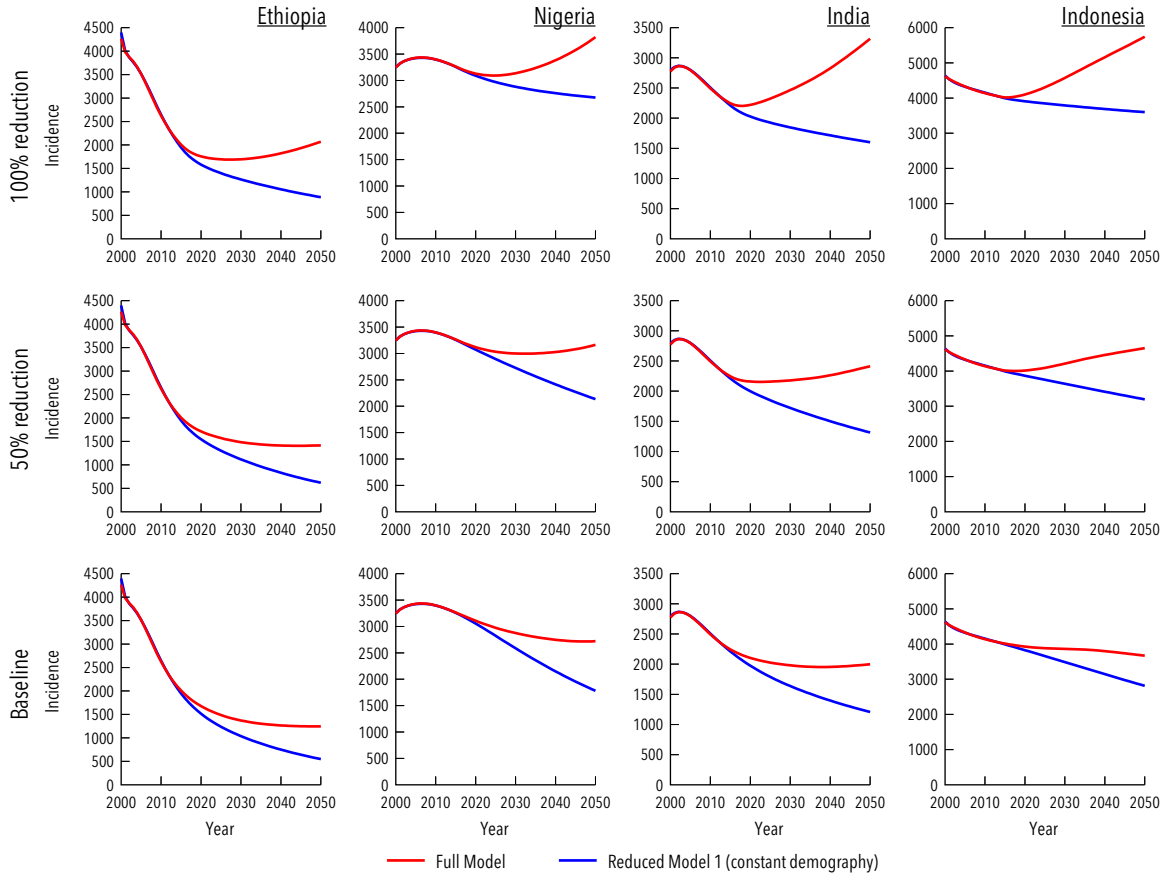

Figure S7: Incidence Rates obtained from the full and the reduced model 1, for 2 alternative scenarios where variation rates of the fitted parameters is reduced a 50% and a 100% from their expected behaviour from 2015 on.

Effect of Demographic Evolution without re-fitting parameters

Comparisons between the full and -for example- the reduced model 1 across the text are performed upon independent calibration of each model, to ensure that both reproduce the initial burden trends independently. However, this procedure does not guarantee that the differences between models arise from the demographic dynamics itself, since they might be a consequence of the different parameters and initial conditions that are estimated in each case. To rule out this possibility, we present here a series of simulations where we use the values for fitted parameters and initial conditions that were estimated upon calibration of the full model, also in the reduced model 1.

In Figure S8 and Table S6, we see that the reduced model 1 still reproduces significantly lower burden projections than the full model even though it is not calibrated to reproduce the data before 2015. Note that the difference with respect to the full model is in many cases amplified.

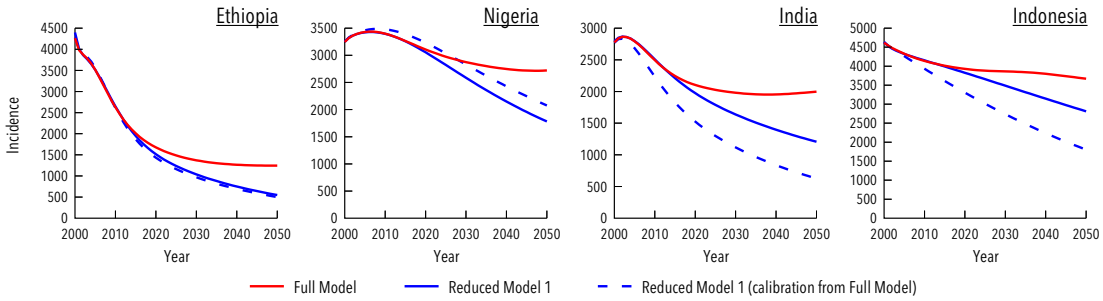

Figure S8: Incidence rate series produced by the full model (red), the reduced model 1 (blue, continuous), and a non-calibrated version of the reduced model 1 which makes use of the same parameters and initial conditions fitted for the full model (blue, dashed line).

Country	Relative Difference in incidence (2050) (%)	Signifcance level
Ethiopia	59.7 (31.7-65.5)	$***$
Nigeria	23.7 (7.4-33.0)	$**$
India	68.6 (59.4-75.7)	$***$
Indonesia	50.8 (37.8 - 61.3)	$***$

Table S6: Relative difference in the incidence rate in 2050 between full and reduced model 1 in Ethiopia, Nigeria, India and Indonesia, when the reduced model 2 is executed using the same initial conditions and fitted parameters inferred upon full model calibration (red minus dashed blue lines in figure S8). Significance levels: $-$: not significant, $*$:95%, $**$:99%, $***$:99.9%.

1.6 Effect of Contact Patterns at the aggregated level

In the main text we have shown that the assumption of homogeneous mixing overestimates the burden of TB in children, and underestimates it among adults. These opposite effects largely cancel each other, which makes the total effect to shrink when considering the aggregated burden across ages. In figure S9 we represent the forecasts of incidence and mortality for the models with heterogeneous and homogeneous mixing patterns (full vs. reduced model 2).

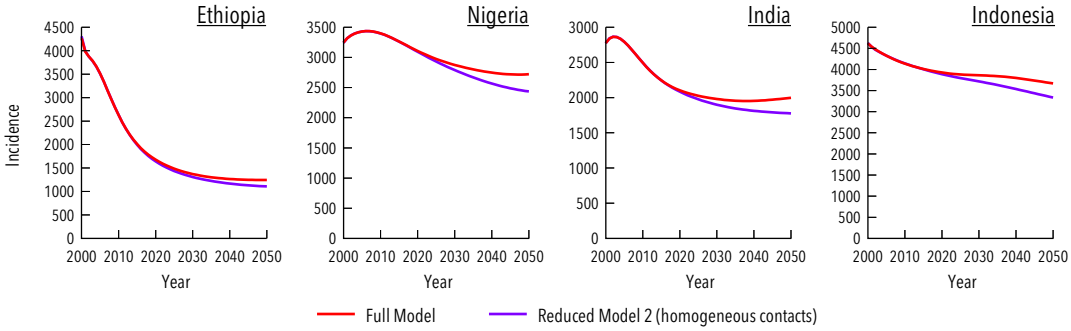

Figure S9: Predictions of incidence and mortality rates for full (red) versus reduced model 2 (violet) in Ethiopia, Nigeria, India and Indonesia

As detailed in Table S7, the usage of empiric contact patterns translates into slightly larger burden rates (relative differences around 10% between the full and the reduced model 2 in 2050). These modest differences are still significant in Ethiopia, India and Indonesia, despite their relatively small values when compared to the magnitude of forecasts' uncertainty. This is not an anomalous behavior, since outcomes from the full and the reduced model are strongly dependent variables, and the uncertainty is propagated to them in a paired fashion

(i.e., the same sources of uncertainty affect both models simultaneously), which is considered when comparing results from model pairs. This allows us to detect significance differences between model behaviors of lower effect sizes than the characteristic uncertainty of each independent model alone.

Country	Relative Difference in incidence (2050) (%)	Significance level
Ethiopia	11.1 (7.8-16.6)	$***$
Nigeria	10.5 (-136-11.7)	–
India	11.1 (2.0-16.6)	$*$
Indonesia	9.1 (2.6 - 15.1)	$**$

Table S7: Relative difference in the incidence rate in 2050 between full and reduced model 2 in Ethiopia, Nigeria, India and Indonesia. Significance levels: –: not significant, $*$:95%, $**$:99%, $***$:99.9%.

1.7 Effect of Contact Patterns on TB burden distribution across age

In the main text we have discussed how the usage of empiric contact patterns (in opposition to the assumption of homogeneous mixing) can change the distribution of TB burden among the different age groups. Regarding this result, it is relevant to note that the contact structures tested in Asian and African countries differ, and thus, a subsequent question is whether or not these empiric data can be interchanged without further effects on TB burden distributions.

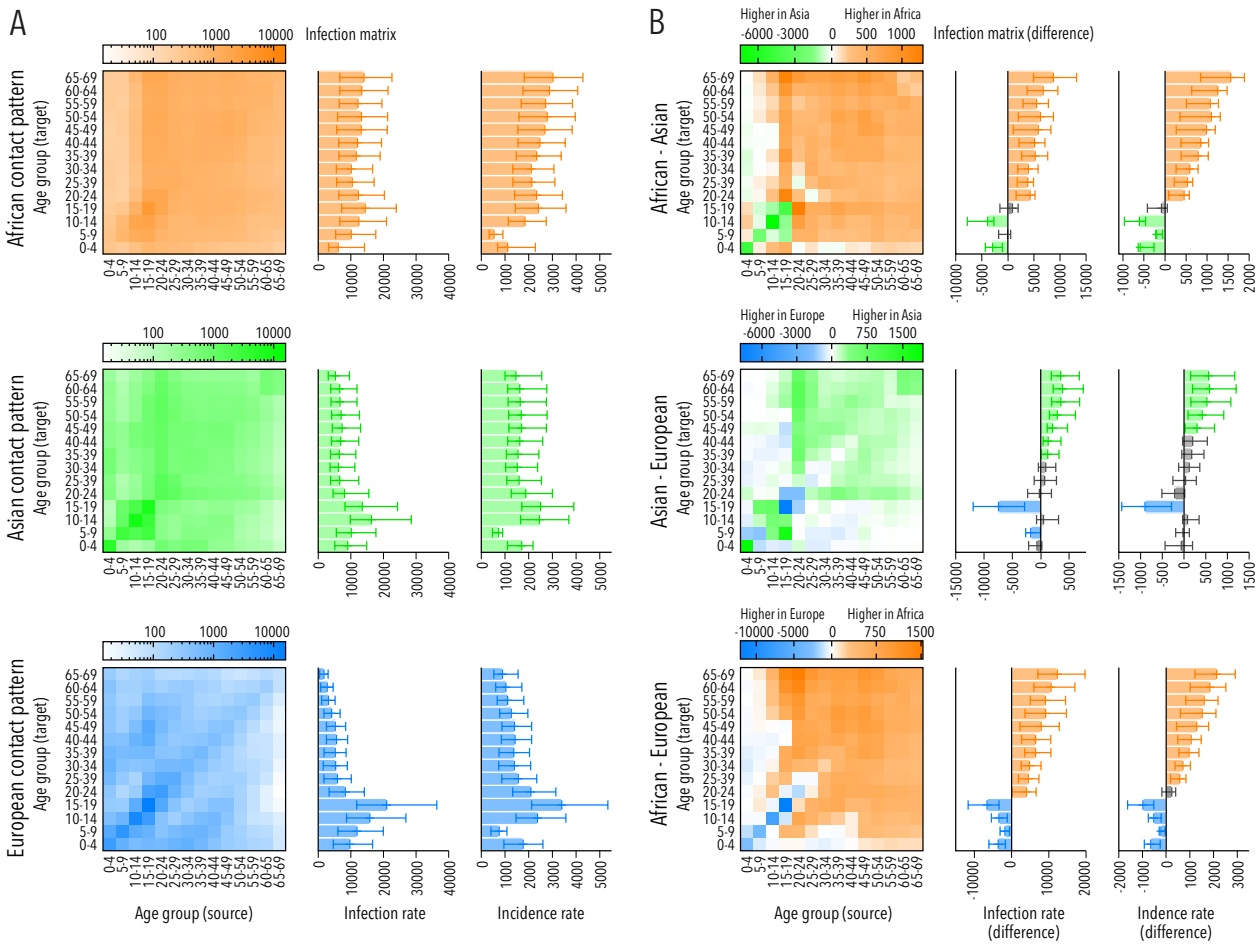

Figure S10: (A): Matrix of infection rates, and age-specific rates of infection and incidence (averaged during the period 2000-2050) corresponding to the use of different contact patterns (African, Asian and European) in Ethiopia. (B): Pairwise differences. Non-significant differences in infection and incidence rates are represented in grey, otherwise they are coloured as the predominant contact pattern.

To answer this question in one particular example, we have chosen one of the countries analyzed, and generated forecasts based on simulations performed using contacts derived from African surveys data (i.e. the default), as well as Asian, and European contact structures. The results of these tests, where the average incidence rate during the period (2000-2050) is reported for each case, are represented in figure S10. In the figure, upon pair-wise comparison between the forecasts associated to each contact matrix, we see that African contacts lead to higher burden among the eldest age-groups, while the European contact patterns tend to induce more infections and TB cases among younger individuals and the matrix used in Asian countries represents an intermediate situation.

1.8 Robustness tests: effects of contact patterns

We also checked the robustness of considering contacts heterogeneity on the age distribution of TB burden in a wide spectrum of different modeling scenarios, as we did for the evolution of demography through section 1.5. In figure S11 we show the age distribution of incidence in 2050 for Ethiopia in 6 scenarios: (A) the base scenario, (B) the case where the reduced model is not recalibrated; (C) a scenario where TB burden data is doubled, (D) a scenario where TB burden data is halved; (E) a scenario where time evolution of fitted parameters is reduced a 50% from 2015, and (F) a scenario where these variation rates are totally arrested from 2015 on. In all these different settings, the assumption of homogeneous mixing between age groups implies the emergence of significant differences between age-specific incidence rates with respect to the hypothesis of homogeneous mixing.

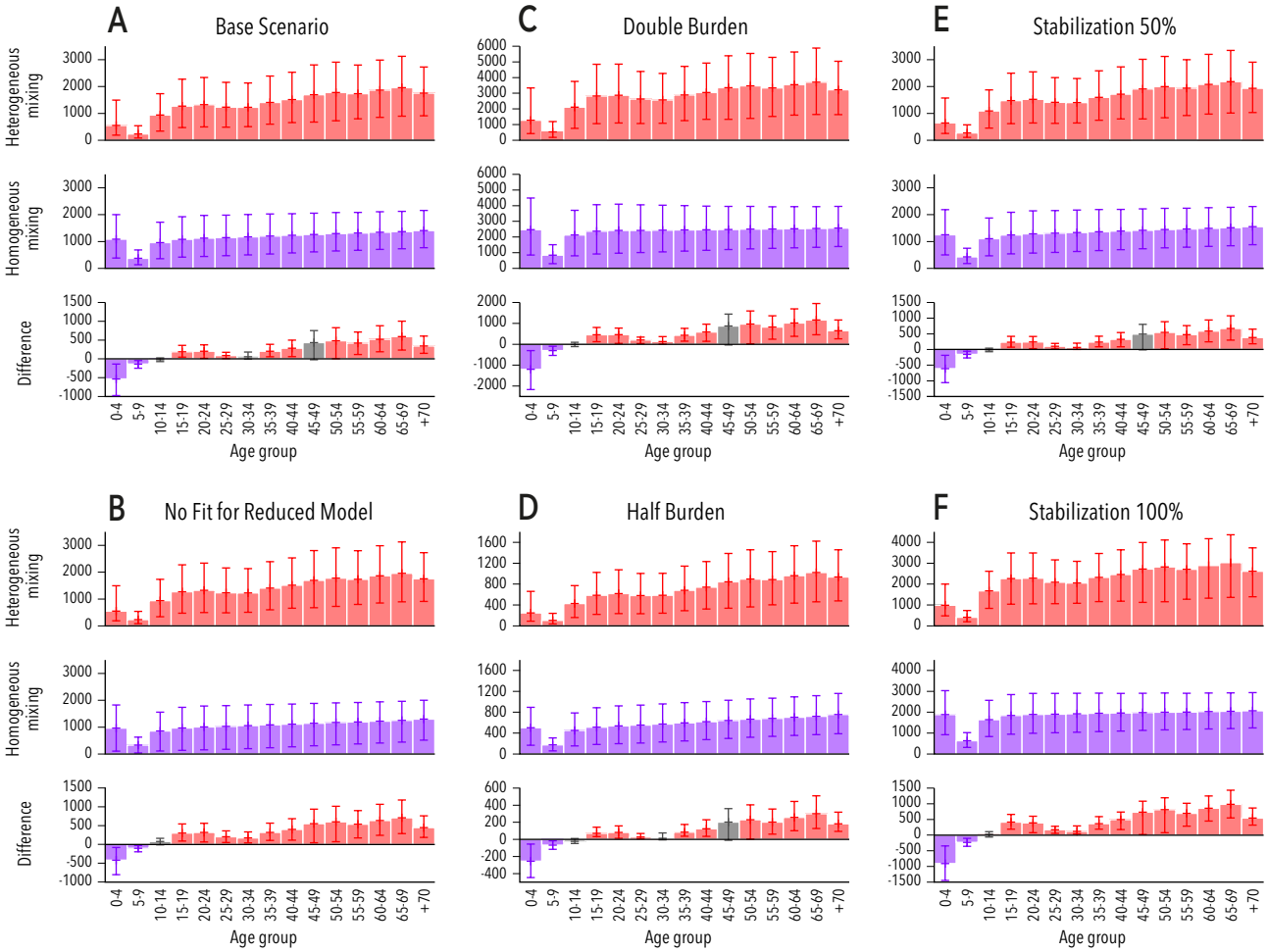

Figure S11: Age specific incidence rates (2050, Ethiopia), for the complete model, the reduced model 2 and the difference between them for different scenarios (same as in section 1.5). Non-significant differences in infection and incidence rates are represented in grey, otherwise they are colored as the predominant model.

2 Model description: technical details

2.1 Natural history of the disease

Our model of Tuberculosis (TB) spreading is essentially based on previous models by C. Dye and colleagues,[3,4] on which new ingredients –heterogeneous contact patterns[5] and an explicit coupling of the demographical evolution and the disease dynamics– have been incorporated. The natural history scheme has also been refined so as to render it more suited to the definitions by the World Health Organization (WHO), mostly in what regards to treatment outcomes.

Summarizing, we deal with an ordinary differential-equations based, age structured model of TB in which we consider a class of unexposed individuals –susceptible–, two different latency paths to disease –fast and slow– and six different kinds of disease, depending on its aetiology: -non pulmonary, pulmonary (smear positive) and pulmonary (smear negative)–, and depending on whether it is untreated or treated. After the disease phase, we explicitly consider the main treatment outcomes contemplated by the WHO data schemes: treatment completion (or success), default, failure and death.[2,6]

The model is structured in 15 age groups, 14 of them covering 5 years of age up to 70 years old, and a last group including all individuals older than 70 years old. When specifying the population of a certain state at a specific age-group and time we will use the notation $X(a,t)$, where X is the concrete state (see list of disease states in table S13), t represents the time, and $a \in [0,14]$ is the index representing any of the fifteen age groups.

In the following, we detail the natural history ingredients and transitions between states that we have considered to build up our model; whose natural history is schematized in figure 1A of the main text and reproduced here in Figure S12 in more detail.

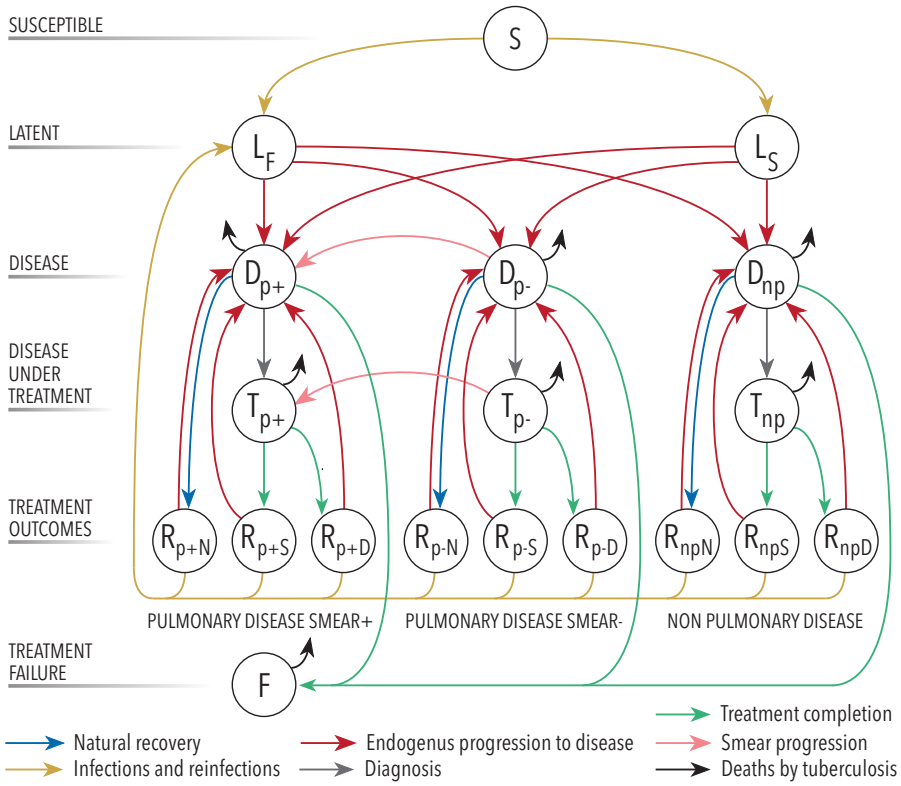

Figure S12: Natural history of the disease: S: susceptible. L: latent. D: (untreated) disease, T (treated) disease, R recovered, F: failed recovery . Types of TB considered: p+: Pulmonary Smear-Positive, p-: Pulmonary Smear-Negative, np: Non-pulmonary. Sub-types of recovery: N: Natural, S: Successful, D: Default (abandon of treatment). A detailed description of each state is provided through section 2.1.

2.1.1 Primary Tuberculosis infection

We call primary the infection of an individual who was not previously exposed to the bacterium: i.e. individuals of class S. If we denote the force of infection $\lambda(a,t)$ as the probability per unit time of any unexposed individual

of age group a of being infected, then the total number of susceptible individuals getting infected per unit time will be approximated by $\lambda(a,t)S(a,t)$. We will address the explicit form of $\lambda(a,t)$ in section 2.2.

Of these newly infected individuals, a fraction $p(a) \in [0,1]$ will experience a quick development of the disease after a short course latency period –fast latency L_f in what follows– characterized by the inability of the host's immune system to restrain mycobacterial growth. In the rest of cases, newly infected individuals' immune system succeeds at containing bacterial proliferation so establishing a host-pathogen dynamic equilibrium that is characterized by an asymptomatic latency state –slow latency L_s in what follows– that can last for the rest of the host's life, or be broken even decades after the infection, typically after an episode of immunosuppression. In conclusion, the primary infection is described as follows:

- Primary infection (to fast latency): transition from $S(a,t)$ to $L_f(a,t)$:
  $p(a)\lambda(a,t)S(a,t)$ individuals/unit time.
- Primary infection (to slow latency): transition from $S(a,t)$ to $L_s(a,t)$:
  $(1-p(a))\,\lambda(a,t)S(a,t)$ individuals/unit time.

It is worth remarking that, within our modeling framework, individuals in latency classes do not have TB disease: they do not develop any disease symptom and they are not infectious at all. Besides, as we will describe in the following sections, they can suffer ulterior re-infections.

Most of the disease parameters that regulate the fluxes of the Natural History are taken from different bibliographic sources. Most of them are taken directly from the works by Dye et al.[3] and Abu-Raddad et al.[4], that form the base of this work. A list with the values of the different disease parameters can be found in section 3. However, the values of $p(a)$ used in this model deserve further explanation.

The risk of fast progression to disease after infection is highly variable, and strongly depends on the age of the individual. In the most recent and complete work on this matter, by Marais et al.[7], a complex outlook regarding this parameter is reported, according to which fast-progression risk is higher in newborns, then decreases in children and increases again during adolescence. This pattern, which is largely accepted in the current literature[8,9,10,11,12], is summarized in Table S8.

Age	**Risk of disease(%)**	$\boldsymbol{p(a)}$
<1 year	50 ± 10	$p(0) = 0.1870\ (0.1474 - 0.2333)$
1-2 years	18.50 ± 5.22	
2-5 years	5.5	
5-10 years	2.25 ± 0.25	$p(1) = 0.0225 \pm 0.0025$
>10 years	15 ± 5	$p(a) = 0.15 \pm 0.05\ \forall a > 1$

Table S8: $p(a)$ values obtained for our model after adapting the values on the risk of developing disease after infection from Marais et al.[7]

The values of Table S8 have been obtained from Marais et al.[7] after summing up pulmonary and non-pulmonary disease risks, and propagating their uncertainties assuming independence. We assign these values as the parameters $p(a)$. For the case of $a = 0$, which corresponds to ages from 0 to 5 years, we have three different values. We construct $p(0)$ as a weighted average:

$$p(0) = \alpha_{<1}p_{<1} + \alpha_{1-2}p_{1-2} + \alpha_{2-5}p_{2-5} \tag{1}$$

The actual values of α_i would depend on the demographic pyramid up to 5 years of age. However that information is not available with the precision required here. Therefore we assume two extreme scenarios: a rectangular pyramid, and a triangular pyramid in which the population of newborns (less than 1 year) doubles that of 4 years old. The former will give us the lower estimate while the latter provides the upper estimate, and the center value is estimated as the average, yielding the final value of $p(0) = 0.1870\ (0.1474 - 0.2333)$.

2.1.2 Progression from latency to (untreated) disease

Either from fast or slow latency, infected individuals can fall sick, progressing to one of the three different active forms of the disease. In the first of these forms, the non-pulmonary disease D_{np}, the pathogen can grow in disparate parts of the host body, including the nervous system, bones, kidneys and other organs foreign to lungs. The main characteristic of this kind of TB is that, since the bacilli can not reach the respiratory tract, the individuals are considered, for the purposes of our model, unable to transmit the disease. However, if the

pathogens proliferate in the lungs, they can eventually reach the upper respiratory tract making its host able to transmit the disease. According to the presence of viable bacilli in the sputum, we have the other variants of TB: pulmonary disease, smear negative D_{p-}, or pulmonary disease smear positive D_{p+}; being the latter more infectious than the former.

This scheme allows six different transitions from the two latency classes to the three untreated TB disease classes:

- Progression from $L_f(a,t)$ to $D_{np}(a,t)$: $\omega_f \rho_{np}(a) L_f(a,t)$ individuals/unit time.
- Progression from $L_f(a,t)$ to $D_{p-}(a,t)$: $\omega_f \left(1 - \rho_{p+}(a) - \rho_{np}(a)\right) L_f(a,t)$ individuals/unit time.
- Progression from $L_f(a,t)$ to $D_{p+}(a,t)$: $\omega_f \rho_{p+}(a) L_f(a,t)$ individuals/unit time.
- Progression from $L_s(a,t)$ to $D_{np}(a,t)$: $\omega_s \rho_{np}(a) L_s(a,t)$ individuals/unit time.
- Progression from $L_s(a,t)$ to $D_{p-}(a,t)$: $\omega_s \left(1 - \rho_{p+}(a) - \rho_{np}(a)\right) L_s(a,t)$ individuals/unit time.
- Progression from $L_s(a,t)$ to $D_{p+}(a,t)$: $\omega_s \rho_{p+}(a) L_s(a,t)$ individuals/unit time.

where ω_f and ω_s represent the rates at which fast and slow progression occur, and $\rho_{p+}(a)$,$\rho_{p-}(a)$ and $\rho_{np}(a)$ represent the probability to develop each of the three different forms of TB previously described: pulmonary smear-positive, pulmonary smear-negative and non-pulmonary, respectively. We are using the closure relation $\rho_{p+}(a) + \rho_{p-}(a) + \rho_{np}(a) = 1$, so we have only two independent parameters –i.e., $\rho_{p+}(a)$ and $\rho_{np}(a)$–. These three probabilities, as the fast progression probability, are age-dependent –children are known to develop more often non-pulmonary forms of TB–. [13]

2.1.3 Tuberculosis related deaths

Individuals in D states suffer the effects of the disease in three ways: 1) they develop disease symptoms; 2) they –except the individuals in D_{np}– infect other individuals and 3) some of them die because of the disease. In the model, we consider that each of the three kinds of disease has a specific mortality rate, so deaths of D individuals are modeled by introducing three independent fluxes:

- Deaths of untreated non pulmonary disease: $\mu_{np} D_{np}(a,t)$ individuals/unit time.
- Deaths of untreated smear negative pulmonary disease: $\mu_{p-} D_{p-}(a,t)$ individuals/unit time.
- Deaths of untreated smear positive pulmonary disease: $\mu_{p+} D_{p+}(a,t)$ individuals/unit time.

where μ_{np}, μ_{p-} and μ_{p+} are the TB-related death rates of D_{np}, D_{p-} and D_{p+} individuals, respectively.

Finally, individuals who were treated in the past that did not respond to treatment will ultimately die of the disease, at the larger rate μ_{p+} regardless of their initial type of disease. Class F is introduced in section 2.1.5 regarding treatment outcomes.

- Deaths of failed recovery individuals: $\mu_{p+} F(a,t)$ individuals/unit time.

2.1.4 TB diagnosis and treatment

In our model, we consider that an individual belongs to D classes until she receives her diagnosis, moment in which she joins the corresponding treated TB class T. This corresponds to the following set of three transitions:

- Diagnosis of non pulmonar TB: transition from $D_{np}(a,t)$ to $T_{np}(a,t)$: $\eta d(t) D_{np}(a,t)$ individuals/unit time
- Diagnosis of smear negative pulmonar TB: transition from $D_{p-}(a,t)$ to $T_{p-}(a,t)$: $\eta d(t) D_{p-}(a,t)$ individuals/unit time
- Diagnosis of smear positive pulmonar TB: transition from $D_{p+}(a,t)$ to $T_{p+}(a,t)$: $d(t) D_{p+}(a,t)$ individuals/unit time

Thus, te diagnosis rate $d(t)$ defines the pace at which undetected individuals in D classes get diagnosed. These diagnosis rates are country specific, as they depend, among other factors, on the capabilities of Public Health systems. Furthermore, the average time needed for TB diagnosis is known to vary depending on the type of disease, partly because the diagnosis criteria used in each type are different too. In our model η represents the variation for the diagnosis rate that is observed for the detection and diagnosis of non smear positive types of disease.

Our estimations for the parameter η are based upon the case detection ratios χ for each type of disease (D_{p+}, D_{p-} and D_{np}) reported by Abu-Raddad and colleagues.[4] The case detection ratio is commonly defined as the ratio of the number of notified cases of TB to the number of incident TB cases in a given year. In Abu-Raddad et al.[4], estimations for the case detection ratios are provided for each type of disease and WHO region: χ_{p+}, χ_{p-} and χ_{np}; and it turns out that according to that source $\chi_{np} \simeq \chi_{p-}$ in all regions. Therefore, if we compare the case detection ratios of non smear positive and smear positive types of the disease we can obtain an estimation for the parameter η for each region:

$$\eta = \frac{\chi_{p-}}{\chi_{p+}} \left(\simeq \frac{\chi_{np}}{\chi_{p+}} \right) \tag{2}$$

The errors have been estimated by considering a 15% as the typical uncertainty of both χ_{p+} and χ_{p-}, as was done in Abu-Raddad et al.[4] for several parameters of the Natural History. We obtain the Confidence Interval for η by propagating errors.

In the table S9 the values of η calculated are listed for the different regions defined by the WHO.

Regions	$\boldsymbol{\chi_{p+}}$	$\boldsymbol{\chi_{p-}}$	$\boldsymbol{\eta}$
AFRH	0.51	0.43	0.843 (0.664-1.022)
EMR	0.45	0.53	1.178 (0.928-1.428)
SEAR	0.64	0.51	0.797 (0.628-0.966)
WPR	0.78	0.50	0.641 (0.505-0.777)

Table S9: Values of χ_{p+} and χ_{p-} considered in Abu-Raddad et al.[4] and the values of η for each region. In this work we have studied 5 countries from the AFRH region (Nigeria, South Africa, Democratic Republic of the Congo, Ethiopia and Tanzania), 1 from EMR (Pakistan), 4 from SEAR (India, Indonesia, Bangladesh and Myanmar) and 2 from WPR (China and Philippines).

In this work we have studied 5 countries from the AFRH region (Nigeria, South Africa, Democratic Republic of the Congo, Ethiopia and Tanzania), 1 from EMR (Pakistan), 4 from SEAR (India, Indonesia, Bangladesh and Myanmar) and 2 from WPR (China and Philippines).

The diagnosis rate is allowed to vary in time, as it has been done in other previous models (see section 2.8 for details).

2.1.5 Treatment outcomes

Right after diagnosis, and supposing that antibiotic treatments are available immediately, sick individuals start their treatment. In terms of our model, individuals under current treatment lie into T_{np}, T_{p-} or T_{p+}, depending on the type of disease they receive treatment to be cured from. During their stage at T classes, either by the effect of treatment or by the common isolation measures that use to follow a TB diagnosis, individuals are not considered to be able to spread the disease.

Typical antibiotic series last six months; let Ψ be the rate associated to the inverse of that treatment time. Once the treatment is completed, different results are possible, and the WHO classifies these treatment outcomes into four main groups:

- Success: the treatment has been completed and bacilli are not present in the sputum.
- Default: the treatment has been abandoned before completion.
- Death.
- Failure: bacilli persist -or appear- in the sputum at the end of the treatment (month five or later).

Therefore, let us denote as f_S^{p+}, f_D^{p+}, f_F^{p+} and f_μ^{p+}, the fraction of pulmonary, smear positive TB sick individuals who finish their treatments belonging respectively to success, default, failure and death groups, as they are available in the WHO database.[6] We will have the closure relationship $f_S^{p+} + f_D^{p+} + f_F^{p+} + f_\mu^{p+} = 1$ that allows us to substitute $f_S^{p+} = 1 - (f_D^{p+} + f_F^{p+} + f_\mu^{p+})$ so as to work just with these three fractions of unsuccessful treatment outcomes. For pulmonary smear negative and non pulmonary TB cases, the WHO database does not differentiate the fractions of treatment outcomes,[6] and so we have f_S^{p-}, f_D^{p-}, f_F^{p-} and f_μ^{p-} standing for the fraction of individuals undertaking each outcome both from pulmonary smear negative and from non pulmonary classes of TB. Again, we have the closure relationship $f_S^{p-} + f_D^{p-} + f_F^{p-} + f_\mu^{p-} = 1$ that yields the substitution $f_S^{p-} = 1 - (f_D^{p-} + f_F^{p-} + f_\mu^{p})$. The values of the fractions of non successful outcomes have been averaged during the fitting time window and their values are provided in table S15, where confidence intervals correspond to two typical deviations of a multinomial distribution.

Therefore, we can enumerate all the possible treatment outcomes from all the different kinds of patients to get:

- Early treatment abandon (default) of smear positive TB: transition from $T_{p+}(a,t)$ to $R_{p+D}(a,t)$: $\Psi f_D^{p+} T_{p+}(a,t)$ individuals/unit time.
- Failed treatment completion of smear positive TB: transition from $T_{p+}(a,t)$ to $F(a,t)$: $\Psi f_F^{p+} T_{p+}(a,t)$ individuals/unit time.
- Death during treatment of smear positive TB: $\Psi f_\mu^{p+} T_{p+}(a,t)$ individuals/unit time.
- Successful treatment completion of smear positive TB: transition from $T_{p+}(a,t)$ to $R_{p+S}(a,t)$: $\Psi(1 - f_D^{p+} - f_F^{p+} - f_\mu^{p+}) T_{p+}(a,t)$ individuals/unit time.
- Early treatment abandon (default) of smear negative TB: transition from $T_{p-}(a,t)$ to $R_{p-D}(a,t)$: $\Psi f_D^{p-} T_{p-}(a,t)$ individuals/unit time.
- Failed treatment completion of smear negative TB: transition from $T_{p-}(a,t)$ to $F(a,t)$: $\Psi f_F^{p-} T_{p-}(a,t)$ individuals/unit time.
- Death during treatment of smear negative TB: $\Psi f_\mu^{p-} T_{p-}(a,t)$ individuals/unit time.
- Successful treatment completion of smear negative TB: transition from $T_{p-}(a,t)$ to $R_{p-S}(a,t)$: $\Psi(1 - f_D^{p-} - f_F^{p-} - f_\mu^{p-}) T_{p-}(a,t)$ individuals/unit time.
- Early treatment abandon (default) of non pulmonary TB: transition from $T_{np}(a,t)$ to $R_{npD}(a,t)$: $\Psi f_D^{p-} T_{np}(a,t)$ individuals/unit time.
- Failed treatment completion of non pulmonary TB: transition from $T_{np}(a,t)$ to $F(a,t)$: $\Psi f_F^{p-} T_{np}(a,t)$ individuals/unit time.
- Death during treatment of non pulmonary TB: $\Psi f_\mu^{p-} T_{np}(a,t)$ individuals/unit time.
- Successful treatment completion of non pulmonary TB: transition from $T_{np}(a,t)$ to $R_{npS}(a,t)$: $\Psi(1 - f_D^{p-} - f_F^{p-} - f_\mu^{p-}) T_{np}(a,t)$ individuals/unit time.

where the different R_{xy} variables stand for the groups of individuals that have completed their treatment for disease of type x (pulmonary smear positive $p+$ or negative, $p-$, or non-pulmonary np) with an outcome denoted by y (Success, S, default D and fail F). We have also used the subindex μ when naming the fraction of deaths that occurs during treatment, but this outcome does not have a recovery class associated –these individuals die and leave the system–.

2.1.6 Natural recovery

In certain occasions, natural recovery from TB is possible without medical intervention or treatment.[4] This is modeled by introducing three new classes of naturally recovered individuals in the first branch: $R_{npN}(a,t)$, $R_{p-N}(a,t)$ and $R_{p+N}(a,t)$. Undiagnosed and sick individuals of each type of TB join these new classes after natural recovery as follows:

- Natural recovery of non pulmonary TB: transition from $D_{np}(a,t)$ to $R_{npN}(a,t)$: $\nu D_{np}(a,t)$ individuals/unit time.
- Natural recovery of smear negative pulmonary TB: transition from $D_{p-}(a,t)$ to $R_{p-N}(a,t)$: $\nu D_{p-}(a,t)$ individuals/unit time.
- Natural recovery of smear positive pulmonary TB: transition from $D_{p+}(a,t)$ to $R_{p+N}(a,t)$: $\nu D_{p+}(a,t)$ individuals/unit time.

where ν is the rate of natural recovery.

2.1.7 Endogenous reactivations after treatment or natural recovery

Nonetheless, naturally recovered individuals may experience an endogenous reactivation of the disease, since generally speaking disease recovery does not suppose the total elimination of the bacilli from the host organism. [14] If we denote by r_N the endogenous relapse rate of naturally recovered individuals we have:

- Endogenous reactivation of non pulmonary TB after natural recovery: transition from $R_{npN}(a,t)$ to $D_{np}(a,t)$: $r_N R_{npN}(a,t)$ individuals/unit time.
- Endogenous reactivation of smear negative TB after natural recovery: transition from $R_{p-N}(a,t)$ to $D_{p-}(a,t)$: $r_N R_{p-N}(a,t)$ individuals/unit time.
- Endogenous reactivation of smear positive TB after natural recovery: transition from $R_{p+N}(a,t)$ to $D_{p+}(a,t)$: $r_N R_{p+N}(a,t)$ individuals/unit time.

Furthermore, endogenous relapse is also possible after antibiotic treatment. Once the treatment has finished the probabilities of experiencing an endogenous reactivation of the disease are related to the treatment outcome of the initial disease episode.

Individuals who have experienced a failed treatment –$F(a,t)$ class–, regardless of the type of TB that they originally had, are considered as infectious as smear positive untreated individuals (because they present bacilli in the sputum at the end of the treatment) and their mortality risk due to TB is also the same of a smear positive untreated individual.

Within our modeling framework, ulterior re-diagnosis, re-infections or re-treatments for $F(a,t)$ individuals are not considered, and so, once an individual joins this class, her dynamics does not depend any more on the type of disease she previously had. However, the fact that these individuals die at a high rate (μ_{p+}) prevents this compartment of highly infectious individuals to become a dead-end in the disease dynamics and becomes a hidden driver of our results: in our simulations, the weight of this class among the totality of infectious individuals never surpasses 10%.

Recovered individuals after successful completion of treatment are considered functionally cured –i.e. they neither present a specific mortality risk due to TB nor they are infectious. However, they may undergo ulterior endogenous reactivations of the disease, caused by the proliferation of the same bacilli of the original episode, if these were not completely eliminated from the host organism. In that case, we have the following transitions:

- Endogenous reactivation of non pulmonary TB after successful treatment: transition from $R_{npS}(a,t)$ to $D_{np}(a,t)$: $r_S R_{npS}(a,t)$ individuals/unit time.
- Endogenous reactivation of smear negative TB after successful treatment: transition from $R_{p-S}(a,t)$ to $D_{p-}(a,t)$: $r_S R_{p-S}(a,t)$ individuals/unit time.
- Endogenous reactivation of smear positive TB after successful treatment: transition from $R_{p+S}(a,t)$ to $D_{p+}(a,t)$: $r_S R_{p+S}(a,t)$ individuals/unit time.

where r_S is the endogenous relapse rate after successful treatment completion. In what regards its estimation, there exist many epidemiological studies based on the surveillance of cohorts of TB patients after treatment completion during defined follow-up periods, which are aimed at determining the relapse rates, as well as the main risk factors associated to its increment.

In the exhaustive meta-analysis by Korenromp and colleagues, [14] an ensemble of such studies is considered. In that work, it is reported that, in all the works re-analized, an average of 4.2% ($3.1-5.3$ c.i.) of HIV uninfected subjects have a TB relapse episode during the follow-up period of the study, of which, 77% ($63-91$ C.I.) is due to endogenous reactivation. This means that the fraction of population that do not develop a relapse is the 96.77% ($95.73-97.80$).

Another relevant result of the meta-analysis is the finding that the risk for TB relapse after treatment decreases with time. This can be seen from the fact that the relapse rates calculated in the different studies considered tend to be lower as the follow-up period of the trials is higher. This would imply that most patients that experiment a relapse after treatment, do it within the first years after the initial episode.

This second result motivates the assumption that the risk of developing a relapse during the follow-up period of an epidemic surveillance study ($100 - 96.77 = 3.23\%$ of the population) can be associated to the total risk of developing such relapse during the entire life of an individual. Hence, our task is to calculate an annual risk of relapse such that, when applied over the whole period of life expectancy of a recovered individual, it yields the same 3.23% of relapse cases. To this end, we estimate that the average life expectancy of individuals within classes R is equal to 35 years, estimation that follows from assuming, as a first order approximation, that infection and further recovery are events that occur uniformly in all ages.

Therefore, and since we are assuming that the relapse rate is constant in age and time, we have an exponential decay describing the relapse of R_{xS} individuals (R_{p+S}, R_{p-S} or R_{npS}) of the form: $R_{xS}(t) \sim e^{-r_S t}$. Thus, after a period $t = 35$ years assimilable to the average life expectancy of an individual that has already entered into class R, from an initial fraction of $R_{xS} = 1$, there remains $R_{xS}(t = 35) \sim e^{-r_S 35} = 0.9677$. This calculation yields the actual value of r_S used in this work, $r_S = 9.4 \cdot 10^{-4}$ 1/year ($6.4 \cdot 10^{-4} - 1.3 \cdot 10^{-3}$). The confidence interval of r_S has been obtained after the propagation of the fraction of not relapsing population as the main source of uncertainty.

Finally, recovered individuals after treatment default are considered partially infectious, although it is assumed that they do not have an explicit mortality risk due to TB. However, their endogenous relapse risk is higher, which can be modeled by introducing a parameter $r_D > r_S$ as follows:

- Endogenous reactivation of non pulmonary TB after treatment default: transition from $R_{npD}(a,t)$ to $D_{np}(a,t)$: $r_D R_{npD}(a,t)$ individuals/unit time.
- Endogenous reactivation of smear negative TB after treatment default: transition from $R_{p-D}(a,t)$ to $D_{p-}(a,t)$: $r_D R_{p-D}(a,t)$ individuals/unit time.
- Endogenous reactivation of smear positive TB after treatment default: transition from $R_{p+D}(a,t)$ to $D_{p+}(a,t)$: $r_D R_{p+D}(a,t)$ individuals/unit time.

r_D stands for the endogenous relapse rate after treatment default, which has been calculated as the product of r_S and the relative risk factor for endogenous relapse related to treatment non-compliance, 4.02 (1.79-9.01 c.i.), taken from Picon et al.[15], which yields the final value of $r_D = 3.8 \cdot 10^{-3}$ 1/year ($1.4 \cdot 10^{-3} - 8.6 \cdot 10^{-3}$).

2.1.8 Exogenous reinfection of infected individuals

Individuals belonging to classes L_s and R have been previously exposed to TB bacilli, although they are not sick while remaining within those classes. In addition, their rates of progression to disease due to eventual endogenous reactivations are slower than the rate ω_f of fast progression to disease from L_f. For these reasons, an eventual exogenous re-infection of an individual in classes L_s or R may cause a faster transition to disease, if fast progression takes place, than endogenous reactivation. This can be modeled by introducing the following transitions:

- Exogenous re-infection of $L_s(a,t)$ individuals yielding fast progression: from $L_s(a,t)$ to $L_f(a,t)$: $p(a)q\lambda(a,t)L_s(a,t)$ individuals/unit time.
- Exogenous re-infection of $R_{npN}(a,t)$ individuals yielding fast progression: from $R_{npN}(a,t)$ to $L_f(a,t)$: $p(a)q\lambda(a,t)R_{npN}(a,t)$ individuals/unit time.
- Exogenous re-infection of $R_{p-N}(a,t)$ individuals yielding fast progression: from $R_{p-N}(a,t)$ to $L_f(a,t)$: $p(a)q\lambda(a,t)R_{p-N}(a,t)$ individuals/unit time.
- Exogenous re-infection of $R_{p+N}(a,t)$ individuals yielding fast progression: from $R_{p+N}(a,t)$ to $L_f(a,t)$: $p(a)q\lambda(a,t)R_{p+N}(a,t)$ individuals/unit time.
- Exogenous re-infection of $R_{npS}(a,t)$ individuals yielding fast progression: from $R_{npS}(a,t)$ to $L_f(a,t)$: $p(a)q\lambda(a,t)R_{npS}(a,t)$ individuals/unit time.
- Exogenous re-infection of $R_{p-S}(a,t)$ individuals yielding fast progression: from $R_{p-S}(a,t)$ to $L_f(a,t)$: $p(a)q\lambda(a,t)R_{p-S}(a,t)$ individuals/unit time.

- Exogenous re-infection of $R_{p+S}(a,t)$ individuals yielding fast progression: from $R_{p+S}(a,t)$ to $L_f(a,t)$: $p(a)q\lambda(a,t)R_{p+S}(a,t)$ individuals/unit time.
- Exogenous re-infection of $R_{npD}(a,t)$ individuals yielding fast progression: from $R_{npD}(a,t)$ to $L_f(a,t)$: $p(a)q\lambda(a,t)R_{npD}(a,t)$ individuals/unit time.
- Exogenous re-infection of $R_{p-D}(a,t)$ individuals yielding fast progression: from $R_{p-D}(a,t)$ to $L_f(a,t)$: $p(a)q\lambda(a,t)R_{p-D}(a,t)$ individuals/unit time.
- Exogenous re-infection of $R_{p+D}(a,t)$ individuals yielding fast progression: from $R_{p+D}(a,t)$ to $L_f(a,t)$: $p(a)q\lambda(a,t)R_{p+D}(a,t)$ individuals/unit time.

where q stands for the variation factor of the infection risk of individuals who have been infected in a previous episode.

On the other hand, if fast progression after the secondary infection does not take place, even if the initial state of the individual is one of the possible R states, the rule is that no transition must be considered from these states to L_s, because an endogenous reactivation from those initial states to disease is always more likely than from L_s, as either r_N, r_S or r_D are greater than the rate of transition from slow latency to disease ω_s. However, we will count these potential transitions when accounting for the total number of infections that occurred in our system.

In summary, re-infection has no effect on our model unless it is followed by fast progression. Nevertheless, we note that the nature of TB infection could be more complicated regarding the influence of repeated exposure: For instance, in Lee et al.[16] it is described how the progression to TB disease increases with the number of exposures.

2.1.9 Smear progression

In certain cases, it is documented that patients of smear negative pulmonary TB progress to smear positive ,[3] even after being treated. In order to describe this phenomenon, we introduce the smear progression by considering the following two transitions:

- Smear progression of untreated individuals: transition from $D_{p-}(a,t)$ to $D_{p+}(a,t)$: $\theta D_{p-}(a,t)$ individuals/unit time.
- Smear progression of individuals under treatment: transition from $T_{p-}(a,t)$ to $T_{p+}(a,t)$: $\theta T_{p-}(a,t)$ individuals/unit time.

where θ stands for the smear progression rate.

2.1.10 Mother-child infection transmission

We model the possibility of mother-child transmission right after birth (peri-natal infection). While most of the newborns that enter the system at each time step $\Delta_N(a=0,t)$ (see section 2.5 to see how we estimate it) will enter the system as susceptiible, a fraction of them will do it directly to the latency classes. This reflects the known fact that a fraction m_c of sick women who are pregnant transmits the disease to the children within the first weeks of their lives[17]. The density of infected newborns depends then on the fraction of mothers who have the disease and are able to transmit it at time step t, which leads to the question of what is the relative risk of transmitting the pathogen to the offspring for women in each of the infectious classes included in our model. In this work we considered that the total number of newborn infections is proportional to $m_d(t)$:

$$m_d(t) = \frac{\sum_{a=3}^{a=7} D_{p+}(a,t) + D_{p-}(a,t) + D_{np}(a,t) + R_{p+D}(a,t) + R_{p-D}(a,t) + R_{npD}(a,t) + F(a,t)}{\sum_{a=3}^{a=7} N(a,t)} \tag{3}$$

that represents the fraction of infected individuals present in the age groups $a \in [3,7]$ associated to women fertility (between 15 and 40 years old), regardless of their relative infectiousnesses. Therefore, this yields following the distribution of the $\Delta_N(a=0,t)$ newborns among S and L classes:

- Birth of $S(0,t)$ individuals (susceptible newborns): $(1-m_c m_d(t))\Delta_N(a=0,t)$
- Birth of $L_f(0,t)$ individuals (infected after birth who develops fast progression): $m_c m_d(t) p(0)\Delta_N(a=0,t)$
- Birth of $L_s(0,t)$ individuals (infected after birth who develops slow progression): $m_c m_d(t)(1-p(0))\Delta_N(a=0,t)$

2.2 Force of infection

The force of infection $\lambda(a,t)$ is, as it has been said before in section 2.1.1, the rate at which infection occurs at time step t for a susceptible individual in age-group a. This magnitude is calculated according to the following expression:

$$\lambda(a,t) = \beta(t) \sum_{a'} \xi_c(a,a',t) \Upsilon(a',t) \tag{4}$$

being $\Upsilon(a',t)$ the weighted density of all the infectious individuals within age-group a' at time step t:

$$\Upsilon(a',t) = \frac{1}{N(a',t)} \left(D_{p+}(a',t) + F(a',t) + \phi_{p-} D_{p-}(a',t) + \phi_D R_{p+D}(a',t) + \phi_{p-}\phi_D R_{p-D} \right) \tag{5}$$

where $N(a,t)$ is the total population of age a, $\phi_{p-} \in [0,1]$ is the coefficient of infectiousness reduction of smear negative sick individuals with respect to smear positive ones, and $\phi_D \in [0,1]$ the infectiousness reduction of individuals who have defaulted the treatment, (R_{p+D} individuals with respect to smear positive, undiagnosed individuals D_{p+}). Diagnosed patients of smear negative TB who failed their treatment, have an infectiousness reduction that is the product of the two terms $\phi_{p-}\phi_D$.

On the other hand, $\xi_c(a,a',t)$ represents the relative contact frequency that an individual of age a has with individuals of age a' at time t, with respect to the overall average of contacts that an individual has per unit time with anyone else, which implies that the following average:

$$< k_c >= \frac{\sum_{a,a'} \xi_c(a,a',t) N_c(a,t)}{N_c(t)} \tag{6}$$

is equal to 1 in any time t and country c (see section 2.3 to see how $\xi_c(a,a',t)$ is built to fulfill this property). Considering this normalization property of the contact matrix $\xi_c(a,a',t)$, the scaled infectiousness $\beta(t)$ provides a global scale for the overall frequency of contacts per unit time that are actually occurring in the system, as well as for how likely an infection is to occur upon one of those contacts. Formally, this can be interpreted considering that β is the product of two indistinguishable nuisance parameters: the average connectivity of the network of contacts (i.e. the average number of epidemiologically relevant interactions that any individual has per unit time) and the "intrinsic" infectiousness, (i.e. the probability of a contagion to occur upon one of those contacts, if established between an infectious and a susceptible individual). Next, in section 2.3, we explain how we obtain the contact matrix $\xi_c(a,a',t)$.

2.3 Contact patterns

It has been previously shown that abandoning the hypothesis of homogeneous age-mixing in favor of data-driven approaches based on empiric data[5,18] improves the descriptive capabilities of epidemiological models of influenza-like diseases.[19,20] Similarly, in a study by Guzzetta and collaborators,[21] the importance of considering heterogeneous contacts in the modeling of TB was assessed in the context of Individual Based Modeling (IBM). Despite these first attempts, in mathematical TB modeling at the level of broad populations (i.e. countries or international macro-regions) the assumption of homogeneous mixing across age-strata still constitutes one of the most pervasive simplifying hypothesis.

In this work, we abandon this hypothesis and make use of country-wise contact matrices among different age groups, denoted as $\xi_c(a,a',t)$. The matrix elements $\xi_c(a,a',t)$ represent the relative contact frequency that an individual of age a has with individuals of age a' at time t, with respect to the overall average of contacts that an individual has per unit time with anyone else. For the computation of the contact matrices used in our model, we have collected different survey studies from several countries: Kenya[22], Zimbabwe[23], Uganda[24], China[25], Japan[26] and Europe (8 countries)[5]. With the first three studies we build a contact matrix that will be adapted to be used in African countries (Nigeria, South Africa, Democratic Republic of the Congo, Ethiopia and Tanzania), while the next two are used to build an Asian Contact Matrix (to be adapted to India, Indonesia, China, Pakistan, Bangladesh, Philippines and Myanmar). Finally, the surveys performed in the Polymod project, that corresponds to 8 European countries will be aggregated into an European Contact Matrix that we will use to check the robustness of our results against different contact structures (used only in the SI Appendix, sections 1.5 and 1.7).

The temporal dependence of the contact matrices $\xi_c(a,a',t)$, which will be explained in detail, comes from considering that contact structures among age-groups are conditioned not only by cultural and socio-economic differences between countries, but also by the underlying demographies of the populations under analysis. As a

consequence, we hypothesize that, if the demography of a population changes over time, the contact structure among different age-strata will change too. In this section, we explain in detail the approach that we follow to formalize our description of contact patterns within our model, around the aforementioned premises.

As a first pre-processing step, contact matrices reported in each study are adapted to the age-structure of 15 age groups used in this article. After this process we obtain a contact matrix for each study, that we will call $\xi_s(a, a')$, where s is an index for the particular study.

The matrices $\xi_s(a, a')$ originally represent the number of contacts per unit time that an individual in age-group a reports, in average, towards individuals in age group a', during a survey. This means that $\xi_s(a, a')$ is not expected to be symmetrical due to the different number of people in each age group: they would be expected to be symmetrical only for a perfectly rectangular demographic pyramid where each age group has exactly the same population. However, they should ideally fulfill the following relationship:

$$\xi_s(a, a')N_s(a) \simeq \xi_s(a', a)N_s(a') \tag{7}$$

where $N_s(a)$ is the population of age a in the setting s (i.e. the specific country at the moment the survey took place). This means that the number of contacts per unit time between a and a', inferred from the reports from both age groups in the survey should not differ substantially. Obviously, these two quantities are not expected to be exactly the same, since the reports from both age groups cannot be, in general, expected to concur exactly. Thus, the next step in the computation of the contact matrices consists of correcting $\xi_s(a, a')$ in order to recover the mentioned symmetry exactly:

$$\xi_s^{\text{sim.}}(a, a') = \frac{1}{N_s(a)} \frac{\xi_c(a, a')N_s(a)n_s(a) + \xi_s(a', a)N_s(a')n_s(a')}{n_s(a) + n_s(a')} \tag{8}$$

where $n_s(a)$ is the number of participants of age a recruited for the study s, and $N_s(a)$ the number of individuals in that age group in the entire population. In other words, we estimate the total number of contacts between age-groups a and a' as an average of the number of contacts that emanates from the reports of both age groups, weighted according to the number of participants in each group; and correct the entries of the matrix $\xi_s^{\text{sim.}}(a, a')$ and $\xi_s^{\text{sim.}}(a', a)$ so they agree around that number of total contacts.

After this process, we have a set of matrices $\xi_s^{\text{sim.}}(a, a')$ for each study that are compatible with the symmetry $\xi_s^{\text{sim.}}(a, a')N_s(a) = \xi_s^{\text{sim.}}(a', a)N_s(a')$, which describe the number of contacts that an individual of age a has with people of age a' during a certain unit of time defined in the particular study. However, those matrices are difficult to compare across studies, since the definition of "contact" being used in each of them differs. To deal with this limitation, we normalize these matrices, so the mean degree of the matrix, defined as the average number of contacts per unit time of an individual, regardless of her age or that of her contacts, equals 1 (and so, they reflect the relative intensity of contacts between age groups a and a' rather than their absolute frequency):

$$\xi_s^{\text{sim.,norm.}}(a, a') = \frac{\xi_s^{\text{sim}}(a, a')}{< k_s^{\text{sim}} >} \tag{9}$$

where $< k_s^{\text{sim}} >$ is the mean degree of $\xi_s^{\text{sim.}}(a, a')$:

$$< k_s^{\text{sim}} >= \frac{\sum_{a,a'} \xi_s^{\text{sim}}(a, a')N_s(a)}{\sum_a N_s(a)} \tag{10}$$

Once we have the normalized and "symmetry-compatible" matrices for each study, we perform a weighted sum per continent (using number of participants in each study as weights) to obtain the correspondent regional matrices, that we will call $\xi_{reg}(a, a')$ (for Africa, Asia and Europe, respectively, see figure S13A). We then use the dispersion between studies to define the uncertainty associated to the contact patterns (which we will be propagating to the final matrices and finally, to model outcomes).

The set of continent-wise matrices $\xi_{reg}(a, a')$ only fulfill the symmetry of contacts in the setting of reference where they have been obtained, defined as the union of the countries being averaged in each case. In order to adapt these matrices to each of the 12 countries where we apply our model, we need to adapt them to their particular demographic pyramids and to their temporal evolution trends.

To obtain this correction, we interpret the matrices $\xi_{reg}(a, a')$ as the product of two nuisance factors: the fraction of individuals in a' that exist in the population: $\frac{N_{reg}(a')}{N_{reg}}$, and an auxiliary matrix $\pi_{reg}(a, a')$. Under this view, the matrix $\pi_{reg}(a, a')$, denoted as intrinsic connectivity matrix (figure S13B), represents the contact structure of each region (Africa, Asia and Europe), once the effect of the demographic structures of the places where the initial studies were performed has been removed. $\pi_{reg}(a, a')$ is obtained as follows:

$$\pi_{reg}(a,a') = \xi_{reg}(a,a')\frac{N_{reg}}{N_{reg}(a')} \tag{11}$$

where $N_{reg} = \sum_a N_{reg}(a)$. Finally, these intrinsic matrices $\pi_{reg}(a,a')$ from which the influence of the populations where the studies were performed have been removed are transformed according to the demography to be described:

$$\tilde{\xi}_c(a,a',t) = \pi_{reg}(a,a')\frac{N_c(a',t)}{N_c(t)} \tag{12}$$

which would constitute a first approximation to the evolution of the contact structure of any country c at any time t. However, one of the features of this matrix is that, by construction, and depending on the time evolution of the demographic structure being considered, as defined by $N_c(a',t)$ in each age-group, it can present, in general, an average contact intensity that changes across time as a consequence of the demographic evolution. In this sense, although it is debatable whether the average connectedness of a population as a whole should depend or not on the demographic changes that it is experiencing, we have decided to rule out this possibility by normalizing these matrices at each time-step (figure S13 C). By doing so, we ensure that the matrices have an average connectivity equal to 1 at every moment, meaning that their entry (a,a') always represent the relative frequency at which individuals of age-group a contact those in age-group a', compared to the overall contact frequency of any individual of the system with anyone else:

$$\xi_c(a,a',t) = \frac{\tilde{\xi}_c(a,a',t)}{<\tilde{k}_c(t)>} = \frac{\pi_{reg}(a,a')N_c(a',t)N_c(t)}{\sum_{a,a'}\pi_{reg}(a,a')N_c(a',t)N_c(a,t)} \tag{13}$$

The reasons supporting this final choice are two-fold.

On the one hand, the main objective that we pursue in this work regarding contact patterns is to quantify the influence that they exert on model forecasts, by comparing the outcomes of our model to a case where contacts are considered to be homogeneous. Since, in the latter case, the mean contact intensity is trivially constant over time, to use heterogeneous contact matrices that share this same property constitutes a conservative choice that makes the comparison between the full and the reduced model easier to interpret.

On the other hand, this procedure implies that the average contact intensity in both cases is not just time-invariant, but equal to 1 in both cases. This makes the scaled infectiousness parameter $\beta(t)$ to be comparable between the full and the reduced model, thus providing an overall scale for the average capability of an infected individual to propagate the disease in both cases.

Under this formulation, the relative ratio between the contact intensities of a group a towards two different groups a' and a'' is represented by the fraction $\xi_c(a,a',t)/\xi_c(a,a'',t)$, which measures how likely are individuals of age a to interact with people of age a' in comparison to age a''. The temporal evolution of this ratio only depends on the relative volumes of the target age groups as time goes by, as specified in the following equation:

$$\frac{\xi_c(a,a',t_1)}{\xi_c(a,a'',t_1)} \Big/ \frac{\xi_c(a,a',t_0)}{\xi_c(a,a'',t_0)} = \frac{N(a',t_1)}{N(a'',t_1)} \Big/ \frac{N(a',t_0)}{N(a'',t_0)} \tag{14}$$

In the figure S13, we summarize this process, showing the contact patterns obtained at each step, normalized to a common scale in each region. As it can be observed, the corrections explained above designed to capture the influence of the different demographic structures across countries and time introduce slight variations when compared to the differences observed between the three geographical areas.

Regarding the reduced model 2, the hypothesis, much simpler, consists of assuming that the probability for two individuals to interact is the same, which means that the contact frequency that an individual in group a has with people in group a' only depends on the frequency of the target group in the population:

$$\xi_c^{RM2}(a,a',t) = N(a',t)/N(t) \tag{15}$$

in such a way that the average number of contacts that an individual of any age has per unit time is always and everywhere equal to 1. Thus, the general contact intensity is modulated by $\beta(t)$ in the same way as in the full model.

In the tables S10,S11 and S12, we specify the values of the $\pi_{reg}(a,a')$ matrices for the three macro-regions considered (Africa, Asia and Europe). Notice that these are not directly the contact patterns considered, but they have to be corrected by the demography of the individual setting (equations 12 and 13).

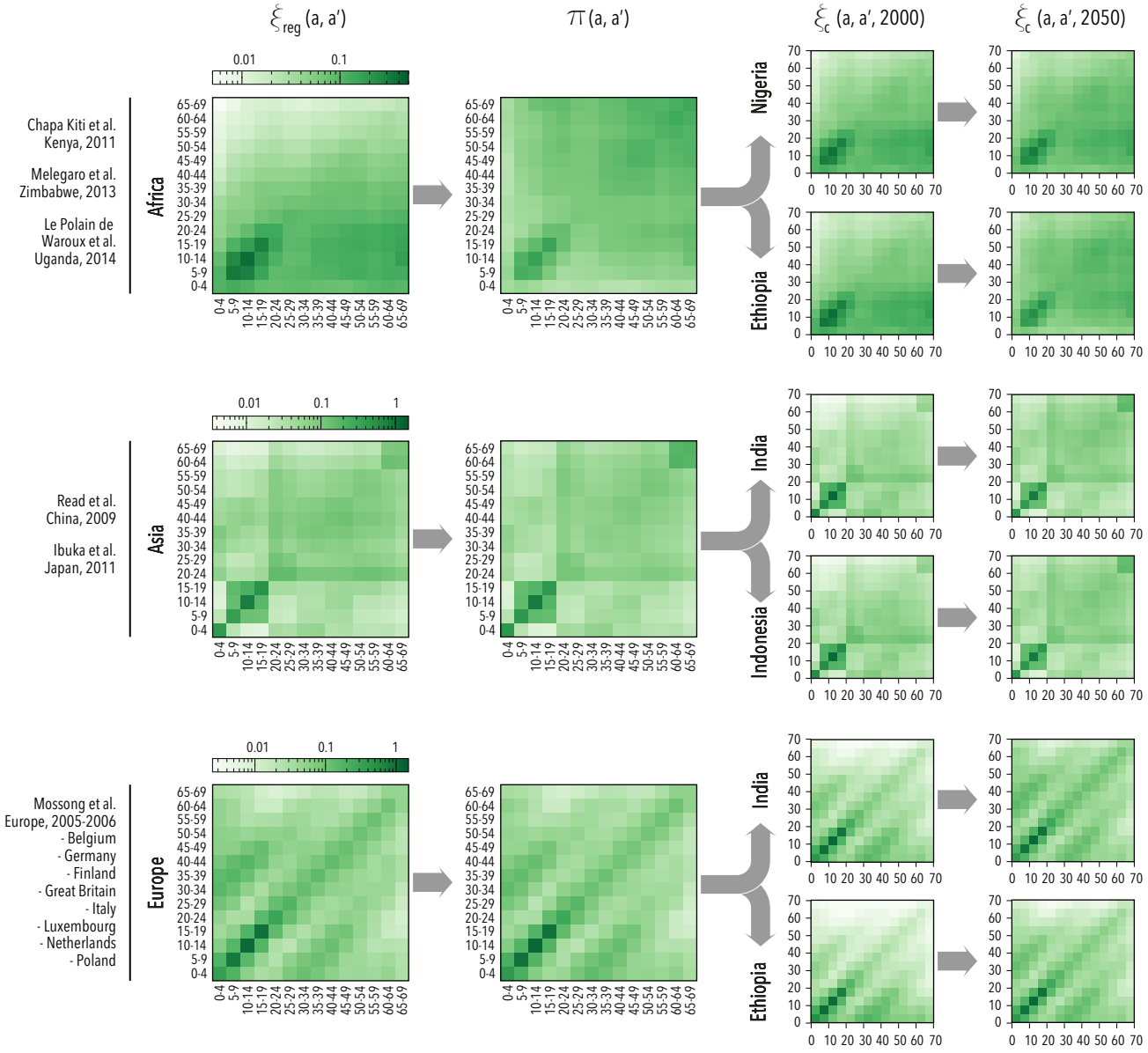

Figure S13: Empiric contact patterns in evolving demographies, as number of contacts that an individual in age-group a (x-axis) has with individuals in age group a' (y-axis) (A) Region-wise contact matrices $\xi_{reg}(a, a')$ derived from weighted averages from studies completed in each continent. (B) Intrinsic contact matrices $\pi_{reg}(a, a')$, capturing the different contact intensities among age-groups once the effects of demographic variability have been removed. (Normalized, in order to display $\langle k \rangle = 1$ in the figure to facilitate comparison to the rest of the matrices) (C): Country-wise, time dependent contact matrices used in this study for the four countries analyzed in the main text $\xi_c(a, a', t)$, obtained after adapting the intrinsic contact matrices $\pi_{reg}(a, a')$ to the specific demographic setup of each country at each time. European contacts are showed when applied to Ethiopia and India, to illustrate that our approach can be used to test the effect of alien contact structures obtained in any setting (these are only used in the SI Appendix, for the analyses presented in figures S6 and S10)

	0-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50	50-55	55-60	60-65	65-70	+70
0-5	1.25	1.50	1.33	1.01	0.93	1.02	1.04	0.93	0.92	0.90	1.03	1.15	1.22	1.15	0.82
5-10	1.50	4.47	4.01	2.08	1.24	1.28	1.33	1.45	1.50	1.67	1.71	1.54	1.72	1.77	1.37
10-15	1.33	4.01	5.94	3.18	1.77	1.33	1.42	1.79	1.83	1.95	1.97	1.77	1.97	2.72	2.26
15-20	1.01	2.08	3.18	5.53	2.84	1.74	1.77	2.20	2.23	2.41	2.45	2.32	2.49	2.90	2.40
20-25	0.93	1.24	1.77	2.84	2.89	1.95	1.88	2.16	2.24	2.53	2.74	2.58	2.83	3.09	2.47
25-30	1.02	1.28	1.33	1.74	1.95	2.31	2.16	2.16	2.27	2.27	2.38	2.20	2.44	2.59	1.85
30-35	1.04	1.33	1.42	1.77	1.88	2.16	2.13	2.25	2.35	2.28	2.35	2.18	2.48	2.48	1.65
35-40	0.93	1.45	1.79	2.20	2.16	2.16	2.25	2.77	2.88	3.01	3.02	2.62	2.87	2.89	1.98
40-45	0.92	1.50	1.83	2.23	2.24	2.27	2.35	2.88	3.14	3.34	3.20	2.70	2.94	3.07	2.20
45-50	0.90	1.67	1.95	2.41	2.53	2.27	2.28	3.01	3.34	4.04	3.86	3.22	3.51	3.80	2.95
50-55	1.03	1.71	1.97	2.45	2.74	2.38	2.35	3.02	3.20	3.86	3.96	3.42	3.76	3.92	3.17
55-60	1.15	1.54	1.78	2.32	2.58	2.20	2.18	2.62	2.70	3.22	3.42	3.82	4.26	3.76	3.07
60-65	1.22	1.72	1.97	2.49	2.83	2.44	2.48	2.87	2.94	3.51	3.76	4.26	4.89	4.33	3.77
65-70	1.15	1.77	2.72	2.90	3.09	2.59	2.48	2.89	3.07	3.80	3.92	3.76	4.33	5.06	4.64
+70	0.82	1.37	2.26	2.40	2.47	1.85	1.65	1.98	2.20	2.95	3.17	3.07	3.77	4.64	3.55

Table S10: Matrix $\pi_{reg}(a, a')$ corresponding to Africa

	0-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50	50-55	55-60	60-65	65-70	+70
0-5	13.89	0.87	0.34	0.31	1.41	1.19	1.88	1.99	1.11	1.18	0.94	0.90	0.97	0.70	0.90
5-10	0.87	13.36	4.60	4.17	0.75	0.62	1.23	1.31	1.35	1.44	0.78	0.76	0.68	0.50	0.50
10-15	0.34	4.60	25.52	5.70	0.77	0.68	1.07	1.09	1.61	1.72	0.90	0.90	0.80	0.59	0.64
15-20	0.31	4.17	5.70	13.23	1.08	1.04	0.90	0.89	1.24	1.32	1.25	1.29	0.86	0.65	1.41
20-25	1.41	0.75	0.77	1.08	3.26	3.18	2.17	2.19	2.20	2.30	2.65	2.67	2.36	1.81	2.13
25-30	1.19	0.62	0.68	1.04	3.18	2.49	1.52	1.63	1.78	1.90	2.14	2.01	1.76	1.29	1.40
30-35	1.88	1.23	1.07	0.90	2.17	1.52	1.92	2.14	1.98	2.13	1.74	1.56	1.46	0.98	0.88
35-40	1.99	1.31	1.09	0.89	2.19	1.63	2.14	2.31	2.06	2.20	1.82	1.67	1.56	1.05	0.98
40-45	1.11	1.35	1.61	1.24	2.20	1.78	1.98	2.06	2.52	2.70	2.16	2.07	1.77	1.25	1.53
45-50	1.18	1.44	1.72	1.32	2.30	1.90	2.13	2.20	2.70	2.87	2.30	2.23	1.88	1.38	1.69
50-55	0.94	0.78	0.90	1.25	2.65	2.14	1.74	1.82	2.16	2.30	2.99	2.87	2.29	1.82	3.04
55-60	0.90	0.76	0.90	1.29	2.67	2.01	1.56	1.67	2.07	2.23	2.87	2.68	2.09	1.61	2.57
60-65	0.97	0.68	0.80	0.86	2.36	1.76	1.46	1.56	1.77	1.88	2.29	2.09	5.99	5.60	3.13
65-70	0.70	0.50	0.59	0.65	1.81	1.29	0.98	1.05	1.25	1.38	1.82	1.61	5.60	6.07	3.48
+70	0.90	0.50	0.64	1.41	2.13	1.40	0.88	0.98	1.53	1.69	3.04	2.57	3.13	3.48	6.16

Table S11: Matrix $\pi_{reg}(a, a')$ corresponding to Asia

	0-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50	50-55	55-60	60-65	65-70	+70
0-5	13.85	7.09	2.32	1.29	1.56	3.18	4.64	3.66	2.26	1.62	1.71	1.54	1.94	1.12	0.65
5-10	7.09	32.05	5.93	1.86	0.98	2.14	3.42	4.35	3.30	1.65	1.57	1.27	1.50	1.03	0.73
10-15	2.32	5.93	39.26	6.06	1.08	0.75	1.67	3.02	4.12	2.48	1.38	0.81	0.88	0.81	0.99
15-20	1.29	1.86	6.06	32.18	4.67	1.47	0.91	1.42	2.78	3.23	1.60	0.94	0.51	0.43	0.73
20-25	1.56	0.98	1.08	4.67	9.04	3.78	1.84	1.07	1.31	1.79	1.58	0.97	0.51	0.33	0.36
25-30	3.18	2.14	0.75	1.47	3.78	5.44	2.59	1.41	1.14	1.38	1.80	1.33	0.98	0.42	0.41
30-35	4.64	3.42	1.67	0.91	1.84	2.59	3.64	2.51	1.60	1.46	1.26	1.30	1.28	0.56	0.40
35-40	3.66	4.35	3.02	1.42	1.07	1.41	2.51	3.74	2.16	1.43	1.12	0.96	1.17	0.81	0.55
40-45	2.26	3.30	4.12	2.78	1.31	1.14	1.60	2.16	3.35	2.46	1.52	0.93	1.15	0.77	0.74
45-50	1.62	1.65	2.48	3.23	1.79	1.38	1.46	1.43	2.46	3.13	2.10	1.30	0.86	0.59	0.90
50-55	1.71	1.57	1.38	1.60	1.58	1.80	1.26	1.12	1.52	2.10	3.22	2.26	1.29	0.63	0.81
55-60	1.54	1.27	0.81	0.94	0.97	1.33	1.30	0.96	0.93	1.30	2.26	3.46	2.00	0.99	0.65
60-65	1.94	1.50	0.88	0.51	0.51	0.98	1.28	1.17	1.15	0.86	1.29	2.00	3.67	1.85	1.04
65-70	1.12	1.03	0.81	0.43	0.33	0.42	0.56	0.81	0.77	0.59	0.63	0.99	1.85	1.84	1.13
+70	0.65	0.73	0.99	0.73	0.36	0.41	0.40	0.55	0.74	0.90	0.81	0.65	1.04	1.13	1.20

Table S12: Matrix $\pi_{reg}(a, a')$ corresponding to Europe

2.4 Aging

The model considers 15 different age groups. Each of these groups comprises an age interval of $\Delta_t = 5$ years, except the last one that corresponds to individuals older that 70 years old. The relevance of such an age structured description of the system comes from the fact that some of the most relevant dynamical parameters take distinct values for each age group. To account for the aging of individuals as time goes by -and thus the

evolution of their age a- we introduce on the system of equations the following aging transitions, that stands for the promotion of individuals within whatever dynamical class of the model $X(a,t)$ to the next age class $X(a+1,t)$.

- Aging of individuals belonging to class $X(a,t)$, transition from $X(a,t)$ to $X(a+1,t)$: $X(a,t)/\Delta_t$ individuals/unit time.

Obviously, each class $X(a,t)$ receives people from $X(a-1,t)$ and sends out people to $X(a+1,t)$, except $X(0,t)$, that only receives newborns and $X(14,t)$, for which no further aging occurs.

2.5 Demographic evolution

Once all the transitions among the different dynamical states of the system have been described, as well as the aging fluxes, it is necessary to provide a global description of the evolution of the demographic structure, given by the evolution of the set of variables $N(a,t)$, which are defined as the total number of individuals within age group a in the population, no matter their states regarding TB dynamics:

$$\begin{aligned} N(a,t) = S(a,t) + L_f(a,t) + L_s(a,t) + D_{p+}(a,t) + D_{p-}(a,t) + D_{np}(a,t) + F(a,t) \\ +T_{p+}(a,t) + T_{p-}(a,t) + T_{np}(a,t) + R_{p+N}(a,t) + R_{p-N}(a,t) + R_{npN}(a,t) \\ +R_{p+S}(a,t) + R_{p-S}(a,t) + R_{npS}(a,t) + R_{p+D}(a,t) + R_{p-D}(a,t) + R_{npD}(a,t) \end{aligned} \tag{16}$$

where the evolution of $N(a,t)$ -in addition to aging and death by TB– is subject to other driving forces related to aspects like vegetative variation of population –births and non-TB deaths– as well as migration. In order to provide a description of the temporal evolution of the demographic structure, previous models have turned to different simplifying hypotheses to describe the system.

One of them consists of forcing the system to preserve, at any time, the total number of individuals $\mathcal{N}(t)$: $\mathcal{N}(t) = \sum_a N(a,t)$ by imposing that $\mathcal{N}(t) = \mathcal{N}(t=0)\forall(t)$.[4] A more sophisticated alternative, adopted in Dye et al.[3], is based on imposing that the system preserves the initial age structure of the population by making that, in each age group: $N(a,t) = N(a,t=t_o)\forall(t)$. Our approach, however, is based on assuming that the temporal evolution of the variables $N(a,t)$ follows the predictions of the United Nations Population Division, available at its on-line databases: $N(a,t) = N_{UN}(a,t)$.[27] From figure 2 in the main text and figure S1, we can see that population aging is a common feature in virtually all the countries under study.

In the following sub-sections, we detail these two different schemes (constant versus evolving demographies, implemented in the reduced model 1 and the full model, respectively), whose influence on model forecasts are analyzed in the paper.

Reduced model 1: Constant demographic structure

A first approach consists of imposing that the demographic structure of the population has to remain constant during the dynamical process: i.e. $N(a,t) = N(a,t_o)\forall t$. As mentioned earlier, this is what is done in some previous works,[3] where the dynamical states indicate densities rather than numbers of individuals. In this case, the force of infection is calculated as an average of the densities of sick individuals in each age group, weighted by the number of individuals within each age class of the demographic structure. In this way, Dye et al.[3] provide a means for calculating infection and mortality rates that takes into account the initial demographic structure of the population, and these rates can be eventually transformed into numbers by using data about the evolution of the total population under consideration.

In order to provide an equivalent description in the context of our model –where states represent number of individuals rather than densities–, we start by calculating the variation of population due to TB and aging in each age group:

$$\begin{aligned} \dot{N}_o(a,t) = ((1-\delta(a))N(a-1,t) - N(a,t))/\Delta_t \\ -\mu_{p+}(D_{p+}(a,t) + F(a,t)) - \mu_{p-}D_{p-}(a,t) - \mu_{np}D_{np}(a,t) \\ -\Psi f_\mu^{p+} T_{p+}(a,t) - \Psi f_\mu^{p-}(T_{p-}(a,t) + T_{np}(a,t)) \end{aligned} \tag{17}$$

being $\delta(a)$ the Dirac delta function. In order to preserve the number of individuals within each age group at any moment, we simply introduce a term $\Delta_N(a,t)$ that is intended to balance $\dot{N}_o(a,t)$ within each age group: $\Delta_N(a,t) = -\dot{N}_o(a,t)$, yielding:

$$\dot{N}(a,t) = \dot{N}_o(a,t) + \Delta_N(a,t) = 0 \ \forall(a,t) \tag{18}$$

The key question is how to distribute these correction terms $\Delta_N(a,t)$ between the different dynamical states $X(a,t)$ within age-class a. These increments have to be distributed between $X(a,t)$ dynamical states keeping the relative volume of these states within the age group so as not to introduce external, undesired biases on states' densities. If we call $\Delta_X(a,t)$ the fraction of $\Delta_N(a,t)$ that is introduced in state $X(a,t)$, we have:

$$\Delta_X(a,t) = \Delta_N(a,t)\frac{X(a,t)}{N(a,t)} \tag{19}$$

and obviously:

$$\Delta_N(a,t) = \sum_X \Delta_X(a,t) \tag{20}$$

This scheme has the advantage, with respect to consider (as in Abu-Raddad et al.[4]), that $\mathcal{N}(t) = \mathcal{N}(t = 0)\forall(t)$, that, at least, the structure of the population is controlled. However, as in that case, population growth is not explicitly considered, and further information about population volume is required so as to scale rates into numbers, as done in Dye et al.[3] Additionally, the main problem with this approach comes from the fact that no variation of the age structure of the population can be considered by proceeding this way, which might introduce significant biases from current demographic forecasts, specially when studying populations subjected to strong processes of demographic aging.

Full model: Evolving demography according to an external constraint

Starting from the last scheme for modeling the demographic evolution, it is easy to obtain a final approach that explicitly considers not only the influence of the age structure into the spreading, but also the population growth and the variation of the age structure itself. To this end, it is necessary to know the actual –or projected– evolution of the demographic structure of the population during the period under analysis. In our case, we are modeling TB dynamics from 2000 to 2050, and the official annual projections for the population per age group of any country are available, up to 2100, at the UN population division database.[27] Thus, from the UN database we obtain the actual annual population series expected by the UN for the populations at each age group, which can be trivially fitted to a continuous function $N_{UN}(a,t)$, from which we can derive an analytical derivative $\dot{N}_{UN}(a,t)$ at any moment. For the purpose of this work, a polynomial of degree 10 is more than enough for building the continuous function $N_{UN}(a,t)$ from the annual data series from UN Database during the period under study.[27]

So, if we recover the variation of population due to TB and aging in each age group $\dot{N}_o(a,t)$, derived from equation 17, we can also introduce a term $\Delta_N(a,t)$, aimed, this time, not at balancing $\dot{N}_o(a,t)$, but at forcing the total temporal evolution of $N(a,t)$ to follow precisely the function $N_{UN}(a,t)$. This is achieved by defining, at each time step:

$$\Delta_N(a,t) = \dot{N}_{UN}(a,t) - \dot{N}_o(a,t) \tag{21}$$

and introducing those $\Delta_N(a,t)$ terms into the system dynamics, thus having: $\dot{N}(a,t) = \dot{N}_o(a,t) + \Delta_N(a,t) = \dot{N}_{UN}(a,t)$. Finally, provided that the initial conditions have been properly set, $N(a,t=0) = N_{UN}(a,t=0)\ \forall a$, this yields the desired behavior for the demographic structure, i.e., $N(a,t) = N_{UN}(a,t)\ \forall(a,t)$.

Again, the $\Delta_N(a,t)$ forcing terms have to be introduced into the different dynamical states within the same age class preserving their proportions, at least in the age groups $a > 0$:

$$\Delta_X(a,t) = \Delta_N(a,t)\frac{X(a,t)}{N(a,t)}\ \forall a > 0 \tag{22}$$

and, under this assumption, the terms $\Delta_N(a,t)$ for $a > 0$, represent the variations of volume of the age group a due to causes other than TB infection and individuals aging. This would include all deaths not caused by TB, and migration, assuming that these factors affect all the dynamical classes regardless of their state with respect to TB infection.

The assumption that migration occurs independently of the disease state is arguable, and, in principle, is hard to anticipate whether TB patients (or latent TB carriers) are more or less prune to migrate than susceptible individuals. However, without specific data that could motivate an informed alternative, we decided to take as the null hypothesis that all individuals are equally prune to migrate regardless of their TB status. Nonetheless, migratory fluxes do not represent the major cause of population variation in any of the twelve countries studied,

where the migratory balance (immigrants-emigrants) supposes less than 25% of the vegetative growth (less than 10% in 8 over 12 countries, all but South Africa, China, Myanmar and Bangladesh) during the period under study. Thus, if the actual reality is more complex than our assumptions, and migrants and not migrants do present different TB prevalence levels, the effects of these differences should be bound by the reduced role of migration in the total variation of the populations under study.

The situation is different for the first age class $a = 0$. In the first age group, the birth of new individuals is the main cause of population variation. For these reasons, and once observed that $\Delta_N(a = 0, t) > 0 \ \forall t$ in all countries under consideration, for simplicity $\Delta_N(a = 0, t)$ is directly associated to the number of newborns and introduced in the S, L_s and L_f states, as described in section 2.1.10.

The uncertainty of UN demographic projections is also reported at UN Database, [27] which allows us to reconstruct the demographic structures at the extremes of the confidence interval (95%) $N_{UN}^{\text{low}}(a,t)$ and $N_{UN}^{\text{high}}(a,t)$. Therefore, its influence on the model forecasts is also measurable, as we will discuss in the section devoted to uncertainty and sensitivity analysis (section 4).

2.6 Ordinary differential equations system

The following system of differential equations describes the evolution of the different dynamical states of the model:

$$\begin{aligned}
\dot{S}(a,t) &= -\lambda(a,t)S(a,t) - ((1-\delta(a-14))S(a,t) - (1-\delta(a))S(a-1,t))/\tau \\
&+ \delta(a)(1-m_c m_d(t))\Delta_N(a,t) + (1-\delta(a))\Delta_N(a,t)S(a,t)/N(a,t)
\end{aligned} \tag{23}$$

$$\begin{aligned}
\dot{L}_s(a,t) &= (1-p(a))\lambda(a,t)S(a,t) - p(a)q\lambda(a,t)L_s(a,t) - \omega_s L_s(a,t) + \delta(a)m_c m_d(t)(1-p(0))\Delta_N(a,t) \\
&- ((1-\delta(a-14))L_s(a,t) - (1-\delta(a))L_s(a-1,t))/\tau + (1-\delta(a))\Delta_N(a,t)L_s(a,t)/N(a,t)
\end{aligned} \tag{24}$$

$$\begin{aligned}
\dot{L}_f(a,t) &= p(a)\lambda(a,t)S(a,t) - \omega_f L_f(a,t) + p(a)q\lambda(a,t)(L_s(a,t) + R_{p+N}(a,t) + R_{p-N}(a,t) + R_{npN}(a,t)) \\
&+ p(a)q\lambda(a,t)(R_{p+S}(a,t) + R_{p-S}(a,t) + R_{npS}(a,t) + R_{p+D}(a,t) + R_{p-D}(a,t) + R_{npD}(a,t)) \\
&- ((1-\delta(a-14))L_f(a,t) - (1-\delta(a))L_f(a-1,t))/\tau + \delta(a)m_c m_d(t)p(0)\Delta_N(a,t) \\
&+ (1-\delta(a))\Delta_N(a,t)L_f(a,t)/N(a,t)
\end{aligned} \tag{25}$$

$$\begin{aligned}
\dot{D}_{p+}(a,t) &= \omega_f \rho_{p+}(a)L_f(a,t) + \omega_s \rho_{p+}(a)L_s(a,t) - \mu_{p+}D_{p+}(a,t) - d(t)D_{p+}(a,t) \\
&- \nu D_{p+}(a,t) + r_N R_{p+N}(a,t) + r_S R_{p+S}(a,t) + r_D R_{p+D}(a,t) + \theta D_{p-}(a,t) \\
&- ((1-\delta(a-14))D_{p+}(a,t) - (1-\delta(a))D_{p+}(a-1,t))/\tau + (1-\delta(a))\Delta_N(a,t)D_{p+}(a,t)/N(a,t)
\end{aligned} \tag{26}$$

$$\begin{aligned}
\dot{D}_{p-}(a,t) &= \omega_f(1-\rho_{p+}(a)-\rho_{np}(a))L_f(a,t) + \omega_s(1-\rho_{p+}(a)-\rho_{np}(a))L_s(a,t) - \mu_{p-}D_{p-}(a,t) \\
&- \eta d(t)D_{p-}(a,t) - \nu D_{p-}(a,t) + r_N R_{p-N}(a,t) + r_S R_{p-S}(a,t) + r_D R_{p-D}(a,t) - \theta D_{p-}(a,t) \\
&- ((1-\delta(a-14))D_{p-}(a,t) - (1-\delta(a))D_{p-}(a-1,t))/\tau + (1-\delta(a))\Delta_N(a,t)D_{p-}(a,t)/N(a,t)
\end{aligned} \tag{27}$$

$$\begin{aligned}
\dot{D}_{np}(a,t) &= \omega_f \rho_{np}(a)L_f(a,t) + \omega_s \rho_{np}(a)L_s(a,t) - \mu_{np}D_{np}(a,t) - \eta d(t)D_{np}(a,t) \\
&- \nu D_{np}(a,t) + r_N R_{npN}(a,t) + r_S R_{npS}(a,t) + r_D R_{npD}(a,t) \\
&- ((1-\delta(a-14))D_{np}(a,t) - (1-\delta(a))D_{np}(a-1,t))/\tau + (1-\delta(a))\Delta_N(a,t)D_{np}(a,t)/N(a,t)
\end{aligned} \tag{28}$$

$$\begin{aligned}
\dot{T}_{p+}(a,t) &= d(t)D_{p+}(a,t) - \Psi T_{p+}(a,t) + \theta T_{p-}(a,t) \\
&- ((1-\delta(a-14))T_{p+}(a,t) - (1-\delta(a))T_{p+}(a-1,t))/\tau + (1-\delta(a))\Delta_N(a,t)T_{p+}(a,t)/N(a,t)
\end{aligned} \tag{29}$$

$$\begin{aligned}
\dot{T}_{p-}(a,t) &= \eta d(t)D_{p-}(a,t) - \Psi T_{p-}(a,t) - \theta T_{p-}(a,t) \\
&- ((1-\delta(a-14))T_{p-}(a,t) - (1-\delta(a))T_{p-}(a-1,t))/\tau + (1-\delta(a))\Delta_N(a,t)T_{p-}(a,t)/N(a,t)
\end{aligned} \tag{30}$$

$$\begin{aligned}
\dot{T}_{np}(a,t) &= \eta d(t)D_{np}(a,t) - \Psi T_{np}(a,t) \\
&- ((1-\delta(a-14))T_{np}(a,t) - (1-\delta(a))T_{np}(a-1,t))/\tau + (1-\delta(a))\Delta_N(a,t)T_{np}(a,t)/N(a,t)
\end{aligned} \tag{31}$$

$$\begin{aligned}\dot{F}(a,t) &= \Psi f_F^{p+} T_{p+}(a,t) + \Psi f_F^{p-}(T_{p-}(a,t) + T_{np}(a,t)) - \mu_{p+} F(a,t) \\ &- ((1-\delta(a-14))F(a,t) - (1-\delta(a))F(a-1,t))/\tau + (1-\delta(a))\Delta_N(a,t)F(a,t)/N(a,t)\end{aligned} \tag{32}$$

$$\begin{aligned}\dot{R}_{p+N}(a,t) &= \nu D_{p+}(a,t) - r_N R_{p+N}(a,t) - p(a)q\lambda(a,t)R_{p+N}(a,t) \\ &- ((1-\delta(a-14))R_{p+N}(a,t) - (1-\delta(a))R_{p+N}(a-1,t))/\tau + (1-\delta(a))\Delta_N(a,t)R_{p+N}(a,t)/N(a,t)\end{aligned} \tag{33}$$

$$\begin{aligned}\dot{R}_{p-N}(a,t) &= \nu D_{p-}(a,t) - r_N R_{p-N}(a,t) - p(a)q\lambda(a,t)R_{p-N}(a,t) \\ &- ((1-\delta(a-14))R_{p-N}(a,t) - (1-\delta(a))R_{p-N}(a-1,t))/\tau + (1-\delta(a))\Delta_N(a,t)R_{p-N}(a,t)/N(a,t)\end{aligned} \tag{34}$$

$$\begin{aligned}\dot{R}_{npN}(a,t) &= \nu D_{np}(a,t) - r_N R_{npN}(a,t) - p(a)q\lambda(a,t)R_{npN}(a,t) \\ &- ((1-\delta(a-14))R_{npN}(a,t) - (1-\delta(a))R_{npN}(a-1,t))/\tau + (1-\delta(a))\Delta_N(a,t)R_{npN}(a,t)/N(a,t)\end{aligned} \tag{35}$$

$$\begin{aligned}\dot{R}_{p+S}(a,t) &= \Psi(1-f_D^{p+}-f_F^{p+}-f_\mu^{p+})T_{p+}(a,t) - r_S R_{p+S}(a,t) - p(a)q\lambda(a,t)R_{p+S}(a,t) \\ &- ((1-\delta(a-14))R_{p+S}(a,t) - (1-\delta(a))R_{p+S}(a-1,t))/\tau + (1-\delta(a))\Delta_N(a,t)R_{p+S}(a,t)/N(a,t)\end{aligned} \tag{36}$$

$$\begin{aligned}\dot{R}_{p-S}(a,t) &= \Psi(1-f_D^{p-}-f_F^{p-}-f_\mu^{p-})T_{p-}(a,t) - r_S R_{p-S}(a,t) - p(a)q\lambda(a,t)R_{p-S}(a,t) \\ &- ((1-\delta(a-14))R_{p-S}(a,t) - (1-\delta(a))R_{p-S}(a-1,t))/\tau + (1-\delta(a))\Delta_N(a,t)R_{p-S}(a,t)/N(a,t)\end{aligned} \tag{37}$$

$$\begin{aligned}\dot{R}_{npS}(a,t) &= \Psi(1-f_D^{p-}-f_F^{p-}-f_\mu^{p-})T_{np}(a,t) - r_S R_{npS}(a,t) - p(a)q\lambda(a,t)R_{npS}(a,t) \\ &- ((1-\delta(a-14))R_{npS}(a,t) - (1-\delta(a))R_{npS}(a-1,t))/\tau + (1-\delta(a))\Delta_N(a,t)R_{npS}(a,t)/N(a,t)\end{aligned} \tag{38}$$

$$\begin{aligned}\dot{R}_{p+D}(a,t) &= \Psi f_D^{p+} T_{p+}(a,t) - r_D R_{p+D}(a,t) - p(a)q\lambda(a,t)R_{p+D}(a,t) \\ &- ((1-\delta(a-14))R_{p+D}(a,t) - (1-\delta(a))R_{p+D}(a-1,t))/\tau + (1-\delta(a))\Delta_N(a,t)R_{p+D}(a,t)/N(a,t)\end{aligned} \tag{39}$$

$$\begin{aligned}\dot{R}_{p-D}(a,t) &= \Psi f_D^{p-} T_{p-}(a,t) - r_D R_{p-D}(a,t) - p(a)q\lambda(a,t)R_{p-D}(a,t) \\ &- ((1-\delta(a-14))R_{p-D}(a,t) - (1-\delta(a))R_{p-D}(a-1,t))/\tau + (1-\delta(a))\Delta_N(a,t)R_{p-D}(a,t)/N(a,t)\end{aligned} \tag{40}$$

$$\begin{aligned}\dot{R}_{npD}(a,t) &= \Psi f_D^{p-} T_{np}(a,t) - r_D R_{npD}(a,t) - p(a)q\lambda(a,t)R_{npD}(a,t) \\ &- ((1-\delta(a-14))R_{npD}(a,t) - (1-\delta(a))R_{npD}(a-1,t))/\tau + (1-\delta(a))\Delta_N(a,t)R_{npD}(a,t)/N(a,t)\end{aligned} \tag{41}$$

where $\delta(a)$ stands for the Dirac delta function ($\delta(x=0)=1$ and $\delta(x\neq 0)=0$). There are three quantities that depend on time: the force of infection $\lambda(a,t)$, the diagnosis rate $d(t)$ and the correction terms $\Delta_N(a,t)$, standing for any demographic variation in the population due to causes foreign to TB and aging.

It is interesting to define two additional variables, fully dependent on the dynamical state of the system, such as the accumulated number of TB incident cases in each age group, from the beginning of the period under analysis $I(a,t)$, and the accumulated number of TB deaths equally defined $M(a,t)$. Their respective temporal evolution reads as follows:

$$\begin{aligned}\dot{I}(a,t) &= \omega_f L_f(a,t) + \omega_s L_s(a,t) + r_N\left(R_{p+N}(a,t) + R_{p-N}(a,t) + R_{npN}(a,t)\right) \\ &+ r_S\left(R_{p+S}(a,t) + R_{p-S}(a,t) + R_{npS}(a,t)\right) + r_D\left(R_{p+D}(a,t) + R_{p-D}(a,t) + R_{npD}(a,t)\right)\end{aligned} \tag{42}$$

$$\begin{aligned}\dot{M}(a,t) &= \mu_{p+}(D_{p+}(a,t) + F(a,t)) + \mu_{p-}D_{p-}(a,t) + \mu_{np}D_{np}(a,t) \\ &+ \Psi f_\mu^{p+} T_{p+}(a,t) + \Psi f_\mu^{p-}(T_{p-}(a,t) + T_{np}(a,t))\end{aligned} \tag{43}$$

From these variables, once summed over all age groups, we explicitly get the incidence rate as the number of new cases per year $i(t)$, and the annual mortality rate as the total number of TB deaths per year $m(t)$, both normalized by 1000000 individuals:

$$i(t) = \frac{100000 \cdot \sum_a \left(I(a,t+1) - I(a,t)\right)}{(\mathcal{N}(t+1) + \mathcal{N}(t))/2} \tag{44}$$

$$m(t) = \frac{100000 \cdot \sum_a \left(M(a,t+1) - M(a,t)\right)}{(\mathcal{N}(t+1) + \mathcal{N}(t))/2} \tag{45}$$

The sums of $I(a,t)$ and $M(a,t)$ over all ages at the end of the period under study provide the total number of cases and deaths due to the disease during the whole period.

2.7 Initial conditions setup

Once we have detailed the forces driving the time evolution of our state variables, it remains to be clarified how the initial conditions $\vec{X}(a, t=0)$ for each possible state $\vec{X}_i$ are set. This problem is traditionally solved just by considering that, at the beginning of the period analyzed, the system is at a stationary state that is reached after fixing the temporal evolution of the time dependent parameters to their values at the beginning of the period: $d(t=0) = d_0$ and $\beta(t=0) = \beta_0$, as well as the demographic boundary conditions $N(a,t) = N(a,0)$, where $N(a,t)$ represents the populations at each age group.[3,4] We denote those stationary levels as $\vec{X}^*(a, d_0, \beta_0, N(a,0))$, so we have $\dot{X}_i^* = 0 \ \forall(i,t)$; provided that all the time-dependent parameters and demographic forcing terms are frozen in their initial values at $t = t_o$. Accordingly, the stationary vector $\vec{X}^*$ is used to set up the initial conditions of the system: $\vec{X}(a,0) = \vec{X}^*$.

In this work, we do not impose that the system must be at stationarity at $t = 0$. Instead, we calculate the stationary values of all states $\vec{X}^*(a, d_0, \beta_0, \vec{N}(a,0))$, and we set up an initial state that can correspond either to higher or lower levels of disease prevalence. In order to map these possible variations on TB burden from the stationary vector of states $\vec{X}^*$, we distinguish the unexposed state, $S(a,t)$, from the rest of the states joined by individuals that have been infected with the bacillus at least once. Finally, we define a parameter $\varsigma \in [-1, 1]$, such that, when $\varsigma < 0$, the initial conditions correspond to a state with lower TB burden than that in the stationary state:

$$X(a, t=0) = (1+\varsigma) X^*(a, d_0, \beta_0, \vec{N}(a,0)) \qquad \forall (X \neq S) \tag{46}$$

$$S(a, t=0) = S^*(a, d_0, \beta_0, \vec{N}(a,0)) \left(1 - \varsigma \sum_{X \neq S} X^*(a, d_0, \beta_0, \vec{N}(a,0))\right) \tag{47}$$

Instead, if $\varsigma > 0$, the initial conditions are set to higher burden levels from stationarity:

$$S(a, t=0) = S^*(a, d_0, \beta_0, \vec{N}(a,0))(1-\varsigma) \tag{48}$$

$$X(a, t=0) = X^*(a, d_0, \beta_0, \vec{N}(a,0)) \left(1 + \frac{\varsigma S^*(a, d_0, \beta_0, \vec{N}(a,0))}{\sum_{X \neq S} X^*(a, d_0, \beta_0, \vec{N}(a,0))}\right) \tag{49}$$

Taking it to their extreme values, $\varsigma = -1$ would mean that every individual is at the susceptible state (pathogen-free situation) while $\varsigma = 1$ would mean that all the population is infected with the bacterium. $\varsigma = 0$ would imply that the initial conditions of the system are those from the stationary state. The previous definition ensures that, at any moment, the sum of individuals in all the states provides the desired population volumes regardless of how far, or in which sense ς shifts the initial condition from the stationary state defined by the vector $\vec{X}^*$.

2.8 Model calibration procedure

The calibration procedure of our model implies the estimation, for each country, of the initial conditions of the system, parametrized through the ς coordinate, along with the diagnosis rate $d(t)$ and the scaled infectiousness $\beta(t)$ that make the model reproduce the TB burden mortality and incidence rates reported by the WHO from $t_o = 2000$ to $t_F = 2015$. Both parameters $d(t)$ and $\beta(t)$ are fitted to half-sigmoid-like curves, as follows:

$$d(t) = \begin{cases} d_0 + (d_{\sup} - d_0)t(t + \frac{1}{d_1})^{-1} & \text{if } d_1 > 0 \\ d_0 & \text{if } d_1 = 0 \\ d_0 - d_0 t(t - \frac{1}{d_1})^{-1} & \text{if } d_1 < 0 \end{cases} \tag{50}$$

$$\beta(t) = \begin{cases} \beta_0 + \beta_0 t \left(t + \frac{1}{\beta_1}\right)^{-1} & \beta_1 > 0 \\ \beta_0 & \beta_1 = 0 \\ \beta_0 - \beta_0 t \left(t - \frac{1}{\beta_1}\right)^{-1} & \beta_1 < 0 \end{cases} \tag{51}$$

Therefore, the diagnosis rate and the scaled infectiousness are, each of them, parameterized by two quantities (d_0, d_1 and β_0, β_1). While d_0 and β_0 give the value of the diagnosis rate and scaled infectiousness at the beginning of the temporal window (i.e. year 2000), d_1 and β_1 define their evolution, either increasing or decreasing with time depending on the sign of d_1 and β_1. In case of a decreasing evolution, both the diagnosis rate and the scaled infectiousness are bounded to be greater than zero, while in the case of increasing evolution the upper bounds are $2 \times \beta_0$ for the scaled infectiousness and $d_{\sup} = 12.17y^{-1}$ for the diagnosis rate. This latter upper bound corresponds to a minimum diagnosis period of one month. We consider this minimum delay as reasonable, since the main symptom of TB is a continuous cough during three weeks, and, after that, there is a diagnostic process which is estimated to last, assuming a conservative lower boundary, at least 10 days. [28]

We have chosen this parameterization of the temporal evolution of the diagnosis rate and the infectiousness because, unlike previous approaches where the evolution of these parameters is assimilated to an exponential curve, [3,4] it provides a bounded growth for them, through a function that is still continuous and differentiable but does not introduce more parameters.

The goal of the calibration procedure is to minimize the overall error H of the model outcome with respect to the input burden measurements (aggregated incidence and mortality rates), calculated as follows:

$$H = \sum_{t=t_o}^{t_F} \left(\left(\frac{i(t) - \bar{i}(t)}{\bar{\Delta}_i(t)} \right)^2 + \left(\frac{m(t) - \bar{m}(t)}{\bar{\Delta}_m(t)} \right)^2 \right) \tag{52}$$

where $\bar{i}(t)$ and $\bar{m}(t)$ stand for the annual incidence and mortality rates, corresponding to the national estimations available at the WHO database for TB. [6] These measurements of TB incidence and mortality have their correspondent confidence intervals $(\bar{i}_{\text{low}}(t), \bar{i}_{\text{high}}(t))$ and $(\bar{m}_{\text{low}}(t), \bar{m}_{\text{high}}(t))$, which are not necessarily symmetrical with respect to the central values $\bar{i}(t)$ and $\bar{m}(t)$. Using these confidence intervals, and taking into consideration their asymmetry, the corresponding terms $\bar{\Delta}_i(t)$ $\bar{\Delta}_m(t)$ are constructed as follows:

$$\bar{\Delta}_i(t) = \begin{cases} \bar{i}(t) - \bar{i}^{\text{low}}(t) & \text{if } i(t) \leq \bar{i}(t) \\ \bar{i}^{\text{high}}(t) - \bar{i}(t) & \text{if } i(t) > \bar{i}(t) \end{cases} \tag{53}$$

$$\bar{\Delta}_m(t) = \begin{cases} \bar{m}(t) - \bar{m}^{\text{low}}(t) & \text{if } m(t) \leq \bar{m}(t) \\ \bar{m}^{\text{high}}(t) - \bar{m}(t) & \text{if } m(t) > \bar{m}(t) \end{cases} \tag{54}$$

In the case of China and Philippines the very small uncertainty on mortality data (directly zero for some particular years) prevents us of using the previous equation 52. In those cases we minimize the absolute distance given by:

$$H = \sum_{t=t_o}^{t_F} \left(\left(\frac{i(t) - \bar{i}(t)}{\langle \bar{i}(t) \rangle} \right)^2 + \left(\frac{m(t) - \bar{m}(t)}{\langle \bar{m}(t) \rangle} \right)^2 \right) \tag{55}$$

where $\langle \bar{i}(t) \rangle$ and $\langle \bar{m}(t) \rangle$ correspond to the averages of incidence and mortality reported by the WHO for the entire period in each country, respectively.

The conceptual scheme for the fitting of these parameters essentially consists in an iterative evaluation of the model across the parameter space $(\varsigma, d_0, \beta_0, d_1, \beta_1)$, which is navigated according to a certain "routing" that eventually guarantees the localization of a parameter set that yields an error H which constitutes a local minimum of the objective function H. In our case, we have used a Levenberg-Marquardt algorithm to solve these multidimensional optimization problem, implemented, as for the rest of the model, in programming language C. [29] See figure S14 for a graphic summary of the procedure.

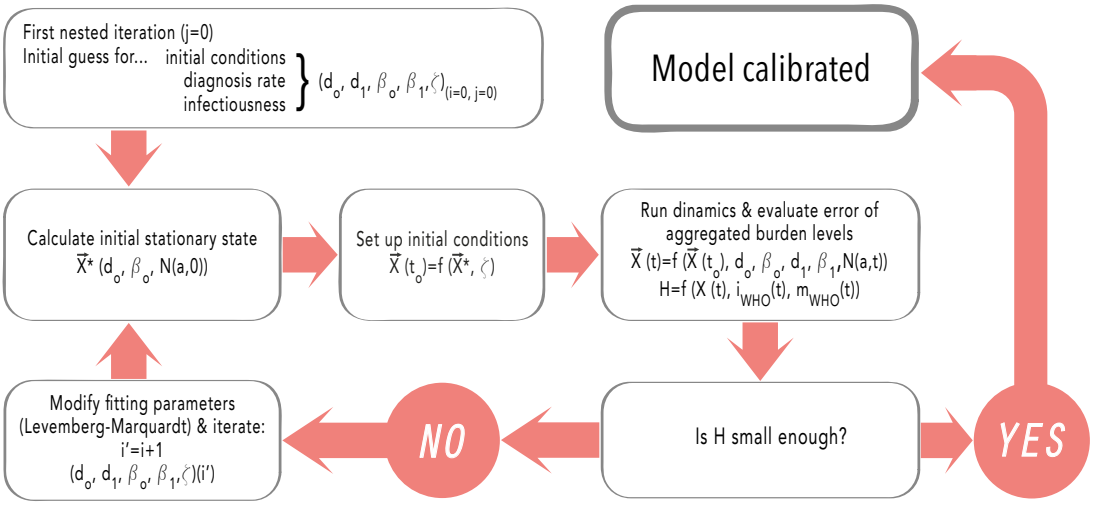

Figure S14: Schematic representation of the calibration algorithm.

3 Model states and parameters summary

In this section, we summarize all the dynamical states and parameters used in the model, along with their values, definitions, confidence intervals and bibliographical sources.

3.1 Dynamic states

State	Definition
$S(a,t)$	Susceptible (not previously exposed to infection) individuals
$L_s(a,t)$	Infected individuals (slow latency)
$L_f(a,t)$	Infected individuals who will develop fast progression
$D_{p+}(a,t)$	Untreated sick individuals: Smear positive pulmonar disease
$D_{p-}(a,t)$	Untreated sick individuals: Smear negative pulmonar disease
$D_{np}(a,t)$	Untreated sick individuals: non pulmonar disease
$T_{p+}(a,t)$	Sick individuals under treatment: Smear positive pulmonar disease
$T_{p-}(a,t)$	Sick individuals under treatment: Smear negative pulmonar disease
$T_{np}(a,t)$	Sick individuals under treatment: non pulmonar disease
$F(a,t)$	Patients who faultily finished their treatment.
$R_{p+S}(a,t)$	Patients of smear positive pulmonary TB who successfully finished their treatment.
$R_{p+D}(a,t)$	Patients of smear positive pulmonary TB who defaulted their treatment by two consecutive months or more.
$R_{p+N}(a,t)$	Patients of smear positive pulmonary TB that naturally recovered –without treatment– from the disease.
$R_{p-S}(a,t)$	Patients of smear negative pulmonary TB who successfully finished their treatment.
$R_{p-D}(a,t)$	Patients of smear negative pulmonary TB who defaulted their treatment by two consecutive months or more.
$R_{p-N}(a,t)$	Patients of smear negative pulmonary TB that naturally recovered –without treatment– from the disease.
$R_{npS}(a,t)$	Patients of non pulmonary TB who successfully finished their treatment.
$R_{npD}(a,t)$	Patients of non pulmonary TB who defaulted their treatment by two consecutive months or more.
$R_{npN}(a,t)$	Patients of non pulmonary TB that naturally recovered –without treatment– from the disease.

Table S13: Description of the different dynamic states considered on the model

3.2 Literature-based epidemiological parameters

Meaning	Parameter	Value	C.I.	Reference	Section
Probability of fast progression	$p(a)$	$(a=0)$ 0.187	(0.1474,0.2333)	Marais et al.[7], this work	2.1.1
		$(a=1)$ 0.0225	(0.0200,0.0250)		
		$(a>1)$ 0.15	(0.10,0.20)		
Rate of fast progression (y^{-1})	ω_f	0.900	(0.765,1.035)	Abu-Raddad et al.[4]	2.1.2
Rate of slow progression (y^{-1})	ω_s	7.500×10^{-4}	$(6.375,8.625)\times10^{-4}$	Abu-Raddad et al.[4]	2.1.2
Probability of developing pulmonary smear-positive disease	$\rho_{p+}(a)$	$(a<3)$ 0.100	(0.085,0.115)	Abu-Raddad et al.[4]	2.1.2
		$(a\geq3)$ 0.500	(0.425,0.575)		
Probability of developing non-pulmonary disease	$\rho_{np}(a)$	$(a<3)$ 0.250	(0.2125,0.2875)	Abu-Raddad et al.[4]	2.1.2
		$(a\geq3)$ 0.100	(0.085,0.115)		
Mortality rate by pulmonary smear positive TB (y^{-1})	μ_{p+}	0.250	(0.213,0.288)	Abu-Raddad et al.[4]	2.1.3
Mortality rate by pulmonary smear negative TB (y^{-1})	μ_{p-}	0.100	(0.085,0.115)	Abu-Raddad et al.[4]	2.1.3
Mortality rate by non-pulmonary TB (y^{-1})	μ_{np}	0.100	(0.085,0.115)	Abu-Raddad et al.[4]	2.1.3
Reduction of infection risk for previously infected individuals	q	0.650	(0.553,748)	Abu-Raddad et al.[4]	2.1.8
Treatment completion rate (y^{-1})	Ψ	2.00	(1.70,2.30)	Abu-Raddad et al.[4]	2.1.5
Smear progression rate (y^{-1})	θ	0.015	(0.007,0.020)	Dye et al.[3]	2.1.9
Relapse rate for individuals who successfully completed treatment (y^{-1})	r_S	9.392×10^{-4}	$(6.364,12.450)\times10^{-4}$	Korenromp et al.[14], this work	2.1.7
Relapse rate for individuals who defaulted treatment (y^{-1})	r_D	3.774×10^{-3}	$(1.354,8.620)\times10^{-3}$	Korenromp et al.[14], Picon et al.[15],this work	2.1.7
Relapse rate for naturally recovered individuals (y^{-1})	r_N	0.030	(0.020,0.040)	Dye et al.[3]	2.1.7
Natural recovery rate (y^{-1})	ν	0.100	(0.085,0.115)	Dye et al.[3]	2.1.6
Infectiousness reduction coefficient of D_{p-} with respect to D_{p+}	ϕ_{p-}	0.250	(0.213,0.288)	Abu-Raddad et al.[4]	2.2
Infectiousness reduction coefficient of R_{p+D} with respect to D_{p+}	ϕ_D	0.500	(0.250,0.750)	Dye et al.[3]	2.2
Proportion of mothers that infect their newborn children	m_c	0.15	(0.10,0.20)	Pillay et al.[17]	2.1.10
Diagnosis rate reduction of D_{p-} and D_{np} with respect to D_{p+}	η	Ethiopia, Nigeria: 0.843	(0.664,1.022)	Abu-Raddad et al.[4], this work	2.1.4
		India, Indonesia: 0.797	(0.628,0.966)		

Table S14: Bibliography-based epidemiological parameters used in this study.

In table S14 we represent the 19 epidemiological parameters used in our model, along with the eventual dependencies each of them show (to age, geographic setting, or none), the bibliographic source and the section of the appendix where the meaning of each parameter is explained.

3.3 Treatment outcomes probabilities

The probabilities of individuals to end their treatment according the four categories defined by the WHO (success, default, failure or death), defined as:

- $(f_D^{p+}, f_F^{p+}, f_\mu^{p+})$: fraction of default, failure and death outcomes for smear positive pulmonary TB. [6]
- $(f_D^{p-}, f_F^{p-}, f_\mu^{p-})$: fraction of default, failure and death outcomes for smear negative pulmonary and non pulmonary TB. [6]

have been obtained from the WHO Treatment Outcomes database for each country, and their values in Ethiopia, Nigeria, India e Indonesia are presented in table S15:

Parameter	Ethiopia	Nigeria	India	Indonesia	Reference	Section
f_D^{p+} (%)	3.84 (3.74,3.94)	8.51 (8.36,8.66)	5.97 (5.95,6.00)	4.62 (4.58,4.66)	WHO Database[6]	2.1.5
f_F^{p+} (%)	1.04 (0.99,1.10)	1.23 (1.16,1.29)	2.07 (2.05,2.08)	0.62 (0.61,0.64)	WHO Database[6]	2.1.5
f_μ^{p+} (%)	3.97 (3.87,4.08)	5.64 (5.51,5.76)	4.48 (4.46,4.51)	2.31 (2.28,2.34)	WHO Database[6]	2.1.5
f_D^{p-} (%)	3.28 (3.22,3.35)	6.58 (6.43,6.72)	6.11 (6.09,6.13)	7.18 (7.12,7.24)	WHO Database[6]	2.1.5
f_F^{p-} (%)	0.12 (0.11,0.13)	0.24 (0.21,0.27)	0.42 (0.41,0.43)	0.27 (0.26,0.28)	WHO Database[6]	2.1.5
f_μ^{p-} (%)	3.53 (3.46,3.59)	6.27 (6.13,6.41)	3.11 (3.09,3.13)	1.98 (1.94,2.01)	WHO Database[6]	2.1.5

Table S15: Values of the treatment outcomes probabilities in Ethiopia, Nigeria, India and Indonesia.

3.4 Initial conditions and fitted parameters (Diagnosis rate, and scaled infectiousness)

Once all the mentioned parameters are fixed, we obtain the time-evolving parameterization of diagnosis rates and infectiousness as the result of the calibration procedure explained in section 2.8, along with the initial conditions of the system in each country. These temporal evolutions are derived from equations 50 and 51, while the fitted values of the parameters $(d_o, d_1, \beta_o, \beta_1)$ are reported in table S16 for the 4 countries discussed in the main text. Confidence Intervals are obtained through the procedure explained in section 4, as we do for any other model outcome.

Country	$d_0\ (y^{-1})$	$d_1 \times 10^{-3}\ (y^{-1})$	$\beta_0\ (y^{-1})$	$\beta_1 \times 10^{-3}\ (y^{-1})$	ς
Ethiopia	0.19 (0.15,0.25)	4.75 (3.94,5.64)	7.10 (4.57,10.90)	73.84 (20.62,103.83)	0.16 (0.16,0.25)
Nigeria	0.045 (0.003,0.088)	0.36 (0.30,0.43)	5.30 (3.36,8.03)	3.71 (-3.93,11.15)	-0.29 (-0.53,-0.02)
India	0.51 (0.12,1.06)	2.98 (1.80,4.26)	10.12 (5.10,16.99)	-1.20 (-12.01,9.19)	-0.30 (-0.44,-0.15)
Indonesia	1.53 (1.31,1.80)	-7.60 (-20.99,-0.42)	16.25 (10.82,25.59)	-12.73 (-17.73,-8.97)	0.05 (0.04,0.07)

Table S16: Fitted parameters for different countries.

In figure S15 we represent the evolution of $d(t)$ in these countries, which describes the average rate at which sick individuals receive their diagnosis in each country and time. The scaled infectiousness $\beta(t)$ has a less immediate epidemiological interpretation, for it is only directly proportional to the number of infections $\tilde{R}_0$, that is caused, on average, by each infectious agent, which also depends on the distribution of individuals among the different infectious classes and age groups. This magnitude (which reduces to the basic reproductive number R_0 when evaluated on a fully susceptible population) is also represented in figure S15.

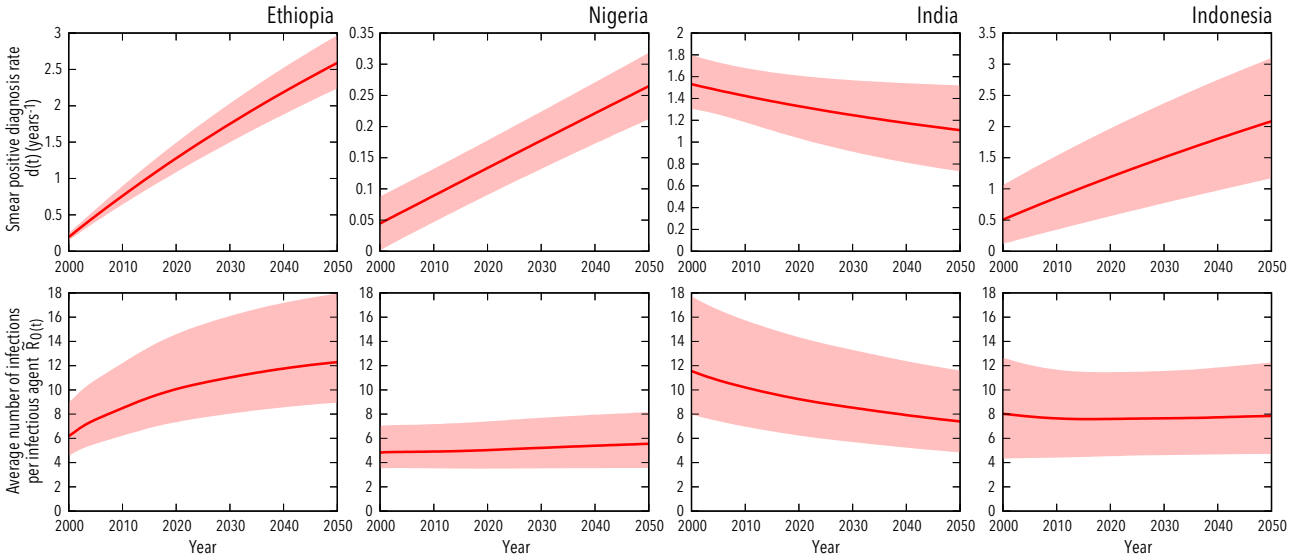

Figure S15: (A) Diagnosis rates for smear-positive individuals (in years^{-1}). (B) Average number of secondary infections per infectious agent $\tilde{R}_0$. The red curve shows the result given by the central fit, while the shadowed area represents the uncertainty obtained as described in section 4

3.5 Fitted parameters in the reduced models

In table S17, we present the fitted parameters for the two reduced models used across the main text in India, Indonesia, Nigeria and Ethiopia, compared to the values from the full model (already reported in S16).

Country	Model	d_0 (y^{-1})	$d_1 \times 10^{-3}$ (y^{-1})	β_0 (y^{-1})	$\beta_1 \times 10^{-3}$ (y^{-1})	ς
Ethiopia	Complete Model	0.19 (0.15,0.25)	4.75 (3.94,5.64)	7.10 (4.57,10.90)	73.84 (20.62,103.83)	0.16 (0.16,0.25)
	Constant Demography	0.22 (0.16,0.27)	4.40 (3.68,5.24)	7.22 (4.64,11.02)	70.97 (28.80,91.43)	0.21 (0.05,0.31)
	Homogeneous Mixing	0.19 (0.15,0.25)	4.56 (3.80,5.36)	7.18 (5.12,10.73)	80.78 (36.04,110.66)	0.23 (0.09,0.34)
Nigeria	Complete Model	0.045 (0.003,0.088)	0.36 (0.30,0.43)	5.30 (3.36,8.03)	3.71 (-3.93,11.15)	-0.29 (-0.53,-0.02)
	Constant Demography	0.046 (0.004,0.090)	0.37 (0.31,0.43)	5.30 (3.35,8.04)	2.71 (-3.67,9.85)	-0.27 (-0.52,-0.01)
	Homogeneous Mixing	0.038 (0.003,0.080)	0.35 (0.14,0.40)	5.52 (3.81,8.15)	2.91 (0.33,10.50)	-0.36 (-0.46,0.10)
India	Complete Model	0.51 (0.12,1.06)	2.98 (1.80,4.26)	10.12 (5.10,16.99)	-1.20 (-12.01,9.19)	-0.30 (-0.44,-0.15)
	Constant Demography	0.51 (0.13,1.04)	2.66 (1.51,4.05)	10.07 (5.12,16.78)	6.16 (-0.39,12.97)	-0.25 (-0.37,-0.10)
	Homogeneous Mixing	0.51 (0.12,1.06)	2.94 (1.83,4.18)	9.12 (4.55,15.20)	-1.04 (-11.38,11.52)	-0.30 (-0.45,-0.15)
Indonesia	Complete Model	1.53 (1.31,1.80)	-7.60 (-20.99,-0.42)	16.25 (10.82,25.59)	-12.73 (-17.73,-8.97)	0.05 (0.04,0.07)
	Constant Demography	1.53 (1.31,1.80)	-8.71 (-22.14,-1.40)	16.14 (10.74,25.42)	-7.51 (-11.74,-4.83)	0.08 (0.06,0.11)
	Homogeneous Mixing	1.53 (1.31,1.80)	-8.09 (-21.64,-0.84)	14.68 (9.63,23.38)	-12.20 (-16.72,-9.07)	0.06 (0.05,0.08)

Table S17: Fitted parameters for different countries and models.

3.6 Data Sources summary

In this section we summarize the structure and origin of different pieces of data used in this work. A scheme on how these data is included in our model can be found in Figure 1D of the main text.

- Annual rates of incidence and mortality from 2000 to 2015 for the different countries studied. WHO TB burden estimates database.[6] These data are used to calibrate, for each country independently, the scaled infectiousness (β_0, β_1), diagnosis rate (d_0,d_1) and initial distance to stationarity ς.
- Treatment outcomes for the different countries. WHO TB treatment outcomes database.[6] Used to extract the parameters: f_D^{p+}, f_F^{p+}, f_μ^{p+}, f_D^{p-}, f_F^{p-} and f_μ^{p-} (fraction of individuals experimenting the different possible treatment outcomes)
- Population of each age group and country from 2000 to 2050. UN population division database.[27] From this data we extract the demographic structures that our populations are forced to follow as explained in section 2.5

- Age contact patterns from different experimental settings. [5,22,23,24,25,26] Used to construct the contact rates between age groups (see sections 2.2 and 2.3).
- Different bibliographical sources, [3,4,12,14,15] from which we extract values for 19 epidemiological parameters (see table S14).

4 Model uncertainty and sensitivity analysis

All the input data sources mentioned in the previous section, and summarized in Figure 1D of the main text, carry intrinsic uncertainties whose influence on both fitted parameters and forecasts has to be evaluated. To this end, we have performed exhaustive uncertainty and sensitivity analyses that allow us to generate confidence intervals for our model outcomes, produce significance estimates (i.e. p-values) as well as to evaluate the part of this uncertainty that is propagated from each of the model inputs.

4.1 Uncertainty sources analysis

In our model, we consider four main different types of inputs, which are associated to independent uncertainty sources for the sake of our sensitivity/uncertainty analysis:

- Parameters associated with the Natural History of the disease: a total amount of 19 parameters, listed in table S14, each of them conservatively treated as totally independent uncertainty sources $u_i, i \in [1, 19]$.
- Burden and treatment outcomes estimations provided by WHO, listed in table S15. Based upon a number of case notifications and treatment outcomes of finite cohorts surveilled in each country, the World Health Organization provides estimations for incidence and mortality rates $\bar{i}(t)$ and $\bar{m}(t)$ and for the treatment outcome fractions $(f_D^{p+}, f_F^{p+}, f_\mu^{p+})$, and $(f_D^{p-}, f_F^{p-}, f_\mu^{p-})$. For the purpose of our model, we have grouped these estimations produced by the WHO TB division as mutually dependent (see figure 1D in the main text), and considered them as one uncertainty source, labelled as u_{20}.
- Demographic structures $N(a, t)$: which are also considered as a single uncertainty source, labeled as u_{21}.
- Contact matrix $\xi(a, a', t)$, whose uncertainty comes from the variability between studies, is the last single uncertainty source u_{22}.

By proceeding in this way, we have 22 uncertainty sources $u_i, i \in [1, 22]$ that are considered independent, whose contributions to the uncertainty of a certain model outcome x is our goal to evaluate. Here, a model outcome can be any possible magnitude that derives from our entire calibration-simulation procedure, as summarized in figure 1D of the main text. This includes, among others, incidence and mortality rates evaluated at any time (or averaged during the entire period), total accumulated number of incident cases or deaths, values for the parameters fitted during the calibration step, and importantly, differences between the full and the reduced models, either absolute or relative, of any of these primary outcomes. Our entire model-based calibration + simulation procedure, summarized in figure 1D of the main text, can be expressed, for what regards the estimation of any generic model outcome x, as a generic functional relationship $x = f(\vec{u})$, where $\vec{u}$ represents the 22-dimensional vector of uncertainty sources (i.e., input data).

Altogether, the computation of model sensitivity to singular input uncertainty sources and its grouping into overall model uncertainty can be summarized according to the following steps:

4.1.1 Estimation of singular sensitivities of model outcome x to individual variations in uncertainty source u_i. (Sensitivity analysis)

First, given a generic model outcome x, its sensitivity to a given uncertainty source $u(i)$ with a 95% confidence interval equal to $(u_i^{\text{low}}, u_i^{\text{high}})$ is defined as its variation in response to a deviation in $u(i)$ towards the lower limit of its confidence interval:

$$d_i^{\text{low}}(x) = x(u_1, ..., u_i^{\text{low}}, ..., u_{21}) - x(\vec{u}) \tag{56}$$

or towards the upper limit:

$$d_i^{\text{high}}(x) = x(u_1, ..., u_i^{\text{high}}, ..., u_{21}) - x(\vec{u}) \tag{57}$$

Importantly, since the variation in u_i that precedes the estimation of $d_i^{\text{low}}(x)$ and $d_i^{\text{high}}(x)$ occurs before model calibration, we are capturing, through this approach, the sensitivity of our entire procedure -calibration included- to the uncertainty input u_i, which implies that the signs of $d_i^{\text{low}}(x)$ and $d_i^{\text{high}}(x)$ cannot be trivially anticipated and can also coincide, as we see, in some cases, in figure S3. In that figure, we use red (blue) bars to represent the variations in the total number of TB cases/deaths that the model produces in 2000-2050 as a consequence of increasing (decreasing) the value of each uncertainty source to the upper (lower) limit of its respective confidence intervals.

Furthermore, it is important to note that some uncertainty sources report several parameters (not just one) whose confidence intervals plausibly carry strong correlations. That is the case of the age dependent parameters, multi-dimensional demographic structures and contact matrices, as well as the WHO estimations, which comprise several measurements of different nature (treatment outcomes fractions, mortalities and incidence rates). In these cases, where single uncertainty sources u_i consist of multi-dimensional correlated data, the sensitivity terms $d_i^{\text{low}}(x)$ and $d_i^{\text{high}}(x)$ are calculated upon variation of all the components of u_i to the limits of their confidence intervals as a block.

4.1.2 Grouping individual sensitivities according type of input data. Generation of confidence intervals and significance levels (Uncertainty analysis)

Once the individual sensitivity of all the n=22 sources of uncertainty are computed following the approach explained in the previous section, for the case of the 19 bibliographical parameters we separate positive versus negative sensitivities (i.e. sensitivity instances where the shift in the uncertainty source translates into an increase or a decrease in the model outcome), represented as red vs. blue bars in figure S3. Then, the square root of the sum of the squares of each type (positive and negative sensitivities) are computed. Denoted as $\Delta(x)^{\text{high}}_{\text{param}}$ and $\Delta(x)^{\text{low}}_{\text{param}}$, respectively, these quantities can be formally defined as follows:

$$\Delta(x)^{\text{high}}_{\text{param}} = \sqrt{\sum_{1}^{19} h(d_i^{\text{low}}(x)) \cdot d_i^{\text{low}}(x)^2 + h(d_i^{\text{high}}(x)) \cdot d_i^{\text{high}}(x)^2} \tag{58}$$

$$\Delta(x)^{\text{low}}_{\text{param}} = \sqrt{\sum_{1}^{19} h(-d_i^{\text{low}}(x)) \cdot d_i^{\text{low}}(x)^2 + h(-d_i^{\text{high}}(x)) \cdot d_i^{\text{high}}(x)^2} \tag{59}$$

where h stands for the Heaviside function (i.e. $h(x) = 1$ when $x > 0$ and 0 otherwise). The properties of this function ensure that only positive d_i terms contribute to $\Delta(x)^{\text{high}}_{\text{param}}$ (regardless of whether they come from an increase d_i^{high} or a decrease d_i^{low} in the uncertainty source), and, at the same time, that only negative d_i terms contribute to $\Delta(x)^{\text{low}}_{\text{param}}$.
In a similar way, we can isolate the contribution to the model uncertainty of the other uncertainty sources. For the uncertainty coming from WHO reports on TB burden and treatment outcomes (uncertainty source $i = 20$), we have:

$$\Delta(x)^{\text{high}}_{\text{WHO}} = \sqrt{h(d_{20}^{\text{low}}(x)) \cdot d_{20}^{\text{low}}(x)^2 + h(d_{20}^{\text{high}}(x)) \cdot d_{20}^{\text{high}}(x)^2} \tag{60}$$

$$\Delta(x)^{\text{low}}_{\text{WHO}} = \sqrt{h(-d_{20}^{\text{low}}(x)) \cdot d_{20}^{\text{low}}(x)^2 + h(-d_{20}^{\text{high}}(x)) \cdot d_{20}^{\text{high}}(x)^2} \tag{61}$$

For the demographic prospects (u_{21}):

$$\Delta(x)^{\text{high}}_{\text{demo}} = \sqrt{h(d_{21}^{\text{low}}(x)) \cdot d_{21}^{\text{low}}(x)^2 + h(d_{21}^{\text{high}}(x)) \cdot d_{21}^{\text{high}}(x)^2} \tag{62}$$

$$\Delta(x)^{\text{low}}_{\text{demo}} = \sqrt{h(-d_{21}^{\text{low}}(x)) \cdot d_{21}^{\text{low}}(x)^2 + h(-d_{21}^{\text{high}}(x)) \cdot d_{21}^{\text{high}}(x)^2} \tag{63}$$

And, finally, for the contact matrices (u_{22}):

$$\Delta(x)^{\text{high}}_{\text{contacts}} = \sqrt{h(d_{22}^{\text{low}}(x)) \cdot d_{22}^{\text{low}}(x)^2 + h(d_{22}^{\text{high}}(x)) \cdot d_{22}^{\text{high}}(x)^2} \tag{64}$$

$$\Delta(x)^{\text{low}}_{\text{contacts}} = \sqrt{h(-d_{22}^{\text{low}}(x)) \cdot d_{22}^{\text{low}}(x)^2 + h(-d_{22}^{\text{high}}(x)) \cdot d_{22}^{\text{high}}(x)^2} \tag{65}$$

Throughout this work, coloured areas around curves, error bars or any Confidence Interval referred to an outcome of our model is calculated by summing up all the contributions:

$$\Delta(x)^{\text{low}} = \sqrt{(\Delta(x)^{\text{low}}_{\text{param}})^2 + (\Delta(x)^{\text{low}}_{\text{WHO}})^2 + (\Delta(x)^{\text{low}}_{\text{demo}})^2 + (\Delta(x)^{\text{low}}_{\text{contacts}})^2} \tag{66}$$

$$\Delta(x)^{\text{high}} = \sqrt{(\Delta(x)^{\text{high}}_{\text{param}})^2 + (\Delta(x)^{\text{high}}_{\text{WHO}})^2 + (\Delta(x)^{\text{high}}_{\text{demo}})^2 + (\Delta(x)^{\text{high}}_{\text{contacts}})^2} \tag{67}$$

In figure 2 of the main text, and figure S1 of this SI, in order to visualize the relative fraction of the total uncertainty that is due to the different 4 main uncertainty sources, we linearly weight the uncertainty error bar as follows:

- Bibliographic parameters contribution, purple area: $(\Delta(x)^{\text{low}} \cdot \frac{\Delta(x)^{\text{low}}_{\text{param}}}{\sum_y \Delta(x)^{\text{low}}_y}, \Delta(x)^{\text{high}} \cdot \frac{\Delta(x)^{\text{high}}_{\text{param}}}{\sum_y \Delta(x)^{\text{high}}_y})$
- WHO contribution, blue area: $(\Delta(x)^{\text{low}} \cdot \frac{\Delta(x)^{\text{low}}_{\text{WHO}}}{\sum_y \Delta(x)^{\text{low}}_y}, \Delta(x)^{\text{high}} \cdot \frac{\Delta(x)^{\text{high}}_{\text{WHO}}}{\sum_y \Delta(x)^{\text{high}}_y})$
- Demography contribution, orange area: $(\Delta(x)^{\text{low}} \cdot \frac{\Delta(x)^{\text{low}}_{\text{demo}}}{\sum_y \Delta(x)^{\text{low}}_y}, \Delta(x)^{\text{high}} \cdot \frac{\Delta(x)^{\text{high}}_{\text{demo}}}{\sum_y \Delta(x)^{\text{high}}_y})$
- Contacts contribution, green area: $(\Delta(x)^{\text{low}} \cdot \frac{\Delta(x)^{\text{low}}_{\text{contacts}}}{\sum_y \Delta(x)^{\text{low}}_y}, \Delta(x)^{\text{high}} \cdot \frac{\Delta(x)^{\text{high}}_{\text{contacts}}}{\sum_y \Delta(x)^{\text{high}}_y})$

The global uncertainty ranges so obtained $(x - \Delta(x)^{\text{low}}, x + \Delta(x)^{\text{high}})$, being propagated from 95% confidence intervals from the different uncertainty sources, are subsequently interpreted as 95% confidence intervals for model outcome x. When this outcome is a difference between the full and the reduced models, its significance level is estimated assuming that the outcome follows a normal distribution centered in x, with the confidence interval width $\Delta(x)^{\text{low}}$ (or $\Delta(x)^{\text{high}}$, should x be negative) defining the standard deviation ($\Delta(x)^{\text{low}} = 1.96\sigma$).

References

[1] Uplekar M, et al. (2016) Mandatory tuberculosis case notification in high tuberculosis-incidence countries: policy and practice. *Eur Respir J* pp. ERJ–00956.

[2] World Health Organization (2016) *Global tuberculosis report 2016*. (Geneva).

[3] Dye C, Garnett GP, Sleeman K, Williams BG (1998) Prospects for worldwide tuberculosis control under the who dots strategy. *Lancet* 352(9144):1886–91.

[4] Abu-Raddad LJ, et al. (2009) Epidemiological benefits of more-effective tuberculosis vaccines, drugs, and diagnostics. *Proc Natl Acad Sci USA* 106(33):13980–5.

[5] Mossong J, et al. (2008) Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med* 5(3):e74.

[6] World Health Organization (2016) Tuberculosis database. http://www.who.int/tb/country/en/index.html (accessed November 2016).

[7] Marais B, et al. (2004) The natural history of childhood intra-thoracic tuberculosis: a critical review of literature from the pre-chemotherapy era [state of the art]. *Int J Tuber Lung Dis* 8(4):392–402.

[8] Donald PR, Marais BJ, Barry III CE (2010) Age and the epidemiology and pathogenesis of tuberculosis.

[9] Dodd PJ, Gardiner E, Coghlan R, Seddon JA (2014) Burden of childhood tuberculosis in 22 high-burden countries: a mathematical modelling study. *Lancet Glob Health* 2(8):e453–e459.

[10] Dodd PJ, Sismanidis C, Seddon JA (2016) Global burden of drug-resistant tuberculosis in children: a mathematical modelling study. *Lancet Infect Dis* 16(10):1193–1201.

[11] Cruz AT, Starke JR (2007) Clinical manifestations of tuberculosis in children. *Paediatr Respir Rev* 8(2):107–117.

[12] Marais BJ, et al. (2006) Childhood pulmonary tuberculosis: old wisdom and new challenges. *American journal of respiratory and critical care medicine* 173(10):1078–1090.

[13] Nelson L, Wells C (2004) Global epidemiology of childhood tuberculosis [childhood tb]. *Int J Tuberc Lung Dis* 8(5):636–47.

[14] Korenromp EL, Scano F, Williams BG, Dye C, Nunn P (2003) Effects of human immunodeficiency virus infection on recurrence of tuberculosis after rifampin-based treatment: an analytical review. *Clin Infect Dis* 37(1):101–12.

[15] Picon PD, et al. (2007) Risk factors for recurrence of tuberculosis. *J Bras Pneum* 33(5):572–8.

[16] Lee RS, Proulx JF, Menzies D, Behr MA (2016) Progression to tuberculosis disease increases with multiple exposures. *Eur Respir J* pp. ERJ–00893.

[17] Pillay T, Khan M, Moodley J, Adhikari M, Coovadia H (2004) Perinatal tuberculosis and HIV-1: considerations for resource-limited settings. *Lancet Inf Dis* 4(3):155–65.

[18] Del Valle SY, Hyman JM, Hethcote HW, Eubank SG (2007) Mixing patterns between age groups in social networks. *Soc Networks* 29(4):539–554.

[19] Miller E, et al. (2010) Incidence of 2009 pandemic influenza A H1N1 infection in England: a cross-sectional serological study. *Lancet* 375(9720):1100–8.

[20] Birrell PJ, et al. (2011) Bayesian modeling to unmask and predict influenza A/H1N1pdm dynamics in London. *Proc Natl Acad Sci USA* 108(45):18238–43.

[21] Guzzetta G, et al. (2011) Modeling socio-demography to capture tuberculosis transmission dynamics in a low burden setting. *J Theor Biol* 289:197–205.

[22] Kiti MC, et al. (2014) Quantifying age-related rates of social contact using diaries in a rural coastal population of Kenya. *PloS one* 9(8):e104786.

[23] Melegaro A, et al. (2017) Social contact structures and time use patterns in the Manicaland province of Zimbabwe. *PloS one* 12(1):e0170459.

[24] le Polain de Waroux O, et al. (2017) Characteristics of human encounters and social mixing patterns relevant to infectious diseases spread by close contact: A survey in southwest Uganda. *bioRxiv* p. 121665.

[25] Read JM, et al. (2014) Social mixing patterns in rural and urban areas of southern China. *Proc R Soc Lond B Biol Sci* 281(1785):20140268.

[26] Ibuka Y, et al. (2015) Social contacts, vaccination decisions and influenza in Japan. *J Epidemiol Community Health* pp. jech–2015.

[27] UN (2016) Population division database. http://esa.un.org/unpd/wpp/index.htm (accessed November 2016).

[28] Millen SJ, Uys PW, Hargrove J, Van Helden PD, Williams BG (2008) The effect of diagnostic delays on the drop-out rate and the total delay to diagnosis of tuberculosis. *PLoS One* 3(4):e1933.

[29] Moré JJ (1978) The Levenberg-Marquardt algorithm: implementation and theory in *Numerical analysis*. (Springer), pp. 105–16.