A data-driven model for the assessment of age-dependent patterns of Tuberculosis burden and impact evaluation of novel vaccines.

Sergio Arregui *[1,2], Joaquín Sanz[3,4], Dessislava Marinova[5,6], María José Iglesias[5,6], Sofía Samper[6,7], Carlos Martín[5,6,8], Yamir Moreno[1,2,9]

[1] Institute for Biocomputation and Physics of Complex Systems (BIFI), University of Zaragoza, Spain.
[2] Department of Theoretical Physics, University of Zaragoza, Spain.
[3] Sainte-Justine Hospital Research Centre, Montreal, Canada.
[4] Department of Biochemistry, University of Montreal, Canada.
[5] Department of Microbiology, Faculty of Medicine, University of Zaragoza, Spain.
[6] CIBER Enfermedades Respiratorias, Instituto de Salud Carlos III, Madrid, Spain.
[7] Instituto Aragonés de Ciencias de la Salud. IIS Aragon.
[8] Service of Microbiology, Miguel Servet Hospital. IIS Aragon.
[9] Complex Networks and Systems Lagrange Lab, Institute for Scientific Interchange, Turin, Italy.

Summary

Background

Among all possible novel epidemiological interventions against tuberculosis (TB), improved preventive vaccines hold the promise of offering substantial reductions of TB burden worldwide. Accordingly, several vaccine candidates are currently under development, each of which, depending on their immunogenic properties, might show differences in protection when applied to different age groups.

Methods

We present a TB epidemiological model which, capitalizing on publicly available data from different sources (World Health Organization (WHO), United Nations (UN) Population Division and published studies on contact-patterns surveys at an international scale), formalizes a data-driven description of most relevant coupling mechanisms between populations' age structure and TB dynamics.

Findings

The global demographic shift projected by the UN for the next decades is to be accompanied, according to our results, by a shift in the age-distribution of TB burden in some of the regions most affected by the disease. We show that for Africa and the East Mediterranean this could lead to revert the projected global decline of TB unless novel epidemiological measures are deployed. Regarding the comparison of different vaccination strategies, an adolescent-focused global immunization campaign appears to be more impactful than newborn vaccination in the short term.

Interpretation

We demonstrated that TB indicators and vaccination strategies remarkably depend on how the disease dynamics is coupled to the demographic structure of the population.

Capitalizing on a data-driven approach, we identified substantial biases in epidemiological forecasts that are rooted on an inadequate description of age-dependent mechanisms, among others. Our findings provide fundamental insights if novel age-focused epidemiological interventions, such as preventive vaccines, are to be considered and established.

Funding

Work partly funded by the Government of Aragón, the MINECO (Spain) and the European Commission.

1. Introduction

The control of TB is one of the largest endeavors of public health authorities ever since the bacterium that causes it –*Mycobacterium tuberculosis*– was discovered. Recently, the development of global strategies for diagnosis and treatment optimization have led to TB burden decay worldwide,[1] to the point that the End TB Strategy has allowed the scientific community to think that its eradication by 2035 is possible.[2] Nonetheless, such goal is yet far away, and TB still represents a major Public Health problem,[3–5] being responsible for 1.8 million deaths worldwide in 2015,[6] which evidences the need of new epidemiological measures and pharmacological resources.

Among all possible interventions, the deployment of new preventive vaccines is currently thought to offer the highest, most immediate impact on disease burden reduction. The old BCG vaccine has been proven inconsistently protective,[1] with the exception of its well-documented role at preventing most severe forms of meningeal and miliary TB in infants.[7] Accordingly, today there are thirteen novel experimental vaccine candidates in clinical trials,[6] designed to substitute BCG or to boost immunogenicity and protection in BCG vaccinated individuals.[8]

In the following years, the clinical trials of some of the candidate vaccines will have finished, and results regarding their efficacy will be available. Modeling techniques will be crucial at that stage, as they will allow forecasting and comparing their expected impacts and cost-effectiveness.[9,10] That will provide guidelines to policy makers on fundamental aspects like the age of the population to target, which is expected to differ for some of the vaccines.[11] While newborns are the primary targets of BCG-substitutive vaccines, adolescents constitute the prototypical target for BCG-boosters.[12]

It is thus imperative that the clinical work be followed by a modeling effort aimed at creating tools to produce vaccine impact forecasts as reliable as possible.[13] Since the age of vaccination target populations is one of the main questions to discuss, these novel modeling tools will have to provide a thorough description of the main mechanisms through which the disease cycle depends on population age structure.

In this sense, key, well-documented aspects of the natural history of TB that are known to depend on age are the reduced probability of fast progression of infants when

compared to adults,[14] and the variations of the odds of developing pulmonary vs extra-pulmonary TB.[15] Nonetheless, it is not clear up to which point these age-dependent parameters are subject to further sources of heterogeneity, which often renders their treatment in TB models a difficult decision.[16] Similarly, other questions are commonly simplified in the TB modeling literature. That is the case of contact rates among individuals of different ages (typically assumed to be age-independent) or the time evolution of the demographic structure of the populations being modeled (see supplementary figure S2), whose influence is traditionally neglected.

Here, we propose a model that overcomes such limitations by providing a data-driven description of the mentioned coupling mechanisms between TB dynamics and population age-structure.

2. Methods

Model Structure and Disease Transmission.

Our model capitalizes on previous works by C. Dye and colleagues,[17,18] on which several new ingredients have been incorporated to avoid biases critically dependent on population age-structure. Specifically, we deal with a compartmental, ordinary differential equation-based, age-structured model of TB in which we consider two different types of individuals, –susceptible and vaccinated–, two different latency paths to disease –fast and slow– and six different situations of disease, depending on its aetiology -non pulmonary, pulmonary (smear positive/negative)- and on whether sick individuals are being treated or not. After disease, we explicitly consider the main treatment outcomes contemplated by 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, smear progression and mother-child transmission,[6,17–20] are sketched in Figure 1A, and thoroughly detailed in the Supplementary Appendix, section 1.

Using our model, we reproduce past TB burden levels and age-distributions reported by the WHO,[15] produce projections until 2050 and evaluate different vaccination schemes in the five regions that are currently affected by the highest TB burden levels world-wide: Africa high HIV prevalence (AFRH), Africa low HIV prevalence (AFRL), Eastern Mediterranean Region (EMR), South-Southeast Asia Region (SEAR) and Western Pacific Region (WPR) (Figure 1B).

Model Calibration.

Model parameters have been obtained from bibliographical sources except the infectiousness, the diagnosis rate and the probability of fast-progression to active disease in each region. The first two parameters, along with the initial conditions of the system (initial prevalence levels of latency and disease) are calibrated so as to fit aggregated incidence and mortality series reported by WHO from 2000 to 2015.[21,22] As

in Abu-Raddad et al.[18], infectiousness and diagnosis rates are assumed to vary over time, so as to capture changes in public health policies, overall socioeconomic conditions, etc. Finally, the age distribution of TB incidence is calibrated by fitting the probabilities of fast-progression in each age group and region. For further details on the calibration procedure, see supplementary appendix, section 2 and figure S3.

Dependences on Age Structure.

Regarding the age structure of the population, the model takes as an explicit input the measured/expected evolution of the demographic structures of each region for the interval 2000-2050, as reported by the UN population division (Figure 1C).[22] Forcing our system to evolve according such projections allows us, unlike previous approaches ,[17,18] to explicitly evaluate the influence of population evolution on the disease burden without making any further demographic assumption.

Similarly, capitalizing on detailed data about age distribution of TB incidence and disease types reported in WHO Database (period analyzed: 2006-2012),[15] we implemented a calibration procedure that unlocked a controlled reproduction of the observed distributions of incidence rates and disease types across age. According to our approach, probabilities of progressing to each type of disease (pulmonary smear positive, smear negative, or extra-pulmonary TB) are directly estimated from country-wise WHO reports, while fast-progression fractions (one of the most relevant Natural-History parameters, yet subject to large age-wise and region-wise heterogeneities),[14,16,23–25] are fitted to reproduce region-specific age-distributions of incidence rates.

Finally, we abandon another classical hypothesis in the modeling of the spread of TB that constitutes a simplification of transmission pathways: the homogeneity of the contact patterns between the different age groups. In this sense, empirical studies have shown that there is a remarkable heterogeneity in the distribution of contacts between age strata of a given population, with real age-mixing patterns being strongly assortative and relatively robust across different countries;[26] an observation of deep epidemiological implications.[27] Capitalizing on these findings, we have incorporated in the model a survey-based, more plausible and realistic heterogeneous contact structure on top of which we model the spreading of the disease (Figure 1D).

For further model details, including definitions of model states, parameters, uncertainty and sensitivity analyses, the reader is referred to sections 3-4 of the Supplementary Appendix, and to supplementary figures S4-S9.

3. Results

Once calibrated, the model was used to estimate the evolution of the incidence and mortality of TB up to 2050 in the five regions analyzed in this study (Figure 2A). Comparing the projections for 2050 to the situation in 2015 (Figure 2B), we found that TB incidence at the end of the simulated period will be marginally higher in AFRH (relative increase 10·3%; 95% CI (-62·5,98·3)), significantly higher in AFRL (81·4%; 95%

CI (55·2,99·0)), and EMR (143·3%; 95% CI (23·1,255·8)); whereas it is estimated to significantly decrease in SEAR (relative decrease -25·2%; 95% CI (-30·6,-19·9)) and WPR (-50·2%; 95% CI (-63·0,-32·0)).

Then, we investigated what are the drivers and main underlying features causing the predicted variations in TB burden levels. To this end, we have repeated our incidence projections using three reduced versions of our model where each one of the three main age-dependent model features -demographic evolution (Reduced model 1), calibration of age/TB type distributions (Reduced model 2) and heterogeneous contact patterns (Reduced model 3)- is removed at a time (Figure 3A). Remarkably enough, in the regions where a relapse in TB burden is foreseen (AFRH, AFRL and EMR), the projected trend shift disappears when the demographic structure is considered to be constant (Reduced model 1), evidencing the primary impact of a proper description of the demographic evolution of the populations on models' forecasts.

Beyond analyzing aggregated burden levels over all age groups, in Figure 3B, we represent the age distribution of incidence rates for the EMR area at different time-points. Aiming at isolating the effects of the age-coupling mechanisms on these measurements, we represent in each column the distributions corresponding to the projected incidence derived from the complete model (left, red bars) versus the equivalent distributions forecasted by each of the three reduced models (right purple bars). For further details about these reduced models and their effects on different model outcomes and regions, see Supplementary Appendix, section 5, and figures S10-S14.

The effect of the demographic evolution on age-distributed incidence forecasts is reflected in the first column, where we see that the most relevant differences arise later in time, specially affecting the eldest age groups. This phenomenon arises mainly as a consequence of a large increase of the infection prevalence in the eldest strata of the population (EMR, 2050: infection prevalence among individuals older than 45: Reduced Model 1: 62·4%; Complete Model: 82·0%; difference 19·6%; 95% CI: (2·0,33·4)), which, along with the global aging of the population, yields a global increase of the infection reservoir (EMR, 2050: total infection prevalence: Reduced Model 1: 27·4%; Complete Model: 56·8%; difference 29·4% 95% CI (11·3,40·7), see supplementary figure S14).

Second, our calibration approach yields a controlled description of the disease burden distribution across age groups, as we can see in the second column of Figure 3B. Comparing the complete model's description at 2010 to that provided by reduced model 2, where fast progression and disease type probabilities are taken from bibliographical estimates instead of being calibrated,[18] we see how this ingredient is needed to reproduce the observed distribution of incidence across age (Figure 3B turquoise symbols in 2010 panels; see Supplementary Appendix, section 5.2.4, for a formal comparison between models and observed distributions in the whole calibrated period). Importantly, these differences are propagated later in time in models' predictions, which underlines the relevance of producing an adequate description of present distributions to produce forecasts and evaluate future vaccines.

Finally, we evaluate the effects of contact patterns heterogeneity, which are more modest than those of other model features. As we see in the third column of figure 3B, incidence levels foreseen by the complete model are similar to those produced by reduced model 3, with the exception of a slight increase in the estimated incidence rates among the eldest groups (EMR, 2050, incidence among individuals older than 60 years: Reduced Model 3: 4342; Complete Model: 5472 annual cases per million, difference: 21 %, 95% CI (1-24)). Similarly, the difference in the fraction of infectious individuals among the eldest age groups is larger in the Complete Model, but significant (EMR, 2050, prevalence fraction of individuals older than 60 years that are infectious: Reduced Model 3: 0·18%; Complete Model: 0·21%, relative difference: 13%, 95% CI (1-15), supplementary figure S14).

The model here presented is the best modeling tool to analyze the effectiveness of different vaccination strategies. To illustrate this question, we modeled two different immunization scenarios where two vaccines with an 80% of efficacy at reducing TB infectiousness are to be deployed on the population in 2025, targeting either newborns, or 15 years old individuals. (See supplementary Appendix, section 6).

Figure 4 shows results obtained for the projected evolution of the incidence rates for each case. Vaccinating adolescents provides the best results, with a faster reduction in incidence rates in all regions. The relative variations in impact when comparing both vaccination strategies (measured as relative variation of total number of cases prevented in 2025-2050) is 53% in AFRH CI (45-59), 57% in AFRL CI (49-62), 62% in EMR CI (56-68), 63% in SEAR CI (55-72), and 81% in WPR CI (76-85). The associated raw impacts foreseen by each vaccination strategy are represented in supplementary figure S15. These differences, which are significantly affected by the inclusion in our model of the mentioned new features (see supplementary figure S16), highlight the importance of achieving a prompt immunization of adolescents and young adults, as it is expected to produce a faster reduction in morbidity and TB-related mortality rates world-wide.[28]

4. Discussion

In this work, we presented a framework for modeling the spreading of TB that accounts for the coupling of the disease dynamics with an evolving demographic structure. Moreover, our model adequately describes the burden distribution across age groups and considers the effects of heterogeneous contact patterns among different age strata. The introduction of those aspects, from a data-driven stand, exerts strong effects on model's forecasts for the years to come. Remarkably enough, taking into explicit consideration the demographic transition foreseen for the next decades by the UN population division in the areas of the world that are currently severely hit by the disease, leads to a shift in age-dependent TB trends. Our model describes how the disease is expected to cease to be prominently associated to young adults to progressively affect elder age segments in several of the low-income populations under study, more and more intensely as the XXIth century unfolds. This previously unnoticed consequence, which gets aggravated when considering

assortative age mixing patterns, might translate into a worrisome unforeseen relapse in incidence and mortality in some parts of the world, notably Africa and Middle East, unless novel epidemiological measures are undertaken.

Regarding the use of our model to compare different vaccination strategies, we found that a vaccine targeting adolescents is expected to have a faster effect in reducing TB burden than an equivalent intervention focused on newborns. This result, which is in line of previous work in the field,[28] must however be considered dependent on, at least, three key considerations which goes beyond the scope of this work. First, when implementing vaccination campaigns, high levels of coverage are easier to achieve when newborns are the vaccination target than in the case of adolescents; which might render any comparison assuming equal levels of coverage –as the ones undertaken here-, unrealistic. Secondly, environmental sensitization to mycobacterial antigens can affect the efficacy of a vaccine, inhibiting its performance if applied in adolescents and adults. This phenomenon, known as "blocking", has been documented and quantified for BCG.[29] Finally, while immunized newborns will eventually reach all age groups, a vaccination strategy focused exclusively on adolescents will always leave in children an unprotected reservoir of individuals, specially where BCG has been considered ineffective.[30] This makes an adolescent-focused vaccination strategy more useful for achieving a faster effect in reducing TB burden, but more problematic in the long term, when the existence of such un-protected reservoir will constitute an impediment for achieving further reductions in TB levels as a necessary step towards its eventual eradication.

Further studies to explore these issues, quantifying the way in which novel vaccines' impact evaluations might depend on those aspects, as well as how the demographic evolution and age-wise distributions of TB burden and contact patterns can affect such evaluations, must therefore be undertaken.

Contributors.

JS, CM and YM conceived the study. SA, JS, DM, MJI, SS, CM and YM participated in the development of the model. SA and JS implemented the model. SA performed the numerical simulations and statistical analyses. SA, JS, and YM analyzed the results and wrote the paper with contributions from DM, MJI, SS and CM. SA, JS and YM prepared the Supplementary Appendix, with contributions from DM, MJI, SS, and CM. All authors approved the final version.

Declaration of interests.

CM is a co-inventor on a composition of matter patent: Title: Tuberculosis Vaccine, Owner entity: Universidad de Zaragoza, Request number: PCT/ES 2007/070051.

Acknowledgements.

SA was supported by the FPI program of the Government of Aragón, Spain, and JS by the postdoctoral training program for non-residents of Quebec from the *Fonds de*

recherche du Québec - Santé (FRQS). This work was partially supported by "*Gobierno de Aragón/Fondo Social Europeo*", by MINECO and FEDER funds through grants FIS2014-55867-P and BIO2014-52580P, by Project TBVAC2020 (643381) funded by the European Commission H2020 to CM and DM, by the European FP7 grant NEWTBVAC 241745 and by the EC Proactive project MULTIPLEX (contract no. 317532) to YM. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We also thank M. Gutierrez for assistance with figures.

References.

1 Borgdorff MW, Floyd K, Broekmans JF. Interventions to reduce tuberculosis mortality and transmission in low- and middle- income countries. *Bulletin of the World Health Organization* 2002; **80(3)**: 217–27.

2 Lienhardt C, Glaziou P, Uplekar M, Lonnroth K, Getahun H , Raviglione M. Cost-effectiveness of novel vaccines for tuberculosis control: a decision analysis study. *BMC Public Health* 2011; **11**:55.

3 Dye C, Williams BG, Williams S. Criteria for the control of drug-resistant tuberculosis. *Proc Natl Acad Sci USA* 2000; **97(14)**: 8180–5 DOI 10.1073/pnas.140102797.

4 Boily MC, Lowndes C, Alary M. The impact of HIV epidemic phases on the effectiveness of core group interventions: insights from mathematical models. *Sexually Transmitted Infections* 2002. **78(suppl 1)**;i78–90 DOI 10.1136/sti.78.suppl_1.i78.

5 Korenromp EL, Scano F, Williams BG, Dye C, Nunn P. Effects of human immunodeficiency virus infection on recurrence of tuberculosis after rifampin- based treatment: an analytical review. *Clin Infect Dis* 2003; **37(1)**: 101–12 DOI 10.1086/375220.

6 WHO. Global Tuberculosis Report. Geneva: WHO Press; 2016. Available at http://www.who.int/tb/publications/global_report/en/.

7 Mangtani P, Abubakar I, Ariti C et al. Protection by BCG vaccine against tuberculosis: a systematic review of randomized controled trials. *Clin Infect Dis* 2014; **58(4)**: 470–80 DOI 10.1093/cid/cit790.

8 Marinova D, Gonzalo-Asensio J, Aguilo N, Martín C. Recent development in tuberculosis vaccines. *Expert Rev Vaccines* 2013; **12(12):** 1431–48 DOI 10.1586/14760584.2013.856765.

9 Dye C, Fine PE. A major event for new tuberculosis vaccines. *Lancet* 2013; **381(9871):** 972–4.

10 Orme IM. Vaccine Development for Tuberculosis: Current Progress. *Drugs* 2013; **73 (10)**: 1015–24.

11 Kaufmann SH, Hussey G, Lambert PH. New vaccines for tuberculosis. *Lancet* 2010; **375(9731)**: 2110–9.

12 Tseng CL, Oxlade O, Menzies D, Aspler A, Schwartzman K. Cost-effectiveness of novel vaccines for tuberculosis control: a decision analysis study. *BMC Public Health* 2011; **11**: 55.

13 Pandemic Prediction and Forecasting Science and Technology Working Group of the National Science and Technology Council of the USA. Towards epidemic prediction: federal efforts and opportunities in outbreak modeling. https://www.whitehouse.gov/sites/default/files/microsites/ostp/NSTC/towards_epidemic_prediction-federal_efforts_and_opportunities.pdf (accessed January 9th, 2017)

14 Vynnycky E, Fine PEM. The natural history of tuberculosis: the implications of age-dependent risks of disease and the role of reinfection. *Epidemiol infect* 1997; **119(02):** 183-201.

15 WHO. Tuberculosis Database. http://www.who.int/tb/country/en/index.html (accesed November 2016)

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

17 Dye C, Garnett GP, Sleeman K, Williams BG. Prospects for worldwide tuberculosis control under the WHO DOTS strategy. *Lancet* 1998; **352:** 1886–91.

18 Abu-Raddad LJ, Sabatelli L, Achterberg JT et al. Epidemiological benefits of more effective tuberculosis vaccines, drugs and diagnostics. *Prod Nat Acad Sci USA* 2009; **106(33):** 13980–5.

19 Picon PD, Bassanesi SL, Caramori ML, Ferreira RL, Jarczewski CA, Vieira PR. Risk factors for recurrence of tuberculosis. *J Bras Pneumol* 2007; **33(5):** 572–8.

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

21 Dye C, Scheele S, Dolin P, Pathania V, Raviglione MC. Global burden of tuberculosis. Estimated incidence, prevalence, and mortality by country. *J Am Med Asoc* 1999; **282(7):** 677–86.

22 United Nations. Population Division Database http://esa.un.org/unpd/wpp/index.htm (accessed November 2016)

23 Murthy VK, Nair SS, Gothi GD, Chakraborty AK. Incidence of tuberculosis among newly infected population and in relation to the duration of infected status. *Ind. J. Tub* 1976; **23**: 3–7.

24 Sutherland I, Švandová E, Radhakrishna S. The development of clinical tuberculosis following infection with tubercle bacilli: 1. A theoretical model for the development of clinical tuberculosis following infection, linking from data on the risk of tuberculous infection and the incidence of clinical tuberculosis in the Netherlands. *Tubercle* 1983; ***63*(4)**: 255–68.

25 Lee RS, Proulx JF, Menzies D, Behr MA. Progression to tuberculosis disease increases with multiple exposures, *Eur Respir J* 2016; ERJ-00893.

26 Mossong J, Hens N, Jit M et al. Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases, *PLoS Med* 2008; **5(3):** e74.

27 Miller E, Hoschler K, Hardelid P, Stanford E, Andrews N, Zambon M. Incidence of 2009 pandemic influenza A H1N1 infection in England: a cross-sectional serological study. *Lancet* 2010; **375(9720):** 1100–8 .

28 Knight GM, Griffiths UK, Sumner T et al. Impact and cost-effectiveness of new tuberculosis vaccines in low- and middle-income countries. *Prod Nat Acad Sci USA* 2009; **111(43):** 15520–5

29 Arregui S, Sanz J, Marinova D, Martín C, Moreno Y, On the impact of Masking and Blocking Hypotheses for measuring efficacy of new tuberculosis vaccines. *PeerJ* 2016; **4**: e1513.

30 Kaufmann SH, Evans TG, Hanekom WA. Tuberculosis vaccines: Time for a global strategy. *Sci Transl Med* 2015; **7(276)**: 276fs8.

FIGURES

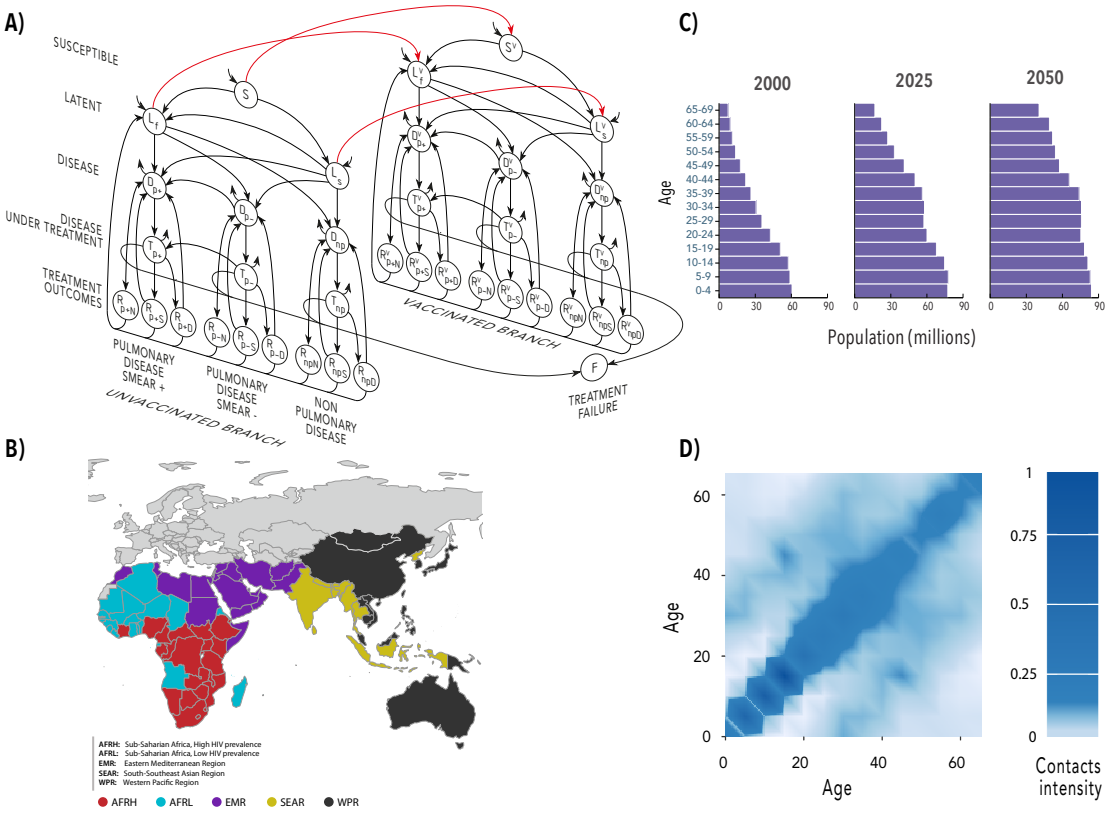

Figure 1. Panel A. Natural History scheme of the TB spreading model. S: susceptible. L: latent. D: (untreated) disease, T (treated) disease, R recovered, F: failed recovery, Super-index v: States of the vaccinated branch. Types of TB considered: p+: Pulmonary Smear-Positive, p-: Pulmonary Smear-Negative, np: Non-pulmonary. Treatment outcomes: N: Natural recovery, S: Successful treatment, D: Default (abandon of treatment), F: treatment failure. For a detailed description of the dynamical variables, transitions and mathematical formulation, see the supplementary appendix, section 1. Panel B. WHO-defined regions analyzed in this study: Africa high HIV prevalence (AFRH), Africa low HIV prevalence (AFRL), Eastern Mediterranean Region (EMR), South-Southeast Asia Region (SEAR) and Western Pacific Region (WPR). These are the five regions with higher current TB burden levels.[15] Panel 3: Prospected evolution of the demographic structure in EMR, as projected by the UN population division.[22] Contact patterns among age groups used in this study. Color indicates the relative intensity of the contacts between age-strata. Data adapted from Mossong et al.[26]

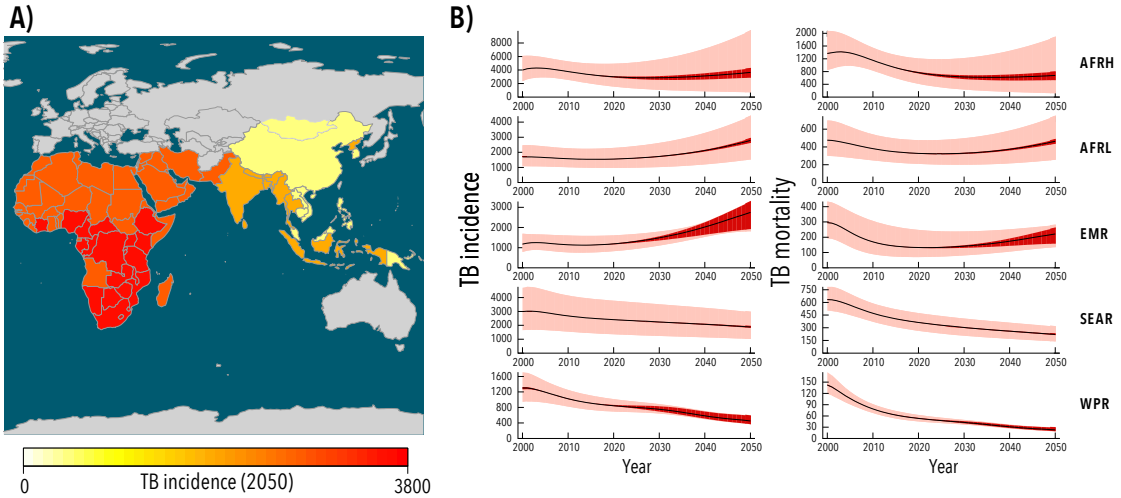

Figure 2. Panel A. Incidence levels in each WHO region projected in 2050 (new TB cases per million per year). Panel B. Projected time evolution of incidence (left) and mortality rates (right, TB-related deaths per million per year) in the five regions analyzed. Uncertainty stripes are propagated from parameters uncertainties (red) and WHO current TB burden estimates (salmon). See supplementary figures S6-S7 for further details on uncertainty and sensitivity analyses.

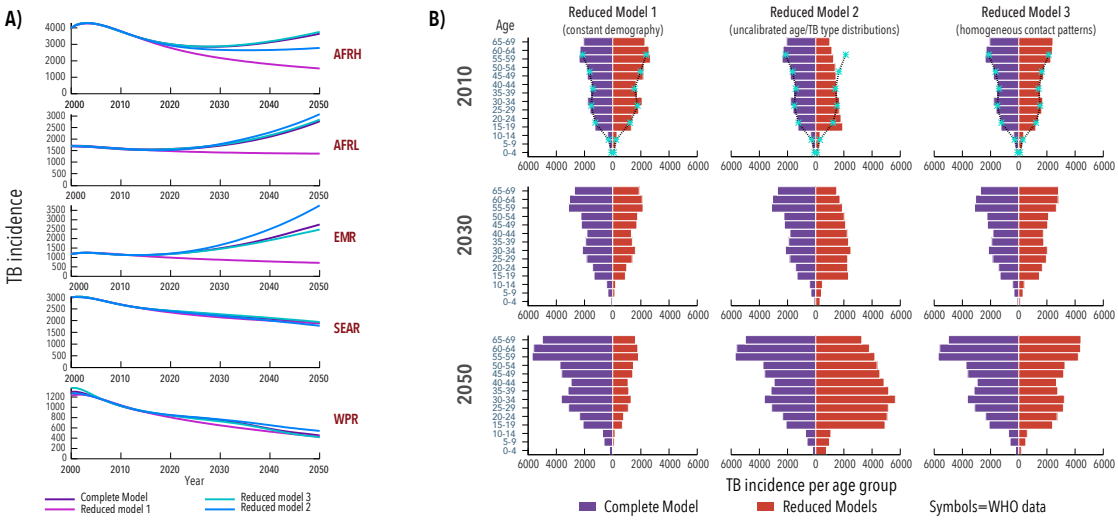

Figure 3. Panel A. Influence of the mechanisms of coupling between age-structure and TB dynamics on incidence rates prospects in all regions. Reduced model 1: the demographic pyramids are considered constant during the period modeled. Reduced model 2: Detailed calibration of age and TB type distributions is not implemented. Reduced model 3: contact patterns among age groups are considered to be homogeneous. Panel B: distribution of incidence rates per age group in the different reduced models in different time-points. In 2010, turquoise symbols represent the data reported by the WHO.[15]

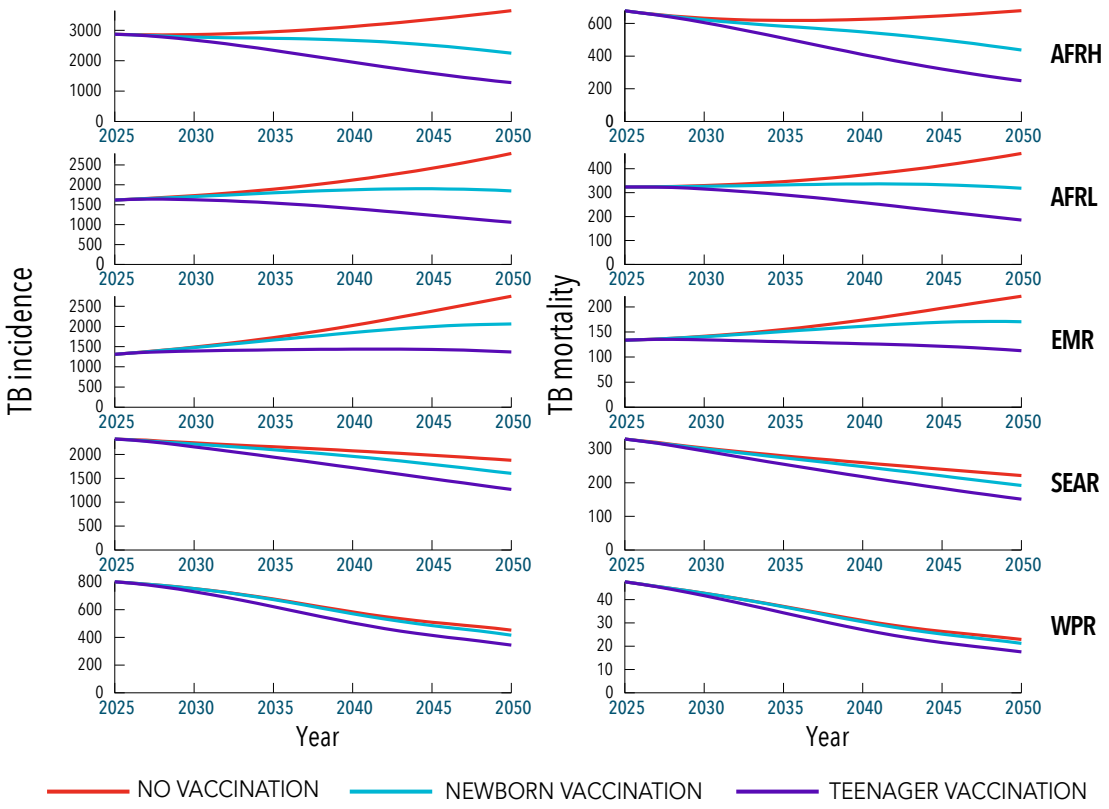

Figure 4. Prospected effects on incidence and mortality rates by newborn and adolescent focused vaccines (15 years old), with an ideal 100% coverage, up to 2050. The hypothetical vaccines being modeled are supposed to be identical, having an 80% protective effect against infection.

A data-driven model for the assessment of age-dependent patterns of Tuberculosis burden and impact evaluation of novel vaccines. Supplementary Appendix

Sergio Arregui[1,2], Joaquín Sanz[3,4], Dessisslava Marinova[5,6], María José Iglesias[5,6], Sofía Samper[6,7], Carlos Martín[5,6,8], and Yamir Moreno[1,2,9]

[1]Institute for Biocomputation and Physics of Complex Systems (BIFI), University of Zaragoza, Spain
[2]Department of Theoretical Physics, University of Zaragoza, Spain
[3]Saint-Justine Hospital Research Center, Montreal, Canada
[4]Department of Biochemistry,University of Montreal, Canada
[5]Department of Microbiology, Faculty of Medicine, University of Zaragoza, Spain
[6]CIBER Enfermedades Respiratorias, Instituto de Salud Carlos III, Madrid, Spain
[7]Instituto Aragonés de Ciencias de la Salud. IIS Aragon
[8]Service of Microbiology, Miguel Servet Hospital, Aragón, Spain
[9]Complex Networks and Systems Lagrange Lab, Institute for Scientific Interchange, Turin, Italy

Contents

1 Model description: technical details **2**
1.1 Natural history of the disease . . . 2
1.1.1 Primary Tuberculosis infection . . . 2
1.1.2 Progression from latency to (untreated) disease . . . 4
1.1.3 Death of untreated individuals . . . 5
1.1.4 TB diagnosis and treatment . . . 5
1.1.5 Treatment outcomes . . . 6
1.1.6 Natural recovery . . . 8
1.1.7 Endogenous reactivations after treatment or natural recovery . . . 9
1.1.8 Exogenous reinfection of infected individuals . . . 10
1.1.9 Smear progression . . . 12
1.1.10 Newborn vaccination. Mother-child disease transmission . . . 12
1.2 Force of infection . . . 13
1.3 Contact patterns . . . 13
1.4 Aging and age-specific vaccination . . . 15
1.5 Demographic evolution . . . 15
1.6 Ordinary differential equations system . . . 18
1.7 Initial conditions setup . . . 22

2 Model calibration procedure **22**
2.1 Global reproduction of aggregated TB burden levels . . . 22
2.2 Detailed calibration of distributions of incident cases among age-groups and types of TB . . . 23
2.2.1 Fitting of fast-progression probabilities per age-group to reproduce TB incidence distribution across age groups . . . 23
2.2.2 Estimation of age-dependent probabilities of TB types from case notifications data . . . 25
2.3 Calibration algorithm summary . . . 26

3 Model states and parameters summary **27**
3.1 Dynamic states . . . 28
3.2 Global parameters . . . 29
3.3 Age-independent regional parameters . . . 29
3.4 Diagnosis rate, infectiousness and initial conditions . . . 30

3.5 Fast progression probabilities . . . 31
3.6 Probabilities of developing each type of TB . . . 32
3.7 Vaccine descriptors . . . 32

4 Model uncertainty and sensitivity analysis **33**
4.1 Uncertainty sources analysis . . . 33
4.2 Model sensitivity . . . 35

5 Model forecasts: basal scenario (no intervention) **36**
5.1 Outcomes of the complete model . . . 36
5.1.1 Aggregated burden . . . 36
5.1.2 Age distribution of incidence rates . . . 36
5.2 Influence of model features . . . 39
5.2.1 Exposition of model features and reduced models . . . 39
5.2.2 Effects on TB aggregated burden . . . 40
5.2.3 Effects on fitted parameters: diagnosis rates and infectiousness . . . 41
5.2.4 Effects on age distributions of TB burden . . . 43

6 Model forecasts: vaccine impact evaluations **47**
6.1 Outcomes from the complete model . . . 47
6.2 Influence of model features on impact evaluations of novel vaccines. . . . 47

7 Conclusion: model caveats and lines of future work **49**

1 Model description: technical details

1.1 Natural history of the disease

Our work is essentially based on previous models by C. Dye and colleagues,[1,2] on which new ingredients – heterogeneous contact patterns,[3] an explicit coupling of the demographical evolution and the disease dynamics and a calibration procedure aimed at reproducing TB incidence distributions across age and disease type– have been incorporated in order to avoid systematic biases that affect certain model predictions and outcomes critically dependent on the age structure of the population. The natural history scheme has also been refined so as to render it more suited to the definitions by WHO, mostly in what regards to treatment outcomes.

Summarizing, we deal with an ordinary differential-equations based, age structured model of TB infection in which we consider two different classes of unexposed individuals –susceptible and vaccinated–, 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 consider explicitly the main treatment outcomes contemplated by WHO data schemes: treatment completion, default, failure and death.

The model presents two branches: the non vaccinated and the vaccinated branch, to which individuals retaining the immunogenic effect of the vaccine belong to. Each state X in the unvaccinated branch has its homologue X^v in the vaccinated branch. For example, at time step t, non vaccinated susceptible individuals, unexposed to TB infection are denoted by $S(a,t)$, while vaccinated susceptible individuals are $S^v(a,t)$. The integer $a \in [0, 13]$, is the index representing any of the fourteen age groups which individuals belong to, each of them covering 5 years; which yields the description of a demographic structure up to 70 years old.

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 S1 for the sake of clarity.

1.1.1 Primary Tuberculosis infection

We call primary the infection of an individual who was not previously exposed to the bacterium: i.e. individuals of classes S and S^v. 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 unvaccinated 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 1.2.

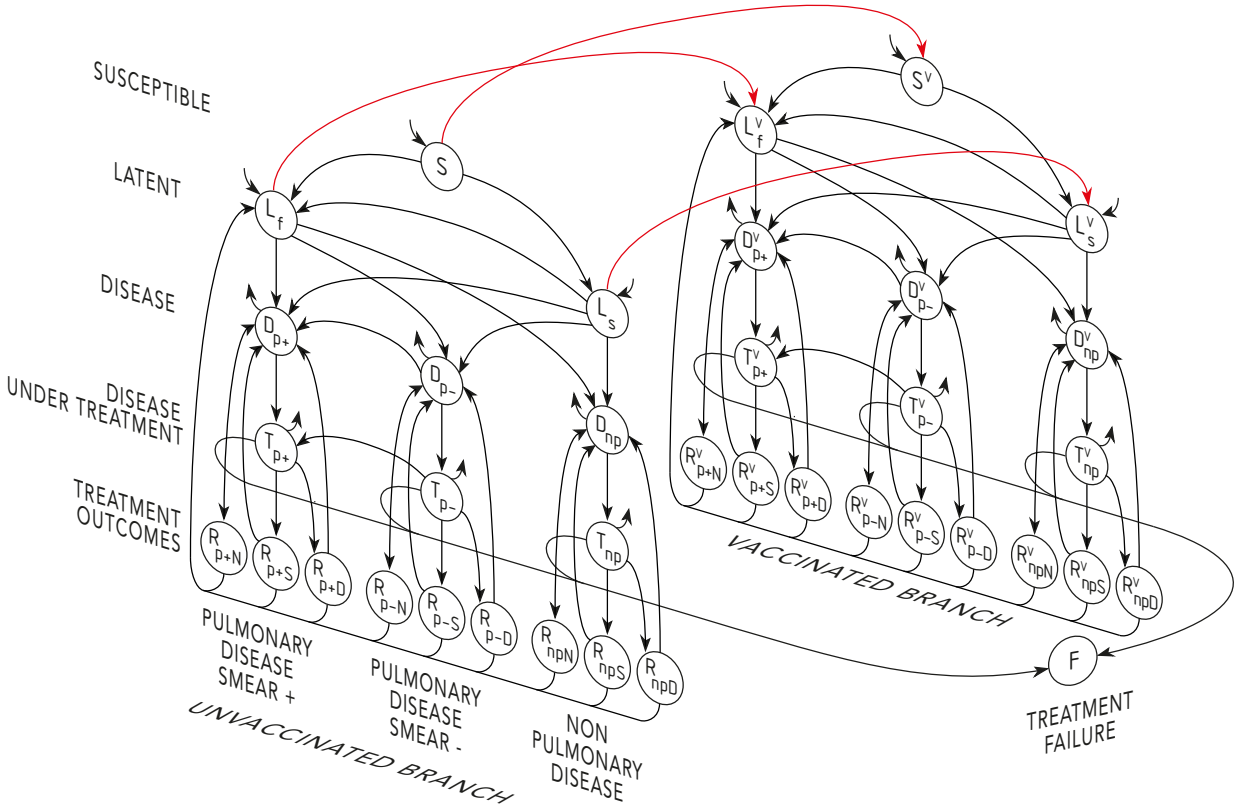

Figure S1: Natural history of the disease: S: susceptible. L: latent. D: (untreated) disease, T (treated) disease, R recovered, F: failed recovery, Super-index v: States of the vaccinated branch. Types of TB considered: p+: Pulmonary Smear-Positive, p-: Pulmonary Smear-Negative, np: Non-pulmonary. Sub-types of recovery: N: Natural, S: Successful, D: Default (abandon of treatment). A detailed description of each state is provided through section 1.1.

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 a 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 of unvaccinated individuals is described as follows:

- Primary infection of unvaccinated individuals (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 of unvaccinated individuals (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.

In what regards the infection of susceptible, vaccinated individuals S^v, the vaccine benefits consist of reducing the probability of an individual to get infected. We model this as the only effect provided by vaccination, by introducing the vaccine efficacy coefficient $\epsilon \in [0, 1]$, which stands for the fraction of the probability of infection that a vaccinated individual presents with respect to a non vaccinated subject. Therefore, the transitions describing primary infections of vaccinated individuals in age group a are described like this:

- Primary infection of vaccinated individuals (to fast latency): transition from $S^v(a,t)$ to $L^v_f(a,t)$: $\epsilon p(a)\lambda(a,t)S^v(a,t)$ individuals/unit time.
- Primary infection of vaccinated individuals (to slow latency): transition from $S^v(a,t)$ to $L^v_s(a,t)$: $\epsilon\,(1-p(a))\,\lambda(a,t)S^v(a,t)$ individuals/unit time.

It worths remarking that individuals in latency classes do not have TB disease: they do not develop any disease symptom and they are not infectious at all. Indeed, as we will see in the following sections, they can even suffer ulterior re-infections, less probably for vaccinated individuals in L^v classes, whom the vaccine also protects against reinfection, being this the reason because of which L^v classes must be differentiated from L classes, a distinction that we must maintain for posterior stages as well–.

Probability of fast progression is an age-dependent parameter, considered to be larger in adults than in children.[2] For our proposed model we will fit these parameters in order to reproduce the proper distribution of TB cases per age group, being this one of the novel ingredients that we introduce with respect to state-of-art TB spreading models.[1,2] The exact process of fitting of $p(a)$ will be made explicit in section 2.

1.1.2 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 $\rho_{np}(a)$– and not three. These three probabilities, as the fast progression probability, are age-dependent −children are known to develop generally non-pulmonary forms of TB.[4] We will obtain these probabilities from the WHO TB database of case notifications,[5] as detailed in section 2.2.2.

In the vaccinated branch, the transitions are the same, not influenced by the vaccine; which in the context of this work only protects against infection:

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

1.1.3 Death of untreated individuals

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

Individuals who got infected despite being vaccinated and develop disease can also die because of TB; as likely as anyone else:

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

1.1.4 TB diagnosis and treatment

In what regards 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. That 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

This means that the diagnosis rate $d(t)$ is the inverse of the average period spent by individuals in the D_{p+} state, as mentioned before. In other words, it essentially corresponds to the time that sick individuals pass ignoring that they have TB disease, either because they suffer symptoms but they have not visited any doctor yet, or because, after recurring to medical services, a TB diagnosis/treatment has not been yet provided. These diagnosis refractory times are region specific, as they depend, among other factors, on the potentiality of the diagnosis services of public health systems. On the other hand, the time needed for TB diagnosis is known to vary depending the type of disease, essentially because the diagnosis tools 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.[2] 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.[2], estimations for the case detection ratios are provided for each type of disease and regions: 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:

$$\eta = \frac{\chi_{p-}}{\chi_{p+}} \left(\simeq \frac{\chi_{np}}{\chi_{p+}} \right) \tag{1}$$

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.[2] for several parameters of the Natural History. From there, we considered the lower and upper estimates of η as:

$$\eta^{\text{up}} = \frac{\chi_{p-}^{\text{up}}}{\chi_{p+}^{\text{low}}} \tag{2}$$

$$\eta^{\text{low}} = \frac{\chi_{p-}^{\text{low}}}{\chi_{p+}^{\text{up}}} \tag{3}$$

In the table 1 the values of η calculated are listed for the different regions under analysis.

Regions	χ_{p+}	χ_{p-}	η
AFRH	0.51	0.43	0.843 (0.623-1.141 c.i.)
AFRL	0.49	0.25	0.510 (0.377-0.690 c.i.)
EMR	0.45	0.53	1.178 (0.870-1.593 c.i.)
SEAR	0.64	0.51	0.797 (0.589-1.278 c.i.)
WPR	0.78	0.50	0.641 (0.474-0.867 c.i.)

Table 1: Values of χ_{p+} and χ_{p-} considered in Abu-Raddad et al.[2] and the values of η for each region.

It is noticeable that the diagnosis rate is allowed to vary in time, as it has been done in other previous models. By doing this, case detection rates are allowed to vary independently to the total volume of classes T and the incidence rates. This is achieved through the introduction of a variation term: $d(t)$, as we will explain in detail in the subsection devoted to the model calibration procedure (section 2).

Once again, the transitions are the same in the vaccinated branch:

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

1.1.5 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 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 World Health Organization groups 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).

The WHO considers an additional treatment result for individuals whose treatment outcome is not well known, either because they have been geographically transferred during the treatment ("transferred") or because their treatment outcome has not been properly evaluated ("not evaluated"). For the sake of clarity, the latter possibilities will be ignored in our model, which implicitly implies that we are assuming, as a first approximation, that these individuals of unknown outcomes would be randomly distributed among all the other four classes if we knew about their final state.

Therefore, let us denote as f_S^{p+}, f_D^{p+}, f_F^{p+} and f_μ^{p+}, the normalized 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,[5] once normalized so as to discard the unknown outcomes to get $f_S^{p+} + f_D^{p+} + f_F^{p+} + f_\mu^{p+} = 1$ what 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,[5] 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 5, in which confidence intervals correspond to two typical deviations of the time average of each parameter.

Therefore, we can enumerate all the possible treatment outcomes from all the different kinds of unvaccinated 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–.

Vaccination, once the infection –and progression to disease and diagnosis– occurred, does not affect the transitions corresponding to treatment outcomes, which are thus equivalent in both branches of the model:

- Early treatment abandon (default) of smear positive TB: transition from $T_{p+}^v(a,t)$ to $R_{p+D}^v(a,t)$: $\Psi f_D^{p+} T_{p+}^v(a,t)$ individuals/unit time.
- Failed treatment completion of smear positive TB: transition from $T_{p+}^v(a,t)$ to $F(a,t)$: $\Psi f_F^{p+} T_{p+}^v(a,t)$ individuals/unit time.

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

It is worth noticing that recovered individuals are still susceptible to suffer further reinfection, that is, as it happens with L individuals, a vaccine might protect them. This is the reason why it is necessary to differentiate R classes from R^v classes, as well as D and T classes from D^v and T^v. We consider an exception for failed recovery (class F), which is the same for the vaccinated and the unvaccinated branch. The behavior of this class is detailed in section 1.1.7.

1.1.6 Natural recovery

In certain occasions, natural recovery from TB is possible without medical intervention or treatment.[2] 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.

Once again, the same stands for the vaccinated branch:

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

1.1.7 Endogenous reactivations after treatment or natural recovery

Nonetheless, naturally recovered individuals may experience an endogenous reactivation of the disease, since disease recovery does not suppose the total elimination of the bacilli from the host organism, generally speaking. [6] If we denote as 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. In consequence, eventual relapses from F to D_{p+} class would not make sense, and they would experience a TB-related mortality described by $\mu_{p+}F(a,t)$ individuals/unit time. 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 or on whether she was vaccinated or not.

Instead, 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 very same bacilli of the original episode, that has not been 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 increasing.

In the exhaustive meta-analysis by Korenromp and colleagues, [6] 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, the 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 of 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: $R_{xS}(t) \sim e^{-r_S t}$. Therefore, 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.[7], which yields the final value of $r_D = 3.8 \cdot 10^{-3}$ 1/year ($1.4 \cdot 10^{-3} - 8.6 \cdot 10^{-3}$).

Finally, endogenous reactivations of disease transitions in the vaccinated branch are equivalent to those of the unvaccinated branch:

- Endogenous reactivation of non pulmonary TB after natural recovery (vaccinated branch): from $R^v_{npN}(a,t)$ to $D^v_{np}(a,t)$: $r_N R^v_{npN}(a,t)$ individuals/unit time.
- Endogenous reactivation of smear negative TB after natural recovery (vaccinated branch): from $R^v_{p-N}(a,t)$ to $D^v_{p-}(a,t)$: $r_N R^v_{p-N}(a,t)$ individuals/unit time.
- Endogenous reactivation of smear positive TB after natural recovery (vaccinated branch): from $R^v_{p+N}(a,t)$ to $D^v_{p+}(a,t)$: $r_N R^v_{p+N}(a,t)$ individuals/unit time.
- Endogenous reactivation of non pulmonary TB after successful treatment (vaccinated branch): from $R^v_{npS}(a,t)$ to $D^v_{np}(a,t)$: $r_S R^v_{npS}(a,t)$ individuals/unit time.
- Endogenous reactivation of smear negative TB after successful treatment (vaccinated branch): from $R^v_{p-S}(a,t)$ to $D^v_{p-}(a,t)$: $r_S R^v_{p-S}(a,t)$ individuals/unit time.
- Endogenous reactivation of smear positive TB after successful treatment (vaccinated branch): from $R^v_{p+S}(a,t)$ to $D^v_{p+}(a,t)$: $r_S R^v_{p+S}(a,t)$ individuals/unit time.
- Endogenous reactivation of non pulmonary TB after treatment default (vaccinated branch): from $R^v_{npD}(a,t)$ to $D^v_{np}(a,t)$: $r_D R^v_{npD}(a,t)$ individuals/unit time.
- Endogenous reactivation of smear negative TB after treatment default (vaccinated branch): from $R^v_{p-D}(a,t)$ to $D^v_{p-}(a,t)$: $r_D R^v_{p-D}(a,t)$ individuals/unit time.
- Endogenous reactivation of smear positive TB after treatment default (vaccinated branch): from $R^v_{p+D}(a,t)$ to $D^v_{p+}(a,t)$: $r_D R^v_{p+D}(a,t)$ individuals/unit time.

1.1.8 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, R or N 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 has been infected in a previous episode.

Regarding the vaccinated branch, the individuals in classes L_s^v and R^v are supposed to have a lower risk of being reinfected, as a consequence of the effect of the vaccine, which is modelized by the presence of the ϵ factor in the following transitions:

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

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.

Therefore, 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.[8] it is described how the progression to TB disease increases with the number of exposures.

1.1.9 Smear progression

In certain cases, it is documented that patients of smear negative pulmonary TB progress to smear positive ,[1] even after being treated. In order to describe this phenomenon, we introduce the smear progression by considering the following two transitions in the non vaccinated branch of the model:

- 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. In the vaccinated branch we have the same behavior:

- Smear progression of untreated individuals (vaccinated branch): from $D^v_{p-}(a,t)$ to $D^v_{p+}(a,t)$: $\theta D^v_{p-}(a,t)$ individuals/unit time.
- Smear progression of individuals under treatment (vaccinated branch): from $T^v_{p-}(a,t)$ to $T^v_{p+}(a,t)$: $\theta T^v_{p-}(a,t)$ individuals/unit time.

1.1.10 Newborn vaccination. Mother-child disease transmission

Our model allows for the description of newborn-focused vaccination campaigns, like that required to introduce a novel BCG-substitutive vaccine. In order to do so, if we have a number of $\Delta_N(a=0,t)$ newborns at time-step t –in section 1.5 we will detail how we obtain this $\Delta_N(a=0,t)$– , we start by splitting them between the non vaccinated branch and the vaccinated branch. $c(a,t)$ represents the coverage of the vaccination campaign at group age a at time t. In our case, we will deal with a substitutive vaccine ideally applied to the 100% of newborns from 2025 on. Let us denote as $c(a=0,t)$ the fraction of newborn individuals that get vaccinated at time t. Among the fraction $(1-c(0,t))$ of children who are not vaccinated, a percentage of them will be infected right after birth by their mothers, as it is a well known fact that an m_t fraction of sick women who are pregnant transmits the disease to their children within the first weeks of their lives.[9] The density of infected newborns obviously depends 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 offsprings for women in each of the infectious classes included in our model. In this work we considered the total number of newborn infections proportional to the fraction m_d:

$$m_d(t) = \frac{\sum_{a=3}^{a=7} D_{p+}(a,t) + D_{p-}(a,t) + D_{np}(a,t) + R_{p+D}(a,t) + R_{p-D}(a,t) + R_{npD}(a,t) + F(a,t)}{\sum_{a=3}^{a=7} N(a,t)} + \\ + \frac{\sum_{a=3}^{a=7} D^v_{p+}(a,t) + D^v_{p-}(a,t) + D^v_{np}(a,t) + R^v_{p+D}(a,t) + R^v_{p-D}(a,t) + R^v_{npD}(a,t)}{\sum_{a=3}^{a=7} N(a,t)} \tag{4}$$

that accounts to 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 $(1-c(0,t))\Delta_N(a=0,t)$ unvaccinated newborns among S and L classes:

- Birth of $S(0,t)$ individuals (susceptible newborns who will not be vaccinated right after birth): $(1-m_t m_d(t))(1-c(0,t))\Delta_N(a=0,t)$
- Birth of $L_f(0,t)$ individuals (infected after birth who will develop fast progression): $m_t m_d(t) p(0)(1-c(0,t))\Delta_N(a=0,t)$

- Birth of $L_s(0,t)$ individuals (infected after birth who will develop slow progression): $m_t m_d(t)(1-p(0))(1-c(0,t))\Delta_N(a=0,t)$

Concerning the vaccinated branch of the system, if we suppose that newborn vaccination has also a protective effect against mother-child transmission –i.e. it diminishes the probability for vaccinated children to get infected from their mothers from m_t to ϵm_t –, we have that the $c(0,t)\Delta_N(a=0,t)$ newborns that are immunized at time step t are distributed among $S^v$and L^v classes as follows:

- Birth of $S^v(0,t)$ individuals (susceptible, vaccinated right after birth): $(1-\epsilon m_t m_d(t))c(0,t)\Delta_N(a=0,t)$
- Birth of $L_f^v(0,t)$ individuals (vaccinated, yet infected after birth who will develop fast progression): $\epsilon m_t m_d(t)p(0)c(0,t)\Delta_N(a=0,t)$
- Birth of $L_s^v(0,t)$ individuals (vaccinates, yet infected after birth who will develop slow progression): $\epsilon m_t m_d(t)(1-p(0))c(0,t)\Delta_N(a=0,t)$

1.2 Force of infection

The force of infection $\lambda(a,t)$ is, as it has been said before in section 1.1.1, the rate at which infection occurs at time step t of a susceptible individual at age-group a. This magnitude is calculated according to the following expression:

$$\lambda(a,t) = \beta(t)\sum_{a'}\xi(a,a')\Upsilon(a',t) \tag{5}$$

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

$$\begin{aligned}\Upsilon(a',t) = D_{p+}(a',t) + F(a',t) + \phi_{p-}D_{p-}(a',t) + \phi_D R_{p+D}(a',t) + \phi_{p-}\phi_D R_{p-D} + \\ +D^v_{p+}(a',t) + \phi_{p-}D^v_{p-}(a',t) + \phi_D R^v_{p+D}(a',t) + \phi_{p-}\phi_D R^v_{p-D}\end{aligned} \tag{6}$$

where $\phi_{p-} \in [0,1]$ 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 a infectiousness reduction that is the product of the two terms $\phi_{p-}\phi_D$.

On the other hand, $\xi(a,a')$ represents the number of contacts per year that an individual within the age group a maintains with individuals of age group a'. In addition, $\beta(t)$ represents the scaled based infectiousness, which is a function of time but not of age. Next, in section 1.3, we explain how we obtain the contact matrix $\xi(a,a')$.

1.3 Contact patterns

Perhaps, one of the most pervasive simplifying hypothesis in the mathematical modeling of TB spreading is the homogeneous mixing of the populations under study. In this work, we abandon this hypothesis by defining an heterogeneous contact matrix among different age groups denoted as $\xi(a,a')$. It has been shown that abandoning the hypothesis of homogeneous age-mixing in favor of a data-driven, more realistic approach,[3,10] improves the descriptive capabilities of epidemiological models of influenza-like diseases.[11,12] Similarly, in a previous work by Guzzetta and collaborators,[13] the importance of considering heterogeneous contacts in the modeling of TB was assessed in the context of Individual Based Modeling (IBM).

In our case, the matrix elements $\xi(a,a') \in [0,1]$ represent the fraction of all possible contacts among individuals of age groups a and a' that actually take place. For example, if we have $N(a,t)$ and $N(a',t)$ in each group at a certain time t, then we count as many as $\xi(a,a')N(a,t)N(a',t)$ contacts between these two groups of individuals. For the computation of the contact matrix used in our model, we take advantage of the data provided by the survey performed in the Polymod project.[3] In this survey, people were asked how many contacts they have had in a certain time window, and the ages of the people they have maintained contact with. Let us denote as $P(a,a')$ the elements of this matrix, which represents the average number of contacts that an individual in age group a has, per unit time, with people in age group a'. These matrices $P(a,a')$ are provided in Mossong et al.[3] for different European countries. In the following, we explain how to get the contact matrix $\xi(a,a')$ from these data.

First of all, it is important to note that, for any country, the matrix $P(a,a')$ is not, –in general–, symmetric, which is due to the different number of people in each age group. The number of total contacts between two age groups a and a' can be computed in two different ways: if we rely on the reports of class a we would write it as the product $P(a,a')N(a,t)$, where $N(a,t)$ is the number of individuals in age group a at time t. Instead, if we rely our estimation of the number of contacts on the reports by individuals in class a', we would obtain $P(a',a)N(a',t)$. Obviously, if the survey has enough quality, one should find:

$$P(a,a')N(a,t) \simeq P(a',a)N(a',t) \Rightarrow \frac{P(a,a')}{N(a',t)} \simeq \frac{P(a',a)}{N(a,t)} \tag{7}$$

So, if t_s is the moment at which surveys took place, we may build up the following matrix, with the same units of $\xi(a,a')$ (i.e. fraction of total contacts between age groups a and a' per unit time):

$$A(a,a') = \frac{P(a,a')}{N(a',t_s)} \tag{8}$$

which is not fully symmetrical. Indeed, what we do is to consider the expected value of the fraction of total contacts as its symmetric part, denoted as $B(a,a')$:

$$B(a,a') = B(a',a) = \frac{A(a,a') + A(a',a)}{2} \tag{9}$$

And use the asymmetric part, which is the measure of the incoherence between the reports of age groups a and a' when referring the number of contacts between them, as an error estimator:

$$\delta B(a,a') = \delta B(a',a) = \frac{1}{\sqrt{2}}\sqrt{\frac{\left(A(a,a') - B(a,a')\right)^2 + \left(A(a',a) - B(a,a')\right)^2}{2}} \tag{10}$$

The Polymod project collected and rendered publicly available the matrices $P(a,a')$ for eight different european countries, as stemmed from the surveys performed between 2005 and 2006.[3] The mixing patterns are highly coherent among the eight countries analyzed, which allows us to build the following average, that we assimilate to $\xi(a,a')$:

$$\xi(a,a') = \frac{\sum_c \left[B(a,a')\right]_c \left[N(a,t_s)\right]_c \left[N(a',t_s)\right]_c}{\sum_c \left[N(a,t_s)\right]_c \left[N(a',t_s)\right]_c} \tag{11}$$

where c is an index that indicates the country. Finally, we consider $B(a,a')$ as the only source of error:

$$\delta\xi(a,a') = \sqrt{\sum_c \left(\frac{\partial \xi(a,a')}{\partial \left[B(a,a')\right]_c}\left[\delta B(a,a')\right]_c\right)^2} = \frac{\sqrt{\sum_c \left(\left[\delta B(a,a')\right]_c \left[N(a,t_s)\right]_c \left[N(a',t_s)\right]_c\right)^2}}{\sum_c \left[N(a,t_s)\right]_c \left[N(a',t_s)\right]_c} \tag{12}$$

Therefore, our hypothesis is that we can use the matrix $\xi(a,a')$ –see figure 1D of the main text– as a first approximation to the description of the heterogeneity of mixing patterns in any part of the world. In relation to this, it is also worth noticing different aspects of our approximation. On the one hand, by scaling the original matrices $P(a,a')$ towards the final matrix $\xi(a,a')$ using the demographic structures of the countries surveyed $N(a,t)$ at $t = t_s$ –available at UN Population Database[14]–, we eliminate the influence of the demographic structure of these countries on the mixing matrix, and so, we come up with an object that is useful to describe the age dependent mixing patterns regardless of region-wise specific demographical details. In conclusion, the fact that other regions have different demographic structures different from Europe imposes no limitation to the generalization of $\xi(a,a')$ out of Europe. Furthermore, the overall intensity of the mixing terms is modulated by the common factor $\beta(t)$, which is fitted independently for each region: this is of utmost relevance, given that the average strength and frequency of social interactions are, undoubtedly, highly dependent on cultural and socio-economic factors.

In any case, even if we account for eventual variations on the overall contact strength by fitting the parameter $\beta(t)$, it is also true that the structure of the specific mixing patterns in regions outside Europe could show differential variations among the different fields of the matrices with respect to the European case calculated as $\xi(a,a')$. In the lack of data about contact patterns in all the regions analyzed, we hypothesize that deviations from the actual patterns occurring in each region from our survey-based estimations will arguably constitute a second order source of error when compared to the use of homogeneous matrices. Unlike the homogeneous mixing hypothesis, our contact patterns exhibit certain features which are reasonably expected to be present in the vast majority of human societies. These are :

1. A strong trend to interact with people of the same age.
2. The presence of more frequent and intense contacts during the childhood and the adolescence.
3. The appearance of inter-generational contacts between parents and children (or grandparents and grandchildren).

1.4 Aging and age-specific vaccination

The model considers 14 different age groups. Each of these groups comprises an age interval of $\Delta_t = 5$ years. We thus consider individuals from 0 to 70 years old, being the latter a reference value for life expectancy worldwide. 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 $X(a+1,t)$. Furthermore, we will couple these transitions with the possible immunization of individuals at an specific age. In that scenario, individuals will progress from classes $X(a,t)$ to classes $X^v(a+1,t)$.

- Aging of individuals belonging to class $X(a,t)$ without vaccination, transition from $X(a,t)$ to $X(a+1,t)$: $(1-c(a,t))\,X(a,t)/\Delta_t$ individuals/unit time.
- Aging of individuals belonging to class $X(a,t)$ with vaccination, transition from $X(a,t)$ to $X^v(a+1,t)$: $c(a,t)X(a,t)/\Delta_t$ individuals/unit time.

where $c(a,t)$ is the coverage of the vaccine for age group a at time t. This vaccination process is valid for $a > 0$, the case of $a = 0$, which corresponds to newborn vaccination, is not coupled with aging but with birth as explained in section 1.1.10. 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 $X(13,t)$, from which the exit flow represents the death of the eldest individuals in the system. In the vaccinated branch, only aging takes place, and thus it makes no sense to consider vaccination.

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

In this work, we consider, apart from the newborn vaccination already explained, adolescent focused vaccination –booster vaccines– that takes places in the transition between the third and the fourth age groups (i.e., vaccination at 15 years old is modeled as a series of fluxes from $X(a=2,t)$ to $X^v(a=3,t)$). Besides, we assume that only individuals that have not suffered an episode of TB are vaccinated. Thus, vaccination only applies to classes S, L_f and L_s.

1.5 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:

$$\begin{aligned} N(a,t) = S(a,t) + L_f(a,t) + L_s(a,t) + D_{p+}(a,t) + D_{p-}(a,t) + D_{np}(a,t) + T_{p+}(a,t) + T_{p-}(a,t) + T_{np}(a,t) \\ +F(a,t) + R_{p+N}(a,t) + R_{p-N}(a,t) + R_{npN}(a,t) + R_{p+S}(a,t) + R_{p-S}(a,t) + R_{npS}(a,t) \\ +R_{p+D}(a,t) + R_{p-D}(a,t) + R_{npD}(a,t) + S^v(a,t) + L_f^v(a,t) + L_s^v(a,t) + D_{p+}^v(a,t) \\ +D_{p-}^v(a,t) + D_{np}^v(a,t) + T_{p+}^v(a,t) + T_{p-}^v(a,t) + T_{np}^v(a,t) + R_{p+N}^v(a,t) + R_{p-N}^v(a,t) \\ +R_{npN}^v(a,t) + R_{p+S}^v(a,t) + R_{p-S}^v(a,t) + R_{npS}^v(a,t) + R_{p+D}^v(a,t) + R_{p-D}^v(a,t) + R_{npD}^v(a,t) \end{aligned} \tag{13}$$

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 $\mathcal{N}(t) = \mathcal{N}(t=0)\forall(t)$.[2] A more sophisticated alternative, adopted in Dye et al.[1], 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: $N(a,t) = N_{UN}(a,t)$, see figure S2.[14] From figure S2 we can see that population aging is a common feature in all the regions under study.

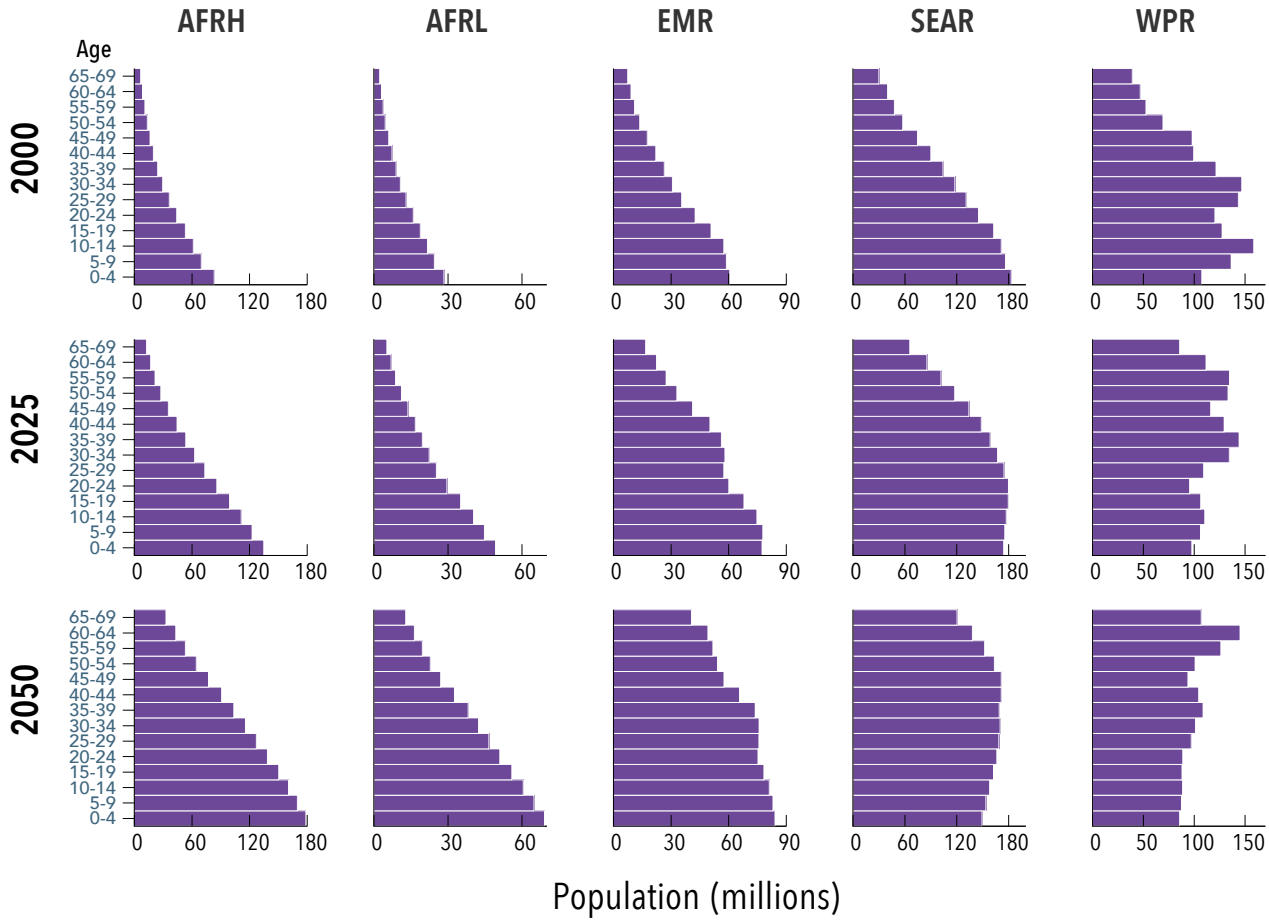

Figure S2: Population structures in 2000, 2025 and 2050 projected by UN population division[14] in each of the region under study.

In the following sub-sections, we detail the last two different schemes, whose influence on model forecasts are analyzed in the paper.

Approach 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,[1] in which 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.[1] 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 –in which 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:

$$\begin{aligned}\dot{N}_o(a,t) = ((1-\delta(a))N(a-1,t) - N(a,t))/\Delta_t - \\ -\mu_{p+}(D_{p+}(a,t) + F(a,t) + D^v_{p+}(a,t)) - \\ -\mu_{p-}(D_{p-}(a,t) + D^v_{p-}(a,t)) - \mu_{np}(D_{np}(a,t) + D^v_{np}(a,t)) - \\ -\Psi f^{p+}_{\mu}(T_{p+}(a,t) + T^v_{p+}(a,t)) - \Psi f^{p-}_{\mu}(T_{p-}(a,t) + T^v_{p-}(a,t) + T_{np}(a,t) + T^v_{np}(a,t))\end{aligned} \tag{14}$$

being $\delta(a)$ the Kronecker 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:

$$\dot{N}(a,t) = \dot{N}_o(a,t) + \Delta_N(a,t) = 0 \ \forall(a,t) \tag{15}$$

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:

$$\Delta_X(a,t) = \Delta_N(a,t)\frac{X(a,t)}{N(a,t)} \tag{16}$$

and obviously:

$$\Delta_N(a,t) = \sum_X \Delta_X(a,t) \tag{17}$$

This scheme has the advantage, with respect to consider (as in Abu-Raddad et al.[2]), 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.[1] 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 regions subjected to strong processes of demographic aging.

Approach 2: Demographic structure as 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 that 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.[14] After adequately grouping the countries, we obtain the actual annual population series expected by the UN for the populations at each age group, that 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.[14]

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 14, 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 $N_{UN}(a,t)$. That is achieved by defining, at each time step:

$$\Delta_N(a,t) = \dot{N}_{UN}(a,t) - \dot{N}_o(a,t) \tag{18}$$

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 $a > 0$:

$$\Delta_X(a,t) = \Delta_N(a,t)\frac{X(a,t)}{N(a,t)} \ \forall a > 0 \tag{19}$$

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 situation is different for the first age class $a = 0$. In the first age group, the introduction of newborn individuals is the main cause of population variation and it is not distributed uniformly among the different dynamical states, but only among S and L states. For these reasons, and once observed that $\Delta_N(a = 0, t) > 0 \ \forall t$ in all regions under consideration, for simplicity $\Delta_N(a = 0, t)$ is directly associated to the number of newborns, as it was previously shown in section 1.1.10, where we addressed newborn vaccination and mother-child disease transmission and how these $\Delta_N(a = 0, t)$ newborns are distributed among classes S, S^v, L and L^v.

The uncertainty of UN demographic projections is also reported at UN Database, [14] which allows us to reconstruct the demographic structures at the extremes of the confidence interval (95%) $N^{\text{low}}_{UN}(a,t)$ and $N^{\text{high}}_{UN}(a,t)$. 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).

1.6 Ordinary differential equations system

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

$$\begin{aligned}
\dot{S}(a,t) &= -\lambda(a,t)S(a,t) - (S(a,t) - (1-c(a,t))(1-\delta(a))S(a-1,t))/\tau \\
&+ \delta(a)(1-m_t m_d(t))(1-c(0,t))\Delta_N(a,t) + (1-\delta(a))\Delta_N(a,t)S(a,t)/N(a,t)
\end{aligned} \tag{20}$$

$$\begin{aligned}
\dot{L}_s(a,t) &= (1-p(a))\lambda(a,t)S(a,t) - p(a)q\lambda(a,t)L_s(a,t) - \omega_s L_s(a,t) \\
&- (L_s(a,t) - (1-c(a,t))(1-\delta(a))L_s(a-1,t))/\tau \\
&+ \delta(a)m_t m_d(t)(1-p(a))(1-c(0,t))\Delta_N(a,t) + (1-\delta(a))\Delta_N(a,t)L_s(a,t)/N(a,t)
\end{aligned} \tag{21}$$

$$\begin{aligned}
\dot{L}_f(a,t) &= p(a)\lambda(a,t)S(a,t) - \omega_f L_f(a,t) + p(a)q\lambda(a,t)(L_s(a,t) + R_{p+N}(a,t) + R_{p-N}(a,t) + R_{npN}(a,t)) \\
&+ p(a)q\lambda(a,t)(R_{p+S}(a,t) + R_{p-S}(a,t) + R_{npS}(a,t) + R_{p+D}(a,t) + R_{p-D}(a,t) + R_{npD}(a,t)) \\
&- (L_f(a,t) - (1-c(a,t))(1-\delta(a))L_f(a-1,t))/\tau \\
&+ \delta(a)m_t m_d(t)p(a)(1-c(0,t))\Delta_N(a,t) + (1-\delta(a))\Delta_N(a,t)L_f(a,t)/N(a,t)
\end{aligned} \tag{22}$$

$$\begin{aligned}
\dot{D}_{p+}(a,t) &= \omega_f \rho_{p+}(a)L_f(a,t) + \omega_s \rho_{p+}(a)L_s(a,t) - \mu_{p+}D_{p+}(a,t) - d(t)D_{p+}(a,t) \\
&- \nu D_{p+}(a,t) + r_N R_{p+N}(a,t) + r_S R_{p+S}(a,t) + r_D R_{p+D}(a,t) + \theta D_{p-}(a,t) \\
&- (D_{p+}(a,t) - (1-\delta(a))D_{p+}(a-1,t))/\tau + (1-\delta(a))\Delta_N(a,t)D_{p+}(a,t)/N(a,t)
\end{aligned} \tag{23}$$

$$\begin{aligned}
\dot{D}_{p-}(a,t) &= \omega_f(1-\rho_{p+}(a)-\rho_{np}(a))L_f(a,t) + \omega_s(1-\rho_{p+}(a)-\rho_{np}(a))L_s(a,t) - \mu_{p-}D_{p-}(a,t) \\
&- \eta d(t)D_{p-}(a,t) - \nu D_{p-}(a,t) + r_N R_{p-N}(a,t) + r_S R_{p-S}(a,t) + r_D R_{p-D}(a,t) - \theta D_{p-}(a,t) \\
&- (D_{p-}(a,t) - (1-\delta(a))D_{p-}(a-1,t))/\tau + (1-\delta(a))\Delta_N(a,t)D_{p-}(a,t)/N(a,t)
\end{aligned} \tag{24}$$

$$\begin{aligned}
\dot{D}_{np}(a,t) &= \omega_f \rho_{np}(a)L_f(a,t) + \omega_s \rho_{np}(a)L_s(a,t) - \mu_{np}D_{np}(a,t) - \eta d(t)D_{np}(a,t) \\
&- \nu D_{np}(a,t) + r_N R_{npN}(a,t) + r_S R_{npS}(a,t) + r_D R_{npD}(a,t) \\
&- (D_{np}(a,t) - (1-\delta(a))D_{np}(a-1,t))/\tau + (1-\delta(a))\Delta_N(a,t)D_{np}(a,t)/N(a,t)
\end{aligned} \tag{25}$$

$$
\begin{aligned}
\dot{T}_{p+}(a,t) &= d(t)D_{p+}(a,t) - \Psi T_{p+}(a,t) + \theta T_{p-}(a,t) \\
&- (T_{p+}(a,t) - (1-\delta(a))T_{p+}(a-1,t))/\tau + (1-\delta(a))\Delta_N(a,t)T_{p+}(a,t)/N(a,t)
\end{aligned} \tag{26}
$$

$$
\begin{aligned}
\dot{T}_{p-}(a,t) &= \eta d(t)D_{p-}(a,t) - \Psi T_{p-}(a,t) - \theta T_{p-}(a,t) \\
&- (T_{p-}(a,t) - (1-\delta(a))T_{p-}(a-1,t))/\tau + (1-\delta(a))\Delta_N(a,t)T_{p-}(a,t)/N(a,t)
\end{aligned} \tag{27}
$$

$$
\begin{aligned}
\dot{T}_{np}(a,t) &= \eta d(t)D_{np}(a,t) - \Psi T_{np}(a,t) \\
&- (T_{np}(a,t) - (1-\delta(a))T_{np}(a-1,t))/\tau + (1-\delta(a))\Delta_N(a,t)T_{np}(a,t)/N(a,t)
\end{aligned} \tag{28}
$$

$$
\begin{aligned}
\dot{F}(a,t) &= \Psi f_F^{p+}(T_{p+}(a,t) + T_{p+}^v(a,t)) + \Psi f_F^{p-}(T_{p-}(a,t) + T_{p-}^v(a,t) + T_{np}(a,t) + T_{np}^v(a,t)) \\
&- (F(a,t) - (1-\delta(a))F(a-1,t))/\tau + (1-\delta(a))\Delta_N(a,t)F(a,t)/N(a,t)
\end{aligned} \tag{29}
$$

$$
\begin{aligned}
\dot{R}_{p+N}(a,t) &= \nu D_{p+}(a,t) - r_N R_{p+N}(a,t) - p(a)q\lambda(a,t)R_{p+N}(a,t) \\
&- (R_{p+N}(a,t) - (1-\delta(a))R_{p+N}(a-1,t))/\tau + (1-\delta(a))\Delta_N(a,t)R_{p+N}(a,t)/N(a,t)
\end{aligned} \tag{30}
$$

$$
\begin{aligned}
\dot{R}_{p-N}(a,t) &= \nu D_{p-}(a,t) - r_N R_{p-N}(a,t) - p(a)q\lambda(a,t)R_{p-N}(a,t) \\
&- (R_{p-N}(a,t) - (1-\delta(a))R_{p-N}(a-1,t))/\tau + (1-\delta(a))\Delta_N(a,t)R_{p-N}(a,t)/N(a,t)
\end{aligned} \tag{31}
$$

$$
\begin{aligned}
\dot{R}_{npN}(a,t) &= \nu D_{np}(a,t) - r_N R_{npN}(a,t) - p(a)q\lambda(a,t)R_{npN}(a,t) \\
&- (R_{npN}(a,t) - (1-\delta(a))R_{npN}(a-1,t))/\tau + (1-\delta(a))\Delta_N(a,t)R_{npN}(a,t)/N(a,t)
\end{aligned} \tag{32}
$$

$$
\begin{aligned}
\dot{R}_{p+S}(a,t) &= \Psi(1 - f_D^{p+} - f_F^{p+} - f_\mu^{p+})T_{p+}(a,t) - r_S R_{p+S}(a,t) - p(a)q\lambda(a,t)R_{p+S}(a,t) \\
&- (R_{p+S}(a,t) - (1-\delta(a))R_{p+S}(a-1,t))/\tau + (1-\delta(a))\Delta_N(a,t)R_{p+S}(a,t)/N(a,t)
\end{aligned} \tag{33}
$$

$$
\begin{aligned}
\dot{R}_{p-S}(a,t) &= \Psi(1 - f_D^{p-} - f_F^{p-} - f_\mu^{p-})T_{p-}(a,t) - r_S R_{p-S}(a,t) - p(a)q\lambda(a,t)R_{p-S}(a,t) \\
&- (R_{p-S}(a,t) - (1-\delta(a))R_{p-S}(a-1,t))/\tau + (1-\delta(a))\Delta_N(a,t)R_{p-S}(a,t)/N(a,t)
\end{aligned} \tag{34}
$$

$$
\begin{aligned}
\dot{R}_{npS}(a,t) &= \Psi(1 - f_D^{p-} - f_F^{p-} - f_\mu^{p-})T_{np}(a,t) - r_S R_{npS}(a,t) - p(a)q\lambda(a,t)R_{npS}(a,t) \\
&- (R_{npS}(a,t) - (1-\delta(a))R_{npS}(a-1,t))/\tau + (1-\delta(a))\Delta_N(a,t)R_{npS}(a,t)/N(a,t)
\end{aligned} \tag{35}
$$

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

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

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

$$
\begin{aligned}
\dot{S}^v(a,t) &= -\epsilon\lambda(a,t)S^v(a,t) - (S^v(a,t) - (1-\delta(a))\left[S^v(a-1,t) + c(a,t)S(a-1,t)\right])/\tau \\
&+ \delta(a)(1-\epsilon m_t m_d(t))c(0,t)\Delta_N(a,t) + (1-\delta(a))\Delta_N(a,t)S^v(a,t)/N(a,t)
\end{aligned} \tag{39}
$$

$$
\begin{aligned}
\dot{L}_s^v(a,t) &= (1-p(a))\epsilon\lambda(a,t)S^v(a,t) - \omega_s L_s^v(a,t) - p(a)q\epsilon\lambda(a,t)L_s^v(a,t) \\
&- (L_s^v(a,t) - (1-\delta(a))\left[L_s^v(a-1,t) + c(a,t)L_s(a-1,t)\right])/\tau + (1-\delta(a))\Delta_N(a,t)L_s^v(a,t)/N(a,t) \\
&+ \delta(a)\epsilon m_t m_d(t)(1-p(a))c(0,t)\Delta_N(0,t)
\end{aligned} \tag{40}
$$

$$
\begin{aligned}
\dot{L}_f^v(a,t) &= p(a)\epsilon\lambda(a,t)S^v(a,t) - \omega_f L_f^v(a,t) + p(a)q\epsilon\lambda(a,t)(L_s^v(a,t) + R_{p+N}^v(a,t) + R_{p-N}^v(a,t) + R_{npN}^v(a,t)) \\
&+ p(a)q\epsilon\lambda(a,t)(R_{p+S}^v(a,t) + R_{p-S}^v(a,t) + R_{npS}^v(a,t) + R_{p+D}^v(a,t) + R_{p-D}^v(a,t) + R_{npD}^v(a,t)) \\
&- (L_f^v(a,t) - (1-\delta(a))\left[L_f^v(a-1,t) + c(a,t)L_f(a-1,t)\right])/\tau + (1-\delta(a))\Delta_N(a,t)L_f^v(a,t)/N(a,t) \\
&+ \delta(a)\epsilon m_t m_d(t)p(a)c(0,t)\Delta_N(0,t)
\end{aligned} \tag{41}
$$

$$
\begin{aligned}
\dot{D}_{p+}^v(a,t) &= \omega_f\rho_{p+}(a)L_f^v(a,t) + \omega_s\rho_{p+}(a)L_s^v(a,t) - \mu_{p+}D_{p+}^v(a,t) - d(t)D_{p+}^v(a,t) \\
&- \nu D_{p+}^v(a,t) + r_N R_{p+N}^v(a,t) + r_S R_{p+S}^v(a,t) + r_D R_{p+D}^v(a,t) + \theta D_{p-}^v(a,t) \\
&- (D_{p+}^v(a,t) - (1-\delta(a))D_{p+}^v(a-1,t))/\tau + (1-\delta(a))\Delta_N(a,t)D_{p+}^v(a,t)/N(a,t)
\end{aligned} \tag{42}
$$

$$
\begin{aligned}
\dot{D}_{p+}^v(a,t) &= \omega_f\rho_{p+}(a)L_f^v(a,t) + \omega_s\rho_{p+}(a)L_s^v(a,t) - \mu_{p+}D_{p+}^v(a,t) - d(t)D_{p+}^v(a,t) \\
&- \nu D_{p+}^v(a,t) + r_N R_{p+N}^v(a,t) + r_S R_{p+S}^v(a,t) + r_D R_{p+D}^v(a,t) + \theta D_{p-}^v(a,t) \\
&- (D_{p+}^v(a,t) - (1-\delta(a))D_{p+}^v(a-1,t))/\tau + (1-\delta(a))\Delta_N(a,t)D_{p+}^v(a,t)/N(a,t)
\end{aligned} \tag{43}
$$

$$
\begin{aligned}
\dot{D}_{p-}^v(a,t) &= \omega_f(1-\rho_{p+}(a)-\rho_{np}(a))L_f^v(a,t) + \omega_s(1-\rho_{p+}(a)-\rho_{np}(a))L_s^v(a,t) - \mu_{p-}D_{p-}^v(a,t) \\
&- \eta d(t)D_{p-}^v(a,t) - \nu D_{p-}^v(a,t) + r_N R_{p-N}^v(a,t) + r_S R_{p-S}^v(a,t) + r_D R_{p-D}^v(a,t) - \theta D_{p-}^v(a,t) \\
&- (D_{p-}^v(a,t) - (1-\delta(a))D_{p-}^v(a-1,t))/\tau + (1-\delta(a))\Delta_N(a,t)D_{p-}^v(a,t)/N(a,t)
\end{aligned} \tag{44}
$$

$$
\begin{aligned}
\dot{D}_{np}^v(a,t) &= \omega_f\rho_{np}(a)L_f^v(a,t) + \omega_s\rho_{np}(a)L_s^v(a,t) - \mu_{np}D_{np}^v(a,t) - \eta d(t)D_{np}^v(a,t) \\
&- \nu D_{np}^v(a,t) + r_N R_{npN}^v(a,t) + r_S R_{npS}^v(a,t) + r_D R_{npD}^v(a,t) \\
&- (D_{np}^v(a,t) - (1-\delta(a))D_{np}^v(a-1,t))/\tau + (1-\delta(a))\Delta_N(a,t)D_{np}^v(a,t)/N(a,t)
\end{aligned} \tag{45}
$$

$$
\begin{aligned}
\dot{T}_{p+}(a,t) &= d(t)D_{p+}(a,t) - \Psi T_{p+}(a,t) + \theta T_{p-}(a,t) \\
&- (T_{p+}(a,t) - (1-\delta(a))T_{p+}(a-1,t))/\tau + (1-\delta(a))\Delta_N(a,t)T_{p+}(a,t)/N(a,t)
\end{aligned} \tag{46}
$$

$$
\begin{aligned}
\dot{T}_{p-}^v(a,t) &= \eta d(t)D_{p-}^v(a,t) - \Psi T_{p-}^v(a,t) - \theta T_{p-}^v(a,t) \\
&- (T_{p-}^v(a,t) - (1-\delta(a))T_{p-}^v(a-1,t))/\tau + (1-\delta(a))\Delta_N(a,t)T_{p-}^v(a,t)/N(a,t)
\end{aligned} \tag{47}
$$

$$
\begin{aligned}
\dot{T}_{np}^v(a,t) &= \eta d(t)D_{np}^v(a,t) - \Psi T_{np}^v(a,t) \\
&- (T_{np}^v(a,t) - (1-\delta(a))T_{np}^v(a-1,t))/\tau + (1-\delta(a))\Delta_N(a,t)T_{np}^v(a,t)/N(a,t)
\end{aligned} \tag{48}
$$

$$
\begin{aligned}
\dot{R}_{p+N}^v(a,t) &= \nu D_{p+}^v(a,t) - r_N R_{p+N}^v(a,t) - p(a)q\lambda(a,t)R_{p+N}^v(a,t) \\
&- (R_{p+N}^v(a,t) - (1-\delta(a))R_{p+N}^v(a-1,t))/\tau + (1-\delta(a))\Delta_N(a,t)R_{p+N}^v(a,t)/N(a,t)
\end{aligned} \tag{49}
$$

$$
\begin{aligned}
\dot{R}_{p-N}^v(a,t) &= \nu D_{p-}^v(a,t) - r_N R_{p-N}^v(a,t) - p(a)q\epsilon\lambda(a,t)R_{p-N}^v(a,t) \\
&- (R_{p-N}^v(a,t) - (1-\delta(a))R_{p-N}^v(a-1,t))/\tau + (1-\delta(a))\Delta_N(a,t)R_{p-N}^v(a,t)/N(a,t)
\end{aligned} \tag{50}
$$

$$
\begin{aligned}
\dot{R}_{p-N}^v(a,t) &= \nu D_{p-}^v(a,t) - r_N R_{p-N}^v(a,t) - p(a)q\epsilon\lambda(a,t)R_{p-N}^v(a,t) \\
&- (R_{p-N}^v(a,t) - (1-\delta(a))R_{p-N}^v(a-1,t))/\tau + (1-\delta(a))\Delta_N(a,t)R_{p-N}^v(a,t)/N(a,t)
\end{aligned} \tag{51}
$$

$$
\begin{aligned}
\dot{R}_{npN}^v(a,t) &= \nu D_{np}^v(a,t) - r_N R_{npN}^v(a,t) - p(a)q\epsilon\lambda(a,t)R_{npN}^v(a,t) \\
&- (R_{npN}^v(a,t) - (1-\delta(a))R_{npN}^v(a-1,t))/\tau + (1-\delta(a))\Delta_N(a,t)R_{npN}^v(a,t)/N(a,t)
\end{aligned} \tag{52}
$$

$$
\begin{aligned}
\dot{R}_{p+S}^v(a,t) &= \Psi(1 - f_D^{p+} - f_F^{p+} - f_\mu^{p+})T_{p+}^v(a,t) - r_S R_{p+S}^v(a,t) - p(a)q\epsilon\lambda(a,t)R_{p+S}^v(a,t) \\
&- (R_{p+S}^v(a,t) - (1-\delta(a))R_{p+S}^v(a-1,t))/\tau + (1-\delta(a))\Delta_N(a,t)R_{p+S}^v(a,t)/N(a,t)
\end{aligned} \tag{53}
$$

$$
\begin{aligned}
\dot{R}_{npN}^v(a,t) &= \nu D_{np}^v(a,t) - r_N R_{npN}^v(a,t) - p(a)q\epsilon\lambda(a,t)R_{npN}^v(a,t) \\
&- (R_{npN}^v(a,t) - (1-\delta(a))R_{npN}^v(a-1,t))/\tau + (1-\delta(a))\Delta_N(a,t)R_{npN}^v(a,t)/N(a,t)
\end{aligned} \tag{54}
$$

$$\begin{aligned}\dot{R}^v_{p+S}(a,t) &= \Psi(1-f^{p+}_D-f^{p+}_F-f^{p+}_\mu)T^v_{p+}(a,t)-r_S R^v_{p+S}(a,t)-p(a)q\epsilon\lambda(a,t)R^v_{p+S}(a,t) \\ &- (R^v_{p+S}(a,t)-(1-\delta(a))R^v_{p+S}(a-1,t))/\tau+(1-\delta(a))\Delta_N(a,t)R^v_{p+S}(a,t)/N(a,t)\end{aligned} \tag{55}$$

$$\begin{aligned}\dot{R}^v_{p-S}(a,t) &= \Psi(1-f^{p-}_D-f^{p-}_F-f^{p-}_\mu)T^v_{p-}(a,t)-r_S R^v_{p-S}(a,t)-p(a)q\epsilon\lambda(a,t)R^v_{p-S}(a,t) \\ &- (R^v_{p-S}(a,t)-(1-\delta(a))R^v_{p-S}(a-1,t))/\tau+(1-\delta(a))\Delta_N(a,t)R^v_{p-S}(a,t)/N(a,t)\end{aligned} \tag{56}$$

$$\begin{aligned}\dot{R}^v_{npS}(a,t) &= \Psi(1-f^{p-}_D-f^{p-}_F-f^{p-}_\mu)T^v_{np}(a,t)-r_S R^v_{npS}(a,t)-p(a)q\epsilon\lambda(a,t)R^v_{npS}(a,t) \\ &- (R^v_{npS}(a,t)-(1-\delta(a))R^v_{npS}(a-1,t))/\tau+(1-\delta(a))\Delta_N(a,t)R^v_{npS}(a,t)/N(a,t)\end{aligned} \tag{57}$$

$$\begin{aligned}\dot{R}^v_{p+D}(a,t) &= \Psi f^{p+}_D T^v_{p+}(a,t)-r_D R^v_{p+D}(a,t)-p(a)q\epsilon\lambda(a,t)R^v_{p+D}(a,t) \\ &- (R^v_{p+D}(a,t)-(1-\delta(a))R^v_{p+D}(a-1,t))/\tau+(1-\delta(a))\Delta_N(a,t)R^v_{p+D}(a,t)/N(a,t)\end{aligned} \tag{58}$$

$$\begin{aligned}\dot{R}^v_{p-D}(a,t) &= \Psi f^{p-}_D T^v_{p-}(a,t)-r_D R^v_{p-D}(a,t)-p(a)q\epsilon\lambda(a,t)R^v_{p-D}(a,t) \\ &- (R^v_{p-D}(a,t)-(1-\delta(a))R^v_{p-D}(a-1,t))/\tau+(1-\delta(a))\Delta_N(a,t)R^v_{p-D}(a,t)/N(a,t)\end{aligned} \tag{59}$$

$$\begin{aligned}\dot{R}^v_{npD}(a,t) &= \Psi f^{p-}_D T^v_{np}(a,t)-r_D R^v_{npD}(a,t)-p(a)q\epsilon\lambda(a,t)R^v_{npD}(a,t) \\ &- (R^v_{npD}(a,t)-(1-\delta(a))R^v_{npD}(a-1,t))/\tau+(1-\delta(a))\Delta_N(a,t)R^v_{npD}(a,t)/N(a,t)\end{aligned} \tag{60}$$

where $\delta(a)$ stands for the Kronecker delta function ($\delta(x=0)=1$ and $\delta(x\neq 0)=0$), and the five quantities that depend on time are the force of infection $\lambda(a,t)$, the proportion of mothers with TB $m_d(t)$, the diagnosis rate $d(t)$, the correction terms $\Delta_N(a,t)$ standing for any demographic variation in the population due to causes foreign to TB and aging and the vaccine coverage $c(a,t)$ simply because the vaccine is introduced at a certain moment t_{vac}.

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:

$$\begin{aligned}\dot{I}(a,t) &= \omega_f L_f(a,t)+\omega_s L_s(a,t)+r_N\left(R_{p+N}(a,t)+R_{p-N}(a,t)+R_{npN}(a,t)\right) \\ &+r_S\left(R_{p+S}(a,t)+R_{p-S}(a,t)+R_{npS}(a,t)\right)+r_D\left(R_{p+D}(a,t)+R_{p-D}(a,t)+R_{npD}(a,t)\right)\end{aligned} \tag{61}$$

$$\begin{aligned}\dot{M}(a,t) &= \mu_{p+}(D_{p+}(a,t)+F(a,t)+D^v_{p+}(a,t))+\Psi f^{p+}_\mu(T_{p+}(a,t)+T^v_{p+}(a,t)) \\ &+\mu_{p-}(D_{p-}(a,t)+D^v_{p-}(a,t))+\mu_{np}(D_{np}(a,t)+D^v_{np}(a,t)) \\ &+\Psi f^{p-}_\mu(T_{p-}(a,t)+T^v_{p-}(a,t)+T_{np}(a,t)+T^v_{np}(a,t))\end{aligned} \tag{62}$$

Because of that, 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:

$$i(t)=\frac{1000000\cdot\sum_a\left(I(a,t+1)-I(a,t)\right)}{(\mathcal{N}(t+1)+\mathcal{N}(t))/2} \tag{63}$$

$$m(t)=\frac{1000000\cdot\sum_a\left(M(a,t+1)-M(a,t)\right)}{(\mathcal{N}(t+1)+\mathcal{N}(t))/2} \tag{64}$$

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.

1.7 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.[1,2] 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 also abandon this initial stationarity assumption and 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 in the unvaccinated branch joined by individuals that, at least, have been infected with the bacillus once. Additionally, no individual starts at the vaccinated branch (thus $X^v(a, t = 0) = 0$), since any vaccination campaign is introduced at some time $t > 0$ and the BCG is considered implicitly in the unvaccinated branch (i.e. the protection levels provided by the old vaccine are treated as the background). Finally, we define a parameter $\varsigma \in [-1, 1]$, such as, when $\varsigma < 0$, the initial conditions correspond to a state with lower TB burden than that in the stationary state:

$$X(a, t = 0) = (1 + \varsigma) X^*(a, d_0, \beta_0, \vec{N}(a, 0)) \qquad \forall (X \neq S) \tag{65}$$

$$S(a, t = 0) = S^*(a, d_0, \beta_0, \vec{N}(a, 0)) \left(1 - \varsigma \sum_{X \neq S} X^*(a, d_0, \beta_0, \vec{N}(a, 0)) \right) \tag{66}$$

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

$$S(a, t = 0) = S^*(a, d_0, \beta_0, \vec{N}(a, 0))(1 - \varsigma) \tag{67}$$

$$X(a, t = 0) = X^*(a, d_0, \beta_0, \vec{N}(a, 0)) \left(1 + \frac{\varsigma S^*(a, d_0, \beta_0, \vec{N}(a, 0))}{\sum_{X \neq S} X^*(a, d_0, \beta_0, \vec{N}(a, 0))} \right) \tag{68}$$

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 Model calibration procedure

2.1 Global reproduction of aggregated TB burden levels

The first ingredient of the calibration procedure of our model consists of estimating in each region the initial conditions of the system, parameterized through the ς coordinate along with the diagnosis rate $d(t)$ and the scaled infectiousness $\beta(t)$, that are adequate to make the model reproduce aggregated TB burden measurements (i.e. mortality and incidence) on a certain time window, comprised, in this case, between $t_o = 2000$ to $t_F = 2015$.

Both parameters $d(t)$ and $\beta(t)$ are fitted to half-sigmoid-like curves, as follows:

$$d(t) = \begin{cases} d_0 + (d_{\sup} - d_0) t (t + \frac{1}{d_1})^{-1} & \text{if } d_1 > 0 \\ d_0 & \text{if } d_1 = 0 \\ d_0 - d_0 t (t - \frac{1}{d_1})^{-1} & \text{if } d_1 < 0 \end{cases} \tag{69}$$

$$\beta(t) = \begin{cases} \beta_0 + \beta_0 t \left(t + \frac{1}{\beta_1} \right)^{-1} & \beta_1 > 0 \\ \beta_0 & \beta_1 = 0 \\ \beta_0 - \beta_0 t \left(t - \frac{1}{\beta_1} \right)^{-1} & \beta_1 < 0 \end{cases} \tag{70}$$

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 the 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 $d_{\text{sup}} = 12.17y^{-1}$ 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.[15]

We have chosen this parameterization of the temporal evolution of the diagnosis rate and the infectiousness because, unlike previous approaches in which the evolution of these parameters is assimilated to an exponential curve,[1,2] it provides a bounded growth for them, with a function that is still continuous and differentiable but does not introduce more parameters. The goal of this part 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:

$$H = \sum_{t=t_o}^{t_F} \left(\frac{i(t) - \bar{i}(t)}{\bar{\Delta}_i(t)} \right)^2 + \left(\frac{m(t) - \bar{m}(t)}{\bar{\Delta}_m(t)} \right)^2 \tag{71}$$

where $\bar{i}(t)$ and $\bar{m}(t)$ stand for the annual incidence and mortality rates, calculated by grouping the national estimations available at the the WHO database for TB,[5] corresponding to the different countries within each region. 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 correspondent terms $\bar{\Delta}_i(t)$ $\bar{\Delta}_m(t)$ are constructed as follows:

$$\bar{\Delta}_i(t) = \begin{cases} \bar{i}(t) - \bar{i}^{\text{low}}(t) & \text{if } i(t) \leq \bar{i}(t) \\ \bar{i}^{\text{high}}(t) - \bar{i}(t) & \text{if } i(t) > \bar{i}(t) \end{cases} \tag{72}$$

$$\bar{\Delta}_m(t) = \begin{cases} \bar{m}(t) - \bar{m}^{\text{low}}(t) & \text{if } m(t) \leq \bar{m}(t) \\ \bar{m}^{\text{high}}(t) - \bar{m}(t) & \text{if } m(t) > \bar{m}(t) \end{cases} \tag{73}$$

The conceptual scheme for the fitting of these parameters (see figure S3) 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 is small enough. 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.[16] This iterative process is in turn nested in another loop aimed at reproducing the observed distribution of TB incident cases, as we discuss in the next section.

2.2 Detailed calibration of distributions of incident cases among age-groups and types of TB

In the previous section (2.1), we have illustrated the part of the calibration method used to reproduce the aggregated trends of TB burden, via the adjustment of initial conditions, diagnosis rate and infectiousness. However, this only step would not allow us to control the distribution of the burden in the different age groups, or to reproduce the fraction of cases associated to each type of TB within each age segment. This limitation, that constitutes an important handicap for the proper forecast and comparison of age-focused vaccination campaigns, motivates us to complement the aforementioned calibration approach, aimed at reproducing aggregated burden rates, with the following two additional procedures:

2.2.1 Fitting of fast-progression probabilities per age-group to reproduce TB incidence distribution across age groups

To ensure a proper description of the distribution of TB cases across age groups, we implemented an additional step at the calibration procedure based on the nested fitting of the probabilities of fast progression to disease as a function of age. The choice of this probability as the element to tune comes from the observation that it is arguably one of the aspects of the TB infection cycle whose inter-individual variation and age dependence has

been best studied in literature.[17,18] Admittedly, many epidemiological studies have focused on measuring this probability and its variation with age in epidemiological settings that correspond to different parts of the world and times, in places such as India[19] or the Netherlands[20]; often producing divergent estimations. Furthermore, the variation in the probability of fast progression to disease after infection has also been reported within single populations related to different levels of past exposure to the pathogen.[8] All these observations, even if they back up the idea of fast progression probabilities being heavily dependent on age, make adventurous to accept a single global estimation of this dependence to be used in our model. For example, while we could use the same global estimation of the primo-infection probabilities used in Abu-Raddad et al.[2], and use a risk of fast progression of 5% in children and 15% in adults in all the regions analyzed, other works have reported that the risk of disease is notably higher in children less than 2 years old, and drops drastically afterwards.[4,21] Considering all these issues, we opted for estimating fast-progression probabilities dependencies with age from a data-driven stand, so as to reproduce the proportion of active cases in the different age groups, as reported in the WHO TB Case Notifications Database.[5] The influence of this choice is addressed in the paper by comparing our model forecasts with those produced from a version of the model in which we use the values of this parameter in each group as in Abu-Raddad et al.[2].

Methodologically speaking, the obtention of the fraction of incident cases per age group from the WHO TB Case notifications database[5] for all the regions studied, poses a series of technical difficulties:

- The database is structured as a matrix where the rows represent the countries and years, and the columns represent age group, sex and type of disease. The value of every matrix entry gives us the number of notifications that fulfill the characteristics given its coordinates within the matrix -i.e. country, year, sex, age and type of disease-. In order to avoid bias due to missing data, we only considered country-year combinations (i.e. database's rows) when at least one case is reported in every column of the matrix. Then, using the populations of the countries for which useful information is retrieved for each year, we normalize the incidence rates, and obtain fractions of incident cases per age-group. Using this criterium to discard missing data, we retrieved a reliable estimation for the age distribution of incident cases world-wide between 2006 and 2012. While before 2006 the fraction of missing entries in the Database discouraged to extend the period further in past, after 2012, the notification criteria and the structure of the database changed, so we can not apply the same methodology.

- The age groups used in the case notifications database do not correspond exactly to the age groups used in the propagation model,[5] inherited from the UN population division Database.[14] This fact, in general, does not constitute a serious impediment, since a correspondence between both age groups schemes can be easily derived, as detailed in table 2. However, while the last group ($\alpha = 7$, see table 2) has no upper limit in the WHO TB database, in our model, the last age-group ($a = 13$, in table 2) corresponds to individuals between 65 and 70 years old. This disagreement would cause an over-estimation of the fraction of cases among individuals between 65 and 70 years old, if the proportion of cases reported in the WHO TB database for its last group $\alpha = 7$ was to be attributed to the last age group of our model ($a = 13$). In order to solve this issue, we opted for fitting the probability of fast progression $p(a)$ in order to reproduce the correct age distribution of cases for individuals under 65 years of age. Then, we assume, as a reasonable approximation, that $p(a = 13) = p(a = 12)$.

Considering all that, we define the proportions of cases of the age group α and the year y, as:

$$g(\alpha, y) = \frac{\sum_c \bar{i}(\alpha, c, y)}{\sum_c \sum_{\alpha=0}^{6} \bar{i}(\alpha, c, y)} = \frac{\bar{i}'(\alpha, y)}{\sum_{\alpha=0}^{6} \bar{i}'(\alpha, y)} \tag{74}$$

where $\bar{i}(\alpha, c, y)$ is the number of cases reported for age group α in the country c and the year y. To obtain the total fractions we perform a weighted average:

$$g(\alpha)^{\text{data}} = \frac{\sum_y \left(g(\alpha, y) \sum_\alpha \sum_{c \in \mathcal{C}(y)} N(\alpha, c, y) \right)}{\sum_\alpha \sum_y \sum_{c \in \mathcal{C}(y)} N(\alpha, c, y)} = \frac{\sum_y \left(g(\alpha, y) \sum_\alpha N'(\alpha, y) \right)}{\sum_\alpha \sum_y N'(\alpha, y)} \tag{75}$$

where $N(\alpha, c, y)$ is the population of the group age α, in the country c and the year y. $\mathcal{C}(y)$ stands for the set of countries of the region of interest that have been selected in year y. The notation $\bar{i}'$ or N' is referring to quantities aggregated to all those countries, while the sum over y covers the years between 2006 and 2012.

Model structure		Database structure	
Age	Index: a	Age	Index: α
0-4	0	0-4	0
5-9	1	5-14	1
10-14	2		
15-19	3	15-24	2
20-24	4		
25-29	5	25-34	3
30-34	6		
35-39	7	35-44	4
40-44	8		
45-49	9	45-54	5
50-54	10		
55-59	11	55-64	6
60-64	12		
65-69	13	+65	7

Table 2: Equivalence between age structures. While the UN database[14] and our model use 5 years age groups, the Case Notification Database[5], uses 10 years age groups except for the first age group (which corresponds to 0-5 years old)

Obviously, in parallel to the proportions $g(\alpha)^{\text{data}}$, we can equally get the equivalent fractions as produced by our model $g(\alpha)^{\text{model}}$. Indeed, the fast-progression probabilities per age $p(\alpha)$ are fitted so as to make the ratios $g(\alpha)^{\text{data}}/g(\alpha)^{\text{model}}$ to deviate, at most, a 0.5% from unity for every age-group. Until that condition is reached, the values $p(\alpha)$ corresponding to the age groups where the ratio $g(\alpha)^{\text{data}}/g(\alpha)^{\text{model}}$ is too far away from 1 are iterative multiplied by that same ratio in order to correct such deviations (see section 2.3 for further details).

Considering that, in order to provide an estimation for the confidence interval for $p(\alpha)$, we begin by estimating the uncertainties of $\bar{i}'(\alpha, y)$ assuming a multinomial distribution:

$$\Delta\bar{i}'(\alpha, y) = \sqrt{\bar{i}'(\alpha, y)\,(1 - g(\alpha, y))} \tag{76}$$

and we propagate these errors to obtain confidence intervals for $g(\alpha)$, whose width we will denote as $\Delta g(\alpha)^+$ and $\Delta g(\alpha)^-$, since they are, in general, asymmetric. Finally, we obtain the upper and lower limit of $p(\alpha)$ -called $p(\alpha)^{\text{up}}$ and $p(\alpha)^{\text{low}}$- from the following expressions:

$$p(\alpha)^{\text{up}} = p(\alpha)\frac{g(\alpha) + \Delta g(\alpha)^+}{g(\alpha)^{\text{model}}} \tag{77}$$

$$p(\alpha)^{\text{low}} = p(\alpha)\frac{g(\alpha) - \Delta g(\alpha)^-}{g(\alpha)^{\text{model}}} \tag{78}$$

The election of these values for $p(\alpha)^{\text{up}}$ and $p(\alpha)^{\text{low}}$, considering the approach used to estimate $p(\alpha)$, intends to qualitatively capture the effect on these probabilities of the deviations from the central values in the fractions of incident cases reported by the database, or at least to scale with such deviations.

2.2.2 Estimation of age-dependent probabilities of TB types from case notifications data

The WHO case-notifications database, beyond reporting the age distributions of incident cases, also provides useful information about how these cases are distributed among the different types of TB -pulmonary smear positive, smear negative or non-pulmonary-, also as a function of age. Once again, incorporating the most reliable estimations about the distribution of TB cases among different types of disease, and how this distribution changes with patients' age constitute a fundamental priority from a modeling stand, if we aim at evaluating age-focused epidemiological interventions. In this spirit, instead of accepting global and/or age-independent estimates to describe the probabilities of suffering each type of disease upon transition to active TB, we make use of the information reported at the WHO TB Database to produce data-driven detailed estimates of these probabilities.

The database reports the division of the notified cases among the three disease types. As a first approximation, it would appear that the probabilities of developing each TB type should simply correspond to the

proportion of cases of each type. However, since we have different diagnosis rates associated to each disease type, we have to rescale these numbers of cases with the parameter η (that represents the ratio between the diagnosis rate of smear negative TB with respect to smear positive or non-pulmonary TB). This yields the following definitions of the probabilities to develop smear positive TB ρ_{p+}, smear negative TB ρ_{p-}, or non-pulmonary disease ρ_{np}:

$$\rho_{p+}(\alpha, y) = \frac{d_{p+}(\alpha, y)}{d_{p+}(\alpha, y) + d_{p-}(\alpha, y) + d_{np}(\alpha, y)} \tag{79}$$

$$\rho_{p-}(\alpha, y) = \frac{d_{p-}(\alpha, y)}{d_{p+}(\alpha, y) + d_{p-}(\alpha, y) + d_{np}(\alpha, y)} \tag{80}$$

$$\rho_{np}(\alpha, y) = \frac{d_{np}(\alpha, y)}{d_{p+}(\alpha, y) + d_{p-}(\alpha, y) + d_{np}(\alpha, y)} \tag{81}$$

where:

$$d_{p+}(\alpha, y) = \bar{i}'_{p+}(\alpha, y) \tag{82}$$

$$d_{p-}(\alpha, y) = \frac{\bar{i}'_{p-}(\alpha, y)}{\eta} \tag{83}$$

$$d_{np}(\alpha, y) = \frac{\bar{i}'_{np}(\alpha, y)}{\eta} \tag{84}$$

Finally, we estimate the uncertainty of the incidence reported for TB type x, age group α and time t: $\bar{i}'_x(\alpha, y)$, by assuming a multinomial distribution for the different types of TB:

$$\bar{i}'^{\text{up}}_x(\alpha, y) = \bar{i}'_x(\alpha, y) + \sqrt{\left(1 - \frac{\bar{i}'_x(\alpha, y)}{\sum_x \bar{i}'_x(\alpha, y)}\right) \bar{i}'_x(\alpha, y)} \tag{85}$$

$$\bar{i}'^{\text{low}}_x(\alpha, y) = \bar{i}'_x(\alpha, y) - \sqrt{\left(1 - \frac{\bar{i}'_x(\alpha, y)}{\sum_x \bar{i}'_x(\alpha, y)}\right) \bar{i}'_x(\alpha, y)} \tag{86}$$

where x refers to the type of TB $x = \{p+, p-, np\}$. Then we obtain the errors of $\rho_{p+}(\alpha, y)$,$\rho_{p-}(\alpha, y)$ and $\rho_{np}(\alpha, y)$, by propagating 4 independent sources of error, $\bar{i}'_{p+}(\alpha, y)$, $\bar{i}'_{p-}(\alpha, y)$, $\bar{i}'_{np}(\alpha, y)$ and η. Lastly, it is worth mentioning that the outcomes for the age groups from 25 years old are merged.

Finally, we perform a weighted average of $\rho_x(\alpha, y)$ between 2006 and 2012, similar to equation 75, in order to calculate the final value of the probability of type of disease (ρ_x) that we will consider in our model. The Confidence Interval is chosen as the range between the minimum lower estimate and the maximum upper estimate of the $\rho_x(\alpha, y)$ in the time window considered.

2.3 Calibration algorithm summary

The fitting algorithm is the following:

1. We choose initial values of $p(\alpha)$ as those used in Abu-Raddad et al.[2]
2. We choose initial values of $(d_o, d_1, \beta_o, \beta_1, \varsigma)$.
3. Fixing $d(t) = d_o$, $\beta(t) = \beta_o$ and $N(a, t) = N(a, t = 2000)$ we run the model until the stationary equilibrium is reached, and register its coordinates: $\vec{X}^*(a, d_0, \beta_0, \vec{N}(a, 0))$.
4. Using ς, which describes the deviations from the stationary, we get to estimate the initial conditions of the system (see section 1.7): $\vec{X}(a, t_o = 2000) = f(\vec{X}^*, \varsigma)$
5. Run dynamics and evaluate the error function H (see eq. 71). If H is small enough, the algorithm continues in the next step; otherwise, parameters $(d_o, d_1, \beta_o, \beta_1, \varsigma)$ are modified according to Levenberg-Marquardt algorithm, and steps 3-5 are repeated.

6. Once step 5 produces an error function small enough (i.e. the Levenberg Marquardt algorithm for fitting aggregated burden converged), we calculate the proportion of cases given by the model: $g(\alpha)^{\text{model}}$. We compare the age distribution of the modelled burden proportions $g(\alpha)^{\text{model}}$ to the ones observed in the Database $g(\alpha)^{\text{data}}$ by calculating the ratio $\frac{g(\alpha)^{\text{data}}}{g(\alpha)^{\text{model}}}$:

   (a) If it is constrained between 0.995 and 1.005, for alll $\alpha \in [0, 6]$, we accept the incidence distribution obtained and the calibration concludes.

   (b) Otherwise, we update the vector $p(\alpha)$ for $\alpha \in [0, 6]$, according to the following scheme:

$$p(\alpha)^{\text{new}} = p_\alpha \frac{g(\alpha)^{\text{data}}}{g(\alpha)^{\text{model}}} \tag{87}$$

   and then, assume that $p(\alpha = 7) = p(\alpha = 6)$ (or, equivalently $p(a = 13) = p(a = 12)$), and iterate the steps 3-6 until final convergence.

The entire procedure is schematized in figure S3.

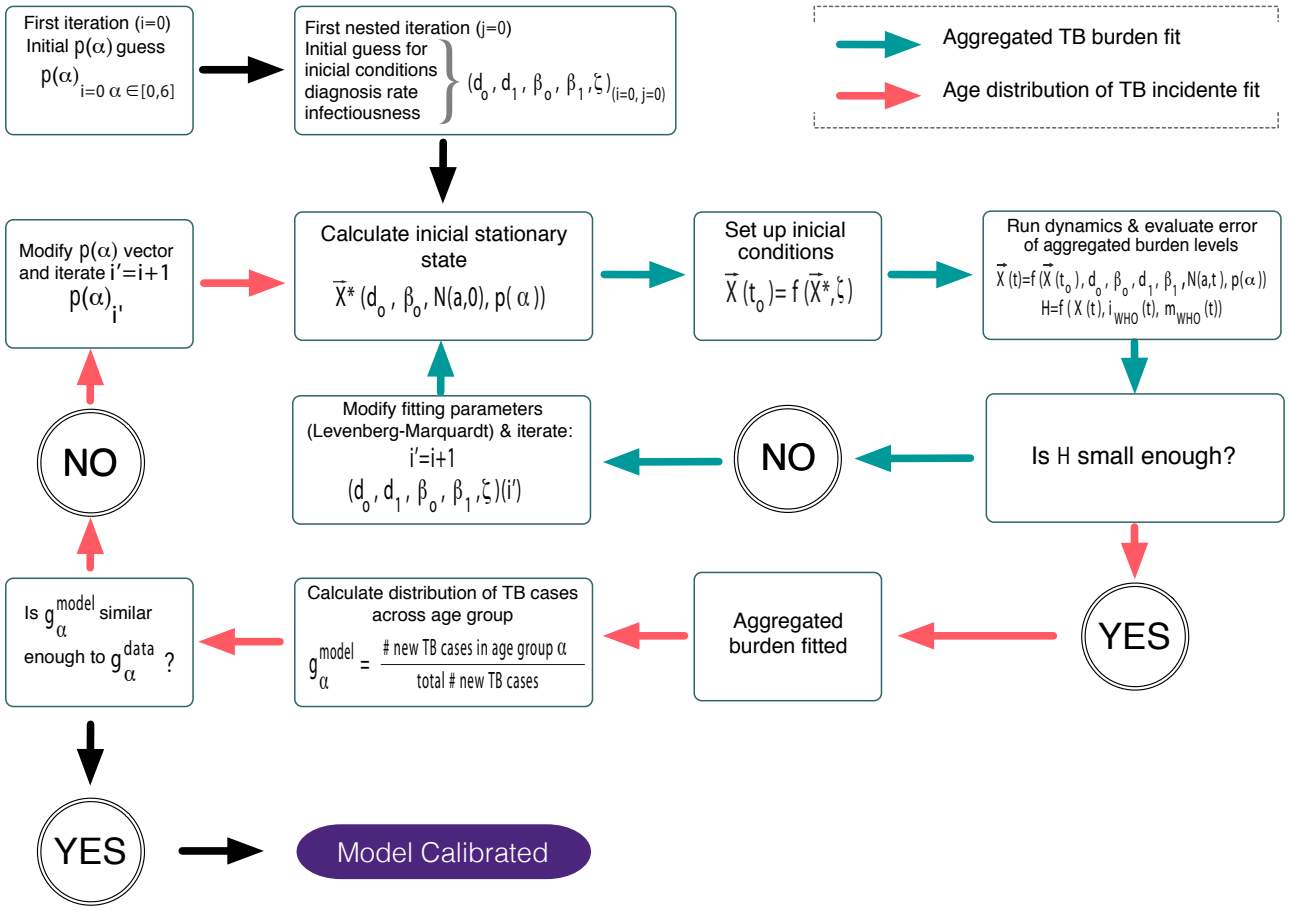

Figure S3: Schematic representation of the calibration algorithm.

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

Table 3: Description of the different dynamic states considered on the model

3.2 Global parameters

In table 4 we represent the global parameters of the disease. The values of these parameters are region independent, and also age independent.

Meaning	Parameter	Value	Confidence interval	Reference
Rate of fast progression (y^{-1})	ω_f	0.900	(0.765,1.035)	Abu-Raddad et al.[2]
Rate of slow progression (y^{-1})	ω_s	7.500×10^{-4}	$(6.375,8.625)\times10^{-4}$	Abu-Raddad et al.[2]
Mortality rate by pulmonary smear positive TB (y^{-1})	μ_{p+}	0.250	(0.213,0.288)	Abu-Raddad et al.[2]
Mortality rate by pulmonary smear negative TB (y^{-1})	μ_{p-}	0.100	(0.085,0.115)	Abu-Raddad et al.[2]
Mortality rate by non-pulmonary TB (y^{-1})	μ_{np}	0.100	(0.085,0.115)	Abu-Raddad et al.[2]
Reduction of infection risk for previously infected individuals	q	0.650	(0.553,748)	Abu-Raddad et al.[2]
Treatment completion rate (y^{-1})	Ψ	2.00	(1.70,2.30)	Abu-Raddad et al.[2]
Smear progression rate (y^{-1})	θ	0.015	(0.007,0.020)	Dye et al.[1]
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.[6], this work
Relapse rate for individuals who defaulted treatment (y^{-1})	r_D	3.774×10^{-3}	$(1.354,8.620)\times10^{-3}$	Korenromp et al.[6], Picon et al.[7],this work
Relapse rate for naturally recovered individuals (y^{-1})	r_N	0.030	(0.020,0.040)	Dye et al.[1]
Natural recovery rate (y^{-1})	ν	0.100	(0.085,0.115)	Dye et al.[1]
Infectivity reduction coefficient of D_{p-} with respect to D_{p+}	ϕ_{p-}	0.250	(0.213,0.288)	Abu-Raddad et al.[2]
Infectiousness reduction coefficient of R_{p+D} with respect to D_{p+}	ϕ_D	0.500	(0.250,0.750)	Dye et al.[1]
Proportion of mothers that infect their newborn children	m_t	0.150	(0.100,0.200)	Pillay et al.[9]

Table 4: Global parameters

3.3 Age-independent regional parameters

The following parameters take different values in each region:

- η: Coefficient of modification of the diagnosis rate for smear positive and extrapulmonary TB types with respect to smear positive (see section 1.1.4).[2]
- $(f_D^{p+}, f_F^{p+}, f_\mu^{p+})$: fraction of default, failure and death outcomes for smear positive pulmonary TB.[5]
- $(f_D^{p-}, f_F^{p-}, f_\mu^{p-})$: fraction of default, failure and death outcomes for smear negative pulmonary and non pulmonary TB.[5]

and their age-independent values, for each of the regions analyzed, are listed in table 5.

Parameter	AFRH	AFRL	EMR	SEAR	WPR
η	0.843 (0.717,0.970)	0.510 (0.434,0.587)	1.178 (1.001,1.354)	0.797 (0.677,0.916)	0.641 (0.545,0.737)
f_D^{p+}	0.081(0.045,0.118)	0.099 (0.044,0.154)	0.067 (0.038,0.095)	0.061 (0.040,0.077)	0.013 (0.008,0.018)
f_F^{p+}	0.016 (0.006,0.026)	0.024 (0.013,0.035)	0.016 (0.013,0.018)	0.019 (0.015,0.021)	0.009 (0.006,0.011)
f_μ^{p+}	0.063(0.023,0.085)	0.061 (0.048,0.074)	0.036 (0.033,0.039)	0.042 (0.032,0.049)	0.017 (0.014,0.020)
f_D^{p-}	0.079 (0.042,0.116)	0.076 (0.044,0.108)	0.076 (0.046,0.107)	0.062 (0.047,0.075)	0.017 (0.008,0.026)
f_F^{p-}	0.003 (0.001,0.006)	0.004 (0.001,0.007)	0.004 (0.001,0.007)	0.005 (0.002,0.009)	0.002 (0.002,0.003)
f_μ^{p-}	0.087 (0.054,0.120)	0.096 (0.055,0.137)	0.033 (0.024,0.042)	0.034 (0.032,0.035)	0.014 (0.011,0.017)

Table 5: Values of the regional parameters with the correspondent confidence intervals.

3.4 Diagnosis rate, infectiousness and initial conditions

Once all the aforementioned parameters are fixed, and as a first result of the calibration procedure explained in section 2, we obtain the time-evolving parameterizations of diagnosis rates and infectiousness, along with the initial conditions of the system in each region. In Figures S4 and S5 we represent the temporal evolution of $d(t)$ and $\beta(t)$ respectively. These temporal evolutions are derived from equations 69 and 70, where the fitted values of the parameters $(d_o, d_1, \beta_o, \beta_1)$ are reported in table 6.

Region	d_0 (y^{-1})	$d_1 \times 10^{-3}$ (y^{-1})	β_0 (y^{-1})	$\beta_1 \times 10^{-3}$ (y^{-1})	ς
AFRH	0.20 (0.10,0.28)	1.65 (1.39,2.22)	14.03 (11.23,17.14)	-27.58 (-46.06,-18.55)	-0.31 (-0.71,-0.03)
AFRL	0.61 (0.48,0.75)	2.70 (1.70,4.25)	33.97 (25.57,43.57)	-27.52 (-33.94,-19.56)	0.00 (-0.05,0.05)
EMR	0.32 (0.24,0.40)	6.50 (2.96,10.60)	24.38 (19.72,29.39)	0.51 (-13.65,26.21)	-0.29 (-0.62,0.21)
SEAR	0.78 (0.33,1.33)	3.77 (2.32,6.69)	4.50 (2.82,5.87)	-18.88 (-31.20,-1.83)	-0.07 (-0.13,0.01)
WPR	1.83 (0.70,2.27)	16.41 (12.36,23.23)	3.46 (2.74,4.24)	-1.49 (-13.98,24.56)	0.00 (0.00,0.02)

Table 6: Fitted parameters for each region.

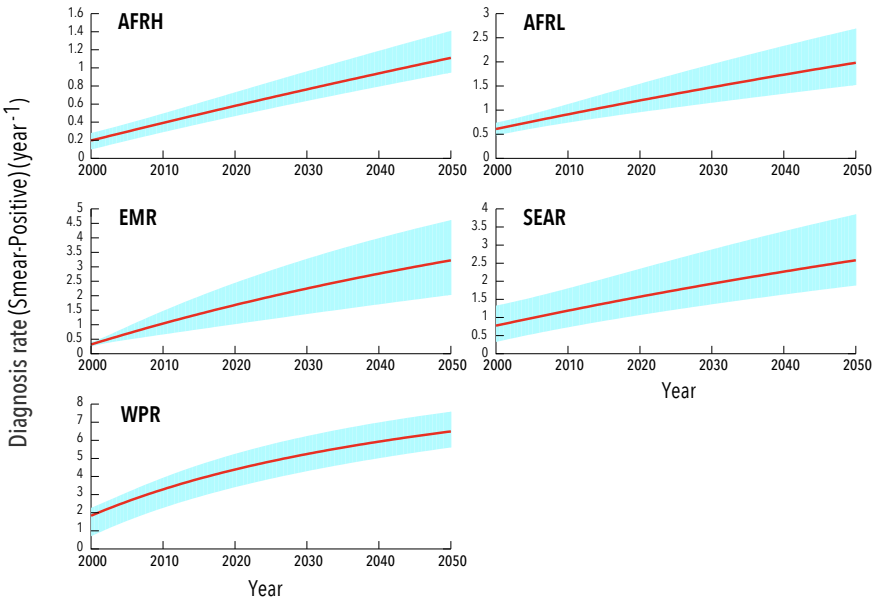

Figure S4: Diagnosis rate for the different regions. The red curve shows the result given by the central fit, while the blue area represents the uncertainty obtained as described in section 4

.

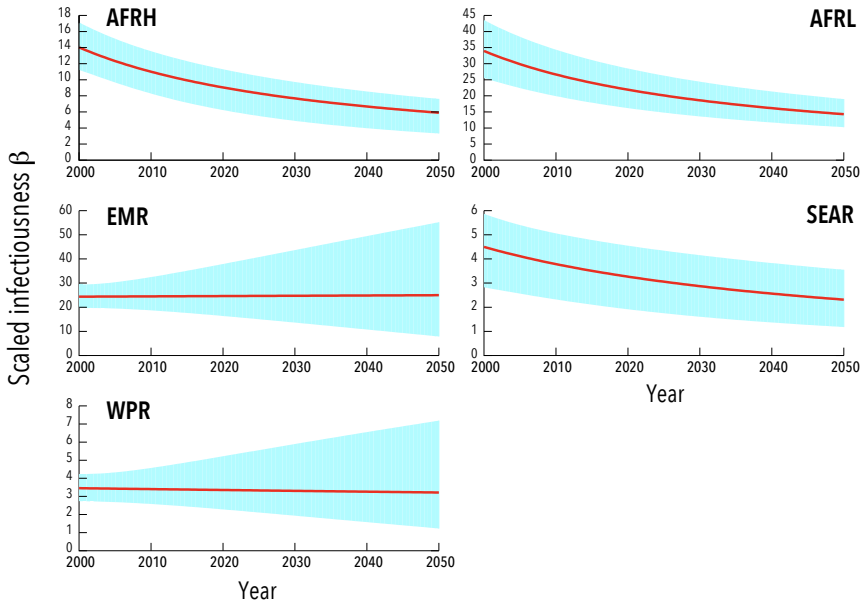

Figure S5: Scaled infectivity for the different regions. The red curve shows the result given by the central fit, while the blue area represents the uncertainty obtained as described in section 4

3.5 Fast progression probabilities

	Age	p	Confidence interval
AFRH	0-4	0.036	(0.030,0.059)
	5-14	0.006	(0.006,0.008)
	15-24	0.035	(0.031,0.040)
	25-34	0.109	(0.102,0.112)
	35-44	0.135	(0.121,0.143)
	45-54	0.120	(0.111,0.123)
	55-64	0.102	(0.077,0.118)
	65-69	0.102	(0.077,0.118)
AFRL	0-4	0.013	(0.006,0.036)
	5-14	0.009	(0.006,0.015)
	15-24	0.036	(0.034,0.039)
	25-34	0.081	(0.076,0.097)
	35-44	0.126	(0.097,0.132)
	45-54	0.122	(0.096,0.129)
	55-64	0.117	(0.090,0.203)
	65-69	0.117	(0.090,0.203)
EMR	0-4	0.005	(0.004,0.008)
	5-14	0.016	(0.014,0.022)
	15-24	0.041	(0.037,0.044)
	25-34	0.053	(0.051,0.054)
	35-44	0.047	(0.044,0.050)
	45-54	0.071	(0.065,0.075)
	55-64	0.120	(0.102,0.130)
	65-69	0.120	(0.102,0.130)

	Age	p	Confidence interval
SEAR	0-4	0.034	(0.026,0.053)
	5-14	0.013	(0.012,0.018)
	15-24	0.050	(0.048,0.053)
	25-34	0.073	(0.071,0.079)
	35-44	0.083	(0.081,0.085)
	45-54	0.126	(0.120,0.129)
	55-64	0.182	(0.173,0.190)
	65-69	0.182	(0.173,0.190)
WPR	0-4	0.002	(0.002,0.006)
	5-14	0.007	(0.006,0.012)
	15-24	0.173	(0.127,0.181)
	25-34	0.160	(0.156,0.208)
	35-44	0.148	(0.137,0.153)
	45-54	0.243	(0.227,0.246)
	55-64	0.430	(0.296,0.452)
	65-69	0.430	(0.296,0.452)

Table 7: Fast progression probabilities for each age group and region.

Regarding the probabilities of fast progression as a function of age, these are fitted to reproduce the observed distribution of incident cases across age, as explained in section 2. The estimated probabilities of fast progression per age group are reported in table 7. Even if we fitted these fast progression probabilities per age group assuming no constraint, it is worth noticing that the accepted trend of children younger than 15 years old having significantly lower risks of fast progression than adults is reproduced by our approach.

3.6 Probabilities of developing each type of TB

Following the procedure detailed in section 2.2.2, we obtained the values for the probabilities of developing each type of disease that are shown in table 8. We introduce only two independent parameters in our model: $\rho_{p+}(a)$ and $\rho_{np}(a)$. The third probability is computed as $\rho_{p-}(a) = 1 - \rho_{p+}(a) - \rho_{np}(a)$.

	Age	ρ_{p+}	ρ_{p-}	ρ_{np}
AFRH	0-4	0.033 (0.018,0.093)	0.892 (0.805,0.939)	0.075 (0.042,0.158)
	5-14	0.206 (0.151,0.274)	0.586 (0.465,0.680)	0.208 (0.142,0.342)
	15-24	0.557 (0.409,0.695)	0.294 (0.180,0.362)	0.150 (0.119,0.236)
	25-69	0.489 (0.402,0.620)	0.345 (0.221,0.410)	0.164 (0.122,0.228)
AFRL	0-4	0.291 (0.098,0.888)	0.544 (0.082,0.729)	0.164 (0.020,0.207)
	5-14	0.475 (0.319,0.760)	0.289 (0.156,0.435)	0.237 (0.057,0.329)
	15-24	0.781 (0.684,0.967)	0.152 (0.017,0.221)	0.067 (0.007,0.112)
	25-69	0.818 (0659,0.955)	0.105 (0.029,0.224)	0.073 (0.007,0.117)
EMR	0-4	0.053 (0.023,0.100)	0.496 (0.280,0.647)	0.452 (0.270,0.689)
	5-14	0.116 (0.077,0.172)	0.225 (0.118,0.326)	0.659 (0.518,0.792)
	15-24	0.354 (0.251,0.469)	0.211 (0.114,0.314)	0.434 (0.361,0.509)
	25-69	0.479 (0.392,0.572)	0.159 (0.135,0.196)	0.358 (0.291,0.437)
SEAR	0-4	0.024 (0.010,0.053)	0.930 (0.912,0.947)	0.046 (0.021,0.060)
	5-14	0.157 (0.126,0.205)	0.654 (0.585,0.787)	0.188 (0.083,0.253)
	15-24	0.658 (0.595,0.714)	0.239 (0.191,0.335)	0.103 (0.038,0.146)
	25-69	0.685 (0.618,0.731)	0.249 (0.201,0.341)	0.065 (0.016,0.092)
WPR	0-4	0.348 (0.045,0.492)	0.592 (0.432,0.724)	0.060 (0.015,0.353)
	5-14	0.330 (0.203,0.439)	0.615 (0.393,0.675)	0.055 (0.012,0.260)
	15-24	0.510 (0.289,0.612)	0.477 (0.384,0.586)	0.013 (0.004,0.151)
	25-69	0.566 (0.404,0.657)	0.416 (0.335,0.527)	0.017 (0.006,0.153)

Table 8: Probability of developing each type of TB for each age group and region (Types of TB: Pulmonary, Smear-positive ($p+$); Pulmonary, Smear-negative ($p-$); Non-pulmonary (np))

3.7 Vaccine descriptors

In principle, a vaccine could modify the disease spreading dynamics at different stages of the Natural History. For example, BCG itself is known to be capable of intervene by protecting either from infection or from progression to disease.[22,23] However, as our goal in this work is to better deal with age dependencies in current TB modeling, and not to explore the landscape of possibilities that a vaccine could provide, we focused on vaccines that act exclusively on the reduction of the infection rate. Therefore the only difference between vaccinated and unvaccinated individuals is that the former experiences a reduction in the force of infection $\lambda(a,t)$ by a factor ϵ with respect to the latter, as we have explained throughout section 1.1. In this work, we are considering an efficacy of 80% which means $\epsilon = 0.2$

However, we introduce an state-of-the-art framework that potentially allows to model more complex vaccines, which could be achieved simply by introducing more changes in the vaccinated branch of the Natural History, modifying the values of the different parameters involved. We will explore this possibility in future works.

The vaccination campaign –i.e. the progression of individuals from the unvaccinated to the vaccinated branch– is described by the parameter $c(a,t) \in [0,1]$, introduced in sections 1.1.10 and 1.4, that accounts for

the fraction of individuals that are covered by the vaccine when reaching the age group a at a certain time t. We are considering vaccination campaigns starting in 2025.

We explore two different vaccination strategies that depend on which age groups are targeted: newborn vaccination –mimicking a BCG-substitutive vaccine– or adolescent vaccination –associated to a BCG-booster vaccine–. In the first case, individuals are vaccinated at birth (therefore reaching $a = 0$) while in the second scenario, individuals are vaccinated when they turn 15 years old ($a = 3$). Therefore, the two vaccine campaigns considered in this work can be described as follows:

- Substitutive vaccine (Newborns):

$$c(a,t) = \begin{cases} 0 & \text{if } t < t_v \text{ OR } a \neq 0 \\ 1 & \text{if } t \geq t_v \text{ AND } a = 0 \end{cases} \tag{88}$$

- Booster vaccine (Adolescents):

$$c(a,t) = \begin{cases} 0 & \text{if } t < t_v \text{ OR } a \neq 3 \\ 1 & \text{if } t \geq t_v \text{ AND } a = 3 \end{cases} \tag{89}$$

Other properties of the vaccines such as immunity waning effects, the appearance of masking and/or blocking effects,[24] and different levels of coverage have still not been considered in this work, but will be taken into account in future works.

4 Model uncertainty and sensitivity analysis

The model parameters –both regional and global–, as well as the demographic data and the TB burden estimations 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, 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, including the global parameters (section 3.2), the parameter η, the probability of fast progression $p(a)$ and the probabilities of types of disease $\rho_{p+}(a)$ and $\rho_{np}(a)$, each of them conservatively treated as if they were totally independent uncertainty sources $u_i, i \in [1, 19]$.
- Burden and treatment outcomes estimations provided by WHO: 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, and considered them as one uncertainty source, labeled as u_{20}.
- Demographic structures $N(a,t)$: which are also considered as a single uncertainty source, labeled as u_{21}.
- Contact matrix $\xi(a,a')$; 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 measurement x can be evaluated by defining the variables $\Delta_i^{\text{low}}(x) = x(u_1, ..., u_i^{\text{low}}, ..., u_{22}) - x(u_1, ..., u_i, ..., u_{22})$ and $\Delta_i^{\text{up}}(x) = x(u_1, ..., u_i^{\text{up}}, ..., u_{22}) - x(u_1, ..., u_i, ..., u_{22})$, that represent the differences of the measurement x (an incidence or mortality rate, either temporal or accumulated, for example), evaluated after varying only the uncertainty source u_i, which takes the values of the lower and upper limits of its confidence interval u_i^{low} and u_i^{up}, respectively.

At this point, two relevant aspects are worth mentioning. On one hand, each uncertainty source is not necessarily a single parameter, but could be a group of them of related origins, whose confidence intervals plausibly carry strong correlations. For example, some of the parameters are age-dependent, WHO estimations

comprise several measurements of different nature (treatment outcomes fractions, mortalities and incidence rates), and either demographic structures or contact matrices are multidimensional objects too. This means that, when evaluating, for example $x(u_1, ..., u_i^{\text{low}}, ..., u_{22})$, all the "components" of u_i (i.e. the value of a parameter on all ages, or all the entries of the contact matrix, for example) are simultaneously evaluated, for the sake of simplicity, at the bottom of their confidence intervals.

On the other hand, the evaluation of $x(u_1, ..., u_i^{\text{low}}, ..., u_{22})$ or $x(u_1, ..., u_i^{\text{up}}, ..., u_{22})$ implies the repetition of the fitting procedure after varying u_i, so as to address the influence of the error associated to the input represented by u_i on the reliability of the whole fitting and forecasting procedure of our model. Again, for simplicity, since most of the uncertainty sources act in block over all age groups, only the inner loop in figure S3 is repeated upon a shift of each uncertainty source u_i from its central value to its extreme values u_i^{low} and u_i^{up}. This implies that, in the context of uncertainty/sensitivity estimations we neither re-evaluate the probabilities of fast progression nor re-calibrate the incidence distribution per age group upon shifting each uncertainty source. Instead, as a first approximation, we only evaluate the effect of the different uncertainty sources on the reproduction of aggregated burden levels, and include, as a conservative assumption, the fast-progression probabilities as an additional, independent source of error to propagate on model forecasts.

Taking that into consideration, we can evaluate the uncertainty in measure x that is due to the uncertainty of the parameters reported in the bibliography. In order to do so, we have to sum up the contributions to the error of the first uncertainty source –the Natural History parameters–, by performing the following partial sums:

$$\Delta_x^{\text{param,up}} = \sqrt{\sum_1^{19} \delta\left(\frac{|d_i^{\text{low}}(x)|}{d_i^{\text{low}}(x)} - 1\right) d_i^{\text{low}}(x)^2 + \sum_1^{19} \delta\left(\frac{|d_i^{\text{up}}(x)|}{d_i^{\text{up}}(x)} - 1\right) d_i^{\text{up}}(x)^2} \tag{90}$$

$$\Delta_x^{\text{param,low}} = \sqrt{\sum_1^{19} \delta\left(\frac{|d_i^{\text{low}}(x)|}{d_i^{\text{low}}(x)} + 1\right) d_i^{\text{low}}(x)^2 + \sum_1^{19} \delta\left(\frac{|d_i^{\text{up}}(x)|}{d_i^{\text{up}}(x)} + 1\right) d_i^{\text{up}}(x)^2} \tag{91}$$

and, in a similar way, we can isolate the contribution to the model uncertainty when determining x of the other uncertainty sources as follows:

$$\Delta_x^{\text{WHO,up}} = \sqrt{\delta\left(\frac{|d_{20}^{\text{low}}(x)|}{d_{20}^{\text{low}}(x)} - 1\right) d_{20}^{\text{low}}(x)^2 + \delta\left(\frac{|d_{20}^{\text{up}}(x)|}{d_{20}^{\text{up}}(x)} - 1\right) d_{20}^{\text{up}}(x)^2} \tag{92}$$

$$\Delta_x^{\text{WHO,low}} = \sqrt{\delta\left(\frac{|d_{20}^{\text{low}}(x)|}{d_{20}^{\text{low}}(x)} + 1\right) d_{20}^{\text{low}}(x)^2 + \delta\left(\frac{|d_{20}^{\text{up}}(x)|}{d_{20}^{\text{up}}(x)} + 1\right) d_{20}^{\text{up}}(x)^2} \tag{93}$$

$$\Delta_x^{\text{demo,up}} = \sqrt{\delta\left(\frac{|d_{21}^{\text{low}}(x)|}{d_{21}^{\text{low}}(x)} - 1\right) d_{21}^{\text{low}}(x)^2 + \delta\left(\frac{|d_{21}^{\text{up}}(x)|}{d_{21}^{\text{up}}(x)} - 1\right) d_{21}^{\text{up}}(x)^2} \tag{94}$$

$$\Delta_x^{\text{demo,low}} = \sqrt{\delta\left(\frac{|d_{21}^{\text{low}}(x)|}{d_{21}^{\text{low}}(x)} + 1\right) d_{21}^{\text{low}}(x)^2 + \delta\left(\frac{|d_{21}^{\text{up}}(x)|}{d_{21}^{\text{up}}(x)} + 1\right) d_{21}^{\text{up}}(x)^2} \tag{95}$$

$$\Delta_x^{\text{contacts,up}} = \sqrt{\delta\left(\frac{|d_{22}^{\text{low}}(x)|}{d_{22}^{\text{low}}(x)} - 1\right) d_{22}^{\text{low}}(x)^2 + \delta\left(\frac{|d_{22}^{\text{up}}(x)|}{d_{22}^{\text{up}}(x)} - 1\right) d_{22}^{\text{up}}(x)^2} \tag{96}$$

$$\Delta_x^{\text{contacts,low}} = \sqrt{\delta\left(\frac{|d_{22}^{\text{low}}(x)|}{d_{22}^{\text{low}}(x)} + 1\right) d_{22}^{\text{low}}(x)^2 + \delta\left(\frac{|d_{22}^{\text{up}}(x)|}{d_{22}^{\text{up}}(x)} + 1\right) d_{22}^{\text{up}}(x)^2} \tag{97}$$

Throughout this work, any area around a curve –representing an uncertainty–, any error bar or any Confidence Interval referred to an outcome of our model is calculated by summing up all the contributions:

$$\Delta_x^{\text{low}} = \sqrt{(\Delta_x^{\text{param,low}})^2 + (\Delta_x^{\text{WHO,low}})^2 + (\Delta_x^{\text{demo,low}})^2 + (\Delta_x^{\text{contacts,low}})^2} \tag{98}$$

$$\Delta_x^{\text{up}} = \sqrt{(\Delta_x^{\text{param,up}})^2 + (\Delta_x^{\text{WHO,up}})^2 + (\Delta_x^{\text{demo,up}})^2 + (\Delta_x^{\text{contacts,up}})^2} \tag{99}$$

In figure S7, we represent the total amount of error due to two main uncertainty sources -Natural History parameters, WHO estimations – to the total model-based estimations of aggregated burden rates:

- Bibliographic parameters contribution, red area: $(\Delta_x^{\text{param,low}}, \Delta_x^{\text{param,up}})$
- WHO contribution, salmon area: $\left(\left(\sqrt{(\Delta_x^{\text{param,low}})^2 + (\Delta_x^{\text{WHO,low}})^2} - \sqrt{(\Delta_x^{\text{param,low}})^2}\right), \left(\sqrt{(\Delta_x^{\text{param,up}})^2 + (\Delta_x^{\text{WHO,up}})^2} - \sqrt{(\Delta_x^{\text{param,up}})^2}\right)\right)$

The contribution to the total error bars due to the uncertainty in the contact matrix patterns and the demographic evolution is negligible when compared to the other two. Besides, in figure S8 error bars for model-based fractions of incident cases in each age groups have also estimated following the aforementioned procedure.

As a final aspect to consider, we note that variables such as the incidence and the mortality (or their age distributions) only require one simulation to be calculated. Thus, in order to estimate their uncertainties, we perform 44 additional simulations corresponding to the 2 limits –low and high– of the confidence interval of the 22 uncertainty sources that we have isolated. This forces us to repeat the fitting for each one of these iterations (inner loop in figure S3) as many times, which ultimately allows us to calculate the uncertainties for the fitted parameters as shown in figures S4 and S5.

We also compute some quantities that need two or more different running settings of our model to be defined. For example, we define the impact of a vaccine as the difference of the accumulated incidence –or mortality– between the model with and without vaccine –and therefore it requires both implementations of the model–. The calculation of the uncertainties of these quantities follows the same scheme: we compute these quantities for each uncertainty iteration independently and we propagate the errors as explained before.

In the case of the calculation of vaccine impacts the fact that two runs are needed to obtain them (i.e. reference run vs. vaccine run) does not imply that the entire fitting procedure has to be repeated too, because the existence or not of a vaccination campaign does not affect the fitting –as we implement vaccination always beyond the fitting window–. However, we define other quantities that do require more fittings, specifically model outcomes comparisons between our model and the reduced versions of it that have their own independent fitting process.

When obtaining differential measurements based on comparisons between two model runs (e.g. vaccine impacts or model-to-model comparisons) and, more specifically, when obtaining the uncertainties of such differential measures, it is important to bear in mind that, since a large fraction of the uncertainties of the two models being compared comes from the same sources, the uncertainties of the two measures being compared are tightly correlated. A consequence of this fact is that the difference of two quantities –e.g. the impacts provide by a substitutive and a booster vaccines– with (even largely) overlapping confidence intervals might still be significant –i.e. may have a confidence interval that do not cross zero–. This is due to the fact that the Confidence Intervals of the two quantities in question are not, generally, independent: when a deviation of the central value of a given uncertainty source causes a deviation from the expected value in –let us say– model 1's forecast for a given measurement, it most likely will cause a shift in the same direction in model 2's estimate. This significantly reduces the variance of the differences between the two models, enhancing our ability to detect statistically significant differences between models as well as vaccines' impacts.

4.2 Model sensitivity

Beyond distinguishing the amount of uncertainty derived from each type of input, we have to isolate the model sensitivity to each of the parameters. More specifically, we estimate the sensitivities ζ_x of model measurement x to the uncertainty source u_i as follows:

$$\zeta_x^{\text{up}}(u_i) = 100 \cdot \frac{\Delta_x^{i,\text{up}}}{x(\{u_i\})} \tag{100}$$

$$\zeta_x^{\text{low}}(u_i) = 100 \cdot \frac{\Delta_x^{i,\text{low}}}{x(\{u_i\})} \tag{101}$$

where $x(\{u_i\})$ represents the value of measurement x when all the inputs u_i are evaluated at their central, expected values, and $(\Delta_x^{i,\text{up}}, \Delta_x^{i,\text{low}})$ represent the deviations registered in variable x when the i-th source of uncertainty is substituted by the lower/upper value of its confidence interval. In figure S6, we show the sensitivities of the total incidence $(\zeta_I^{\text{low}}(u_i), \zeta_I^{\text{up}}(u_i))$ and mortality $(\zeta_M^{\text{low}}(u_i), \zeta_M^{\text{up}}(u_i))$ accumulated over the entire simulated period (2000-2050) with respect to all uncertainty sources, in all the regions studied.

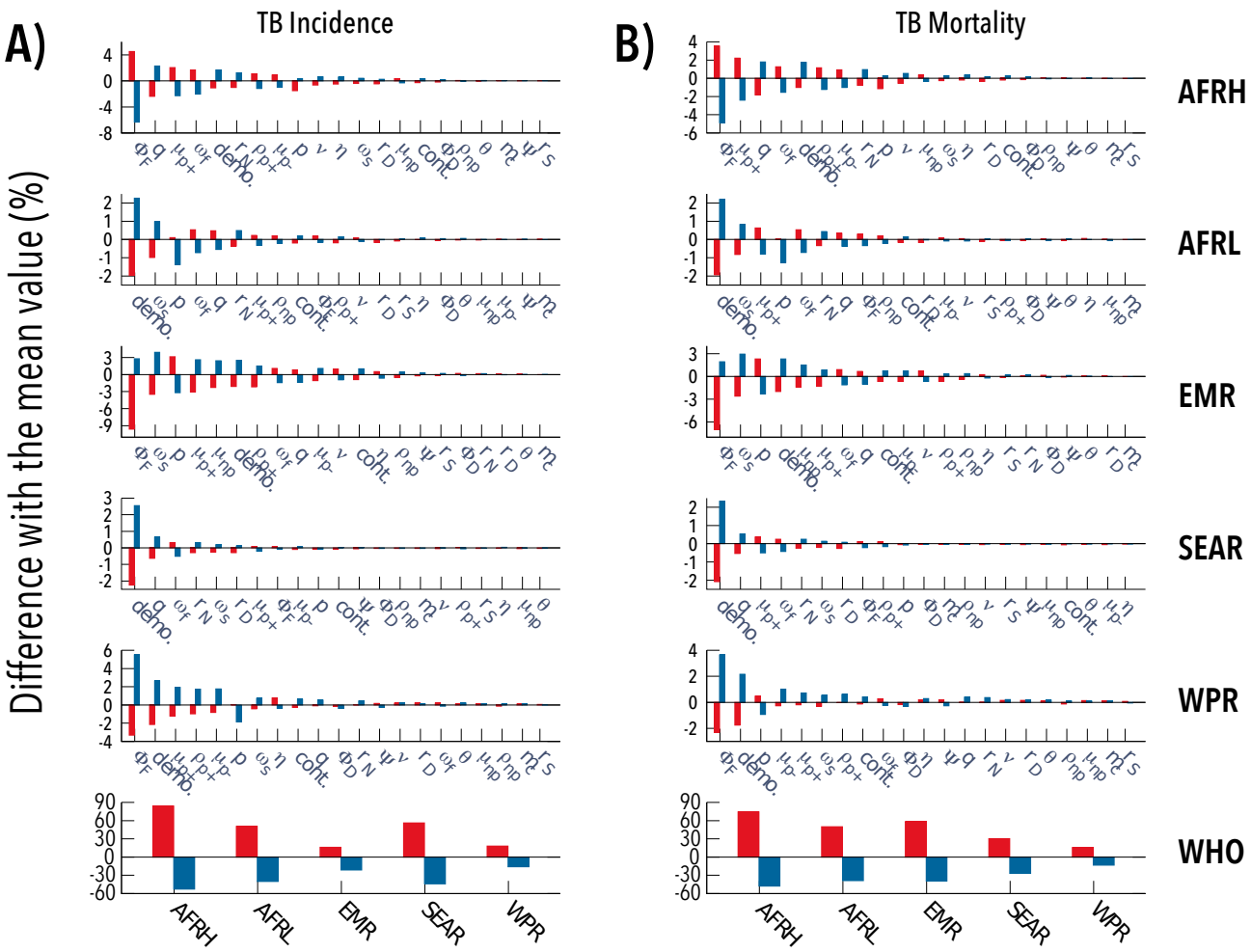

Figure S6: Model sensitivity analysis. Red bars represent the sensitivities when the uncertainty source is replaced by its upper limit, while the blue bars represent the sensitivities for the lower limits. The left panels show the sensitivities for the accumulated incidence, while right panels correspond to the sensitivities for the accumulated mortality, over the period 2000-2050, for the five regions as indicated. The different sources of uncertainty on these panels are sorted in decreasing order of total –upper and lower– sensitivity. The bottom panels, labeled as WHO, contain the sensitivities caused by the WHO estimates for the five regions.

5 Model forecasts: basal scenario (no intervention)

5.1 Outcomes of the complete model

5.1.1 Aggregated burden

In figure S7, we represent the projected incidence and mortality rates for the five regions in the world with highest TB burden levels, along the corresponding uncertainty levels. As we see, our model correctly captures incidence and mortality levels reported in WHO Database,[5] whose variability constitutes the main source of uncertainty of model's predictions, mostly at the initial stages of the simulations.

5.1.2 Age distribution of incidence rates

In Figure S8 we compare the age distribution of cumulated incident cases occurred between 2006 and 2012 for individuals younger than 65 years old (see section 2.2) reported in the database (dark blue bars), to the distribution produced by our model (light blue bars) and a reduced version of it that does not make use of the detailed information of the database regarding TB distribution by age and disease type (grey bars). In this reduced model, data-driven estimates of fast progression probabilities and fractions of cases of each type, the values used in Abu-Raddad et al.[2] are implemented. As we see in the figure, while our model succeeds at correctly reproducing the distribution of incident cases that is reported in the WHO TB database, the reduced

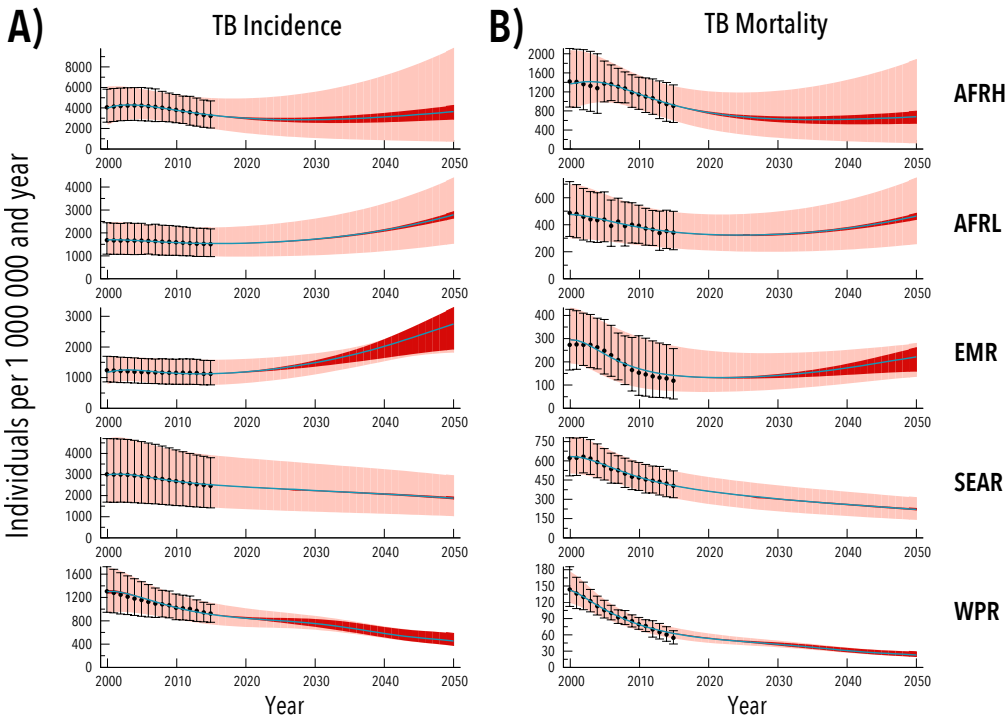

Figure S7: Model projections and uncertainty sources for the different regions, for aggregated TB burden rates in the period 2000-2050. Salmon area represents the uncertainty given by the WHO measurements, red area represents the uncertainty given by the parameters associated to the Natural History. Black filled dots and error bars stand for the WHO estimations of TB burden levels in each region during the training period 2000-2015.

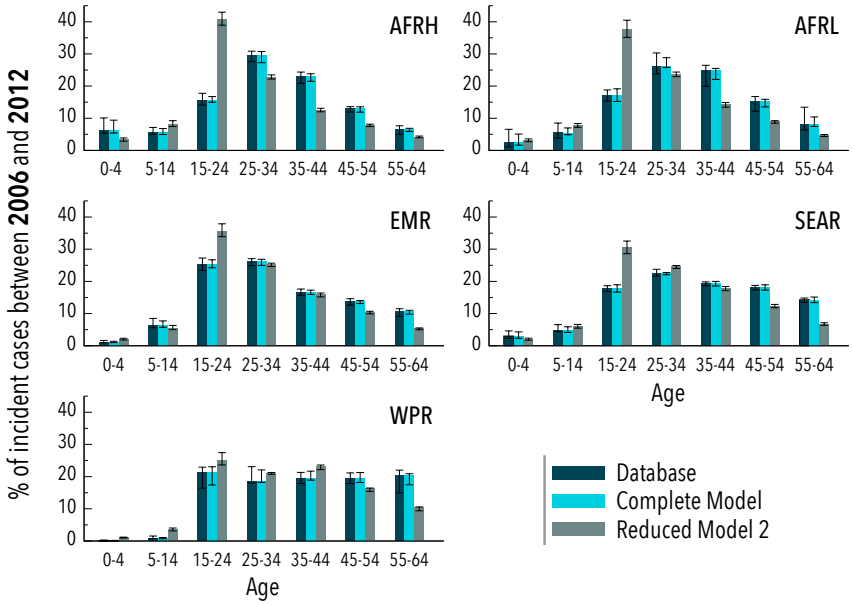

Figure S8: Percentage of cases per age-group between 0 and 65 years old, between 2006 and 2012 given by the database (dark blue),[5] obtained by our complete model (light blue bars), and obtained from reduced model 2 (grey).

model produces an over-estimation of the number of cases that occurs in young adults and an under-estimation of the burden in the eldest individuals, in all 5 regions. For further details on the reduced model and this comparison, see section 5.2.

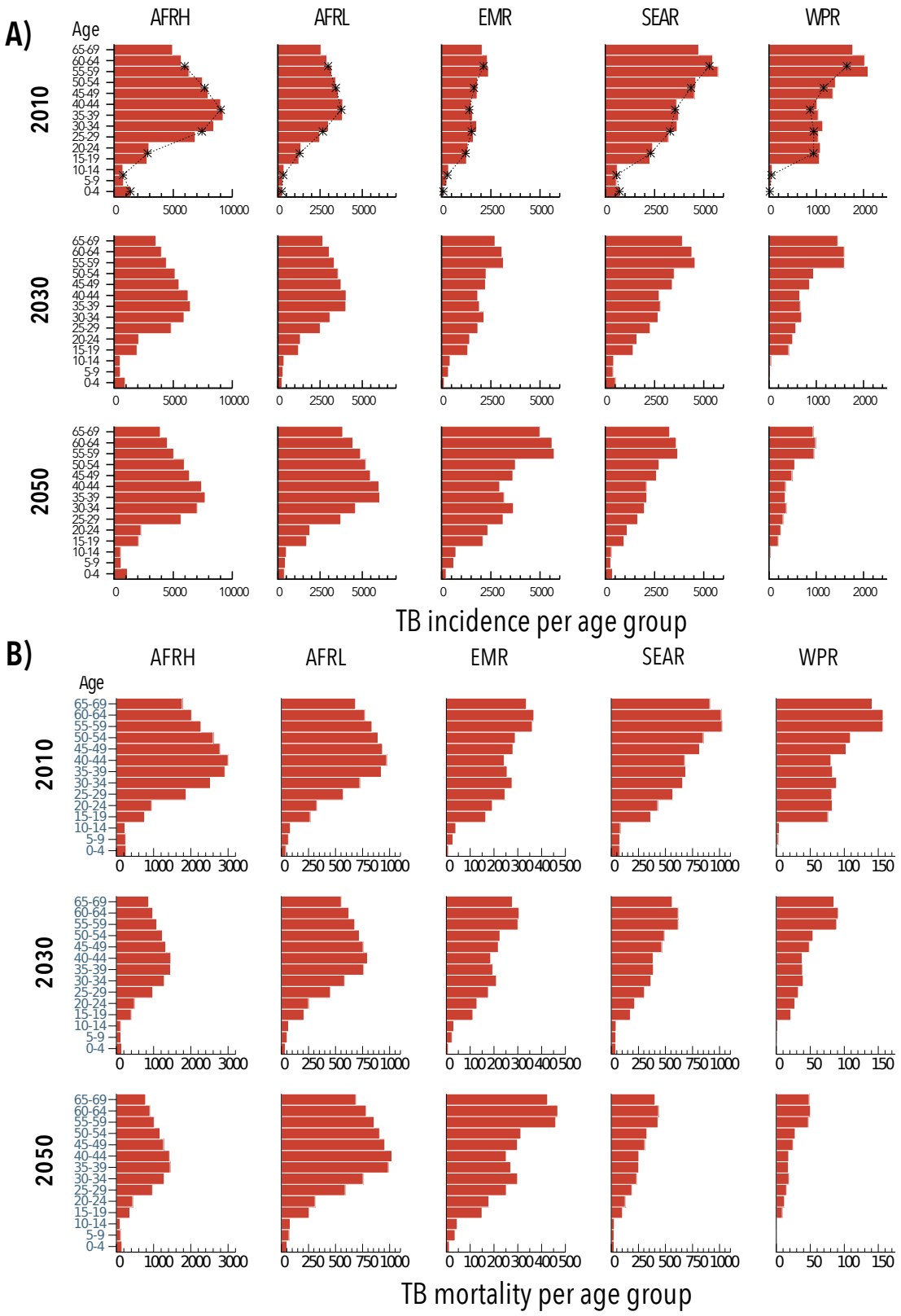

Figure S9: Age distributions of incidence (panel A) and mortality rates (panel B) for the different regions considered in this work at different times as indicated. Symbols in panel A refers to the WHO dataset values.

Once shown that our model correctly reproduces the incidence distribution observed in the training period 2006-2012, in Figure S9, we represent the age distribution of incidence and mortality rates in the different

regions in 2010, 2030 and 2050. We observe different age distributions of the TB burden, which correlate with the demographic evolution expected in each of them. In regions AFRH and AFRL, where the demographic shift –i.e. the transition from high to low birth and death rates– is not expected to be finished by 2050, the peak of incidence and mortality rates remain in adults –from 30 to 45 years old, approximately–. However, in the rest of the regions EMR, SEAR and WPR, where the aging process is more advanced, the peak of incidence and mortality occurs in the eldest individuals. We stress that this finding, with has important consequences for policy practices, is a genuine outcome of our modeling approach that has not been anticipated before.

5.2 Influence of model features

5.2.1 Exposition of model features and reduced models

In this work we have reconsidered certain modeling assumptions that were made by previous models, critically related to the way in which TB intrinsic dynamics and the coupling between populations' age structure is described. To do so, we have adopted a data-driven approach that corresponds to the following three main age-related model ingredients discussed in the main text:

- Time evolution of the demographic structure: unlike previous approaches, where the demographic structure of the populations is considered to remain constant, in this work we explicitly force our dynamical system to evolve in accordance to demographic forecasts by UN (see section 1.5).
- Detailed calibration of age and TB type distributions of incident cases: While model calibration procedures of previous approaches are aimed at reproducing aggregated TB burden levels during a given training period, when it comes to provide a more detailed description of age-dependent burden distributions for the different kinds of disease, the WHO Database for TB provides more information that can (and must) be used in the context of large-scale TB modeling. In this sense, as explained in section 2.2, our model is designed to reproduce: i) age-distributions of incident cases, and ii) fractions of cases that correspond to each type of TB (pulmonary smear positive or negative and non-pulmonary) also per age group. As previously exposed, this is achieved by i) fitting the probabilities of fast-progression per age group (to reproduce the fraction of cases associated to each age-strata (see section 2.2.1), and ii) estimating the probabilities of developing each type of disease from the fractions reported at the case notifications WHO database (see section 2.2.2).
- Heterogeneous contact patterns: Abandoning the classical hypothesis according to which contacts among all age groups are homogeneous, in this work, we rely on international-scale survey data to estimate an heterogeneous matrix for age-mixing contact patterns (section 1.3).[3]

Beyond these three main novel features, which are thoroughly discussed in the main text, an additional aspect that is treated in a way that differs from previous approaches is the procedure for estimating the initial conditions of the dynamical system:

- Initial Conditions out from the stationary equilibrium. Initial conditions –i.e., the distribution of individuals across the different disease stages detailed in table 3 at $t = 2000$– are not forced to correspond to the stationary equilibrium of the system. Instead of that, an additional parameter ς is fitted representing the initial deviations of the system with respect to the equilibrium point.

In this section, we inspect and study the influence of all these different ingredients by switching off each one of them separately, which, alongside the complete model and a base model –which includes none of these new features–, give us a total of 6 different models to compare. Specifically, we have:

- Complete model. All ingredients are implemented.
- Reduced model 1: Constant demographic structure. A version of the complete model where the evolution of the demographic structure is neglected. It uses the demographic structure of 2000 and keeps it constant during the entire time window, as detailed in section 1.5.
- Reduced model 2: Uncalibrated Age/TB type distributions. A version of the complete model where the detailed information about age and TB type distributions of incident cases is ignored. In this model, fast-progression probabilities, as well as the probabilities associated for each type of disease are not estimated from data, and instead, the values used in Abu-Raddad et al.[2] are implemented (see table 9).

Age group	ρ_{p+}	ρ_{np}	p
Children ($a < 15$ y.o.)	0.1 (0.085,0.15 c.i.)	0.25 (0.2125,0.2875 c.i.)	0.05 (0.043,0.058 c.i.)
Adults ($a > 15$ y.o.)	0.5 (0.425,0.575 c.i.)	0.1 (0.085,0.15 c.i.)	0.15 (0.128,0.173 c.i.)

Table 9: Values of $\rho_{p+}(a)$ and $\rho_{np}(a)$ and $p(a)$ used in Reduced model 2, (obtained from Abu-Raddad et al.[2])

- Reduced model 3: Homogeneous contact patterns. A version of the complete model where contacts between individuals of any age are considered equally likely: i.e. $\xi(a_i, a_j) = \xi(a_k, a_l) \forall (i, j, k, l)$. Specifically, we assign a value that is the average contact rate in the heterogeneous mixing scenario –which permit to compare faithfully the fitted infectiousness between models.

And the following two complementary reduced models shown here:

- Reduced model 4: Initial Stationary Conditions. A version of the complete model where the system is forced to begin at its stationary conditions (forcing $\varsigma = 0$)
- Reduced model 5: (Base model): a version of the model that presents:
  - The same description of the Demographic evolution of Reduced model 1.
  - Uncalibrated Age/TB type distributions, as in Reduced model 2.
  - An homogeneous matrix of contact patterns as in Reduced model 3.
  - Initial Conditions as in Reduced model 4.

The Natural History of the disease remains the same for all the models. The implementation of the different reduced models requires the complete repetition of the fitting process, which includes the fitting of the diagnosis rate, the infectivity, the initial distance from stationarity –except for reduced models 4 and 5, for which it is fixed to zero– and the fast-progression probabilities –except for reduced models 2 and 5, for which they are given externally (see table 9)–.

5.2.2 Effects on TB aggregated burden

In figure S10 we show the evolution of aggregated incidence and mortality under these 6 different models. We see that for some regions –SEAR and WPR– all the different models provide approximately the same forecast, however in other cases –AFRH, AFRL and EMR– the behavior may diverge completely from one model to another. As we have discussed in the main text, we see in those regions that for models that do not take into account the evolution of demography, the burden of TB decreases monotonously with time –models 1 and 5– while in the rest of the models –full model and reduced models 2, 3 and 4– the aging of the population is able to reverse that trend. In SEAR and WPR, although aging of the population also takes place, the decreasing trend is not reversed.

Regarding the influence of the contact patterns, we hardly see any influence on the aggregated incidence for any region. Despite existing differences in the age dependency of the force of infection between the full model and the reduced model 3, both are bound to reproduce largely similar aggregated incidence and mortality rates. The main differences between these models, as discussed in the main text, arise late in time, affecting the incidence rates among the eldest group population.

As far as the different settings of initial conditions concern, no significant effect on the aggregated incidence rates are observed when comparing the reduced model 4 with the complete model in five over six regions. The exception is AFRH, which is the region that sets the initial conditions farthest from the stationary state, where we observe that forcing the system at the stationary in $t = 2000$ translates into a more exacerbated relapse in the last decades of our simulations. Nonetheless, the wide confidence intervals of this region –not shown in figure S10 for clarity, but can be found for the complete model in Figure S7–, make these differences not big enough to be called significant (relative variation of incidence rates in 2050: AFRH, Reduced Model 4 minus Complete model: 35%, (95% CI: -3, 210)).

Summing up, two main conclusions can be extracted from the analysis of figure S10. On one hand, the description of the demographic structure is the key model feature as it has the deepest effect on aggregated burden forecasts. On the other hand, this influence is observed in EMR, AFRL and AFRH. These regions are specifically (see figure S2) those where the global aging process that is being experienced by the entire human population all around the world is at its earlier stages.

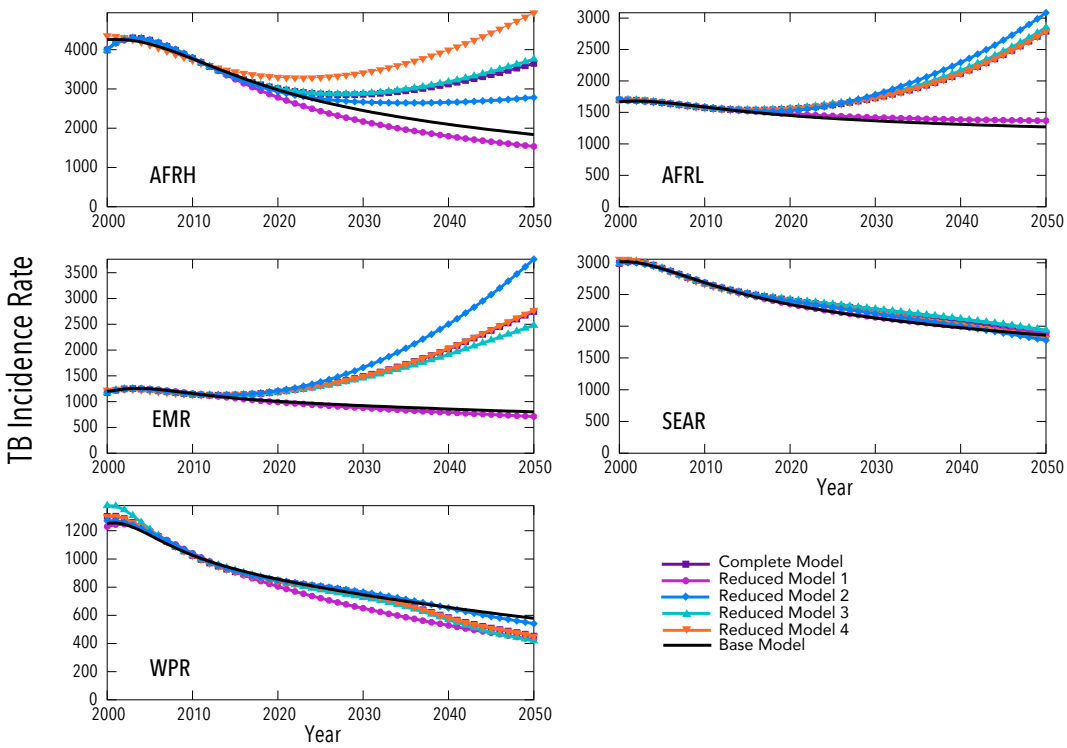

Figure S10: Evolution of incidence rates under different models.

5.2.3 Effects on fitted parameters: diagnosis rates and infectiousness

Under the frameworks defined by each of the reduced models, we also repeat the entire fitting process. In this way, an interesting exercise consists of analyzing the estimates produced for $d(t)$ and $\beta(t)$ after calibrating each of the reduce models, and compare them to those obtained after calibrating the complete model. This allows us to evaluate the possibility that some of the observed differences between the complete model and the reduced ones could be explained due to the different infectiousnesses/diagnoses rates fitted in each case. In Figures S11 and S12, we represent how the diagnosis rate and scaled infectiousnesses obtained in the different reduced models compare to the equivalent parameters of the complete model, by computing the variation of each parameter x as:

$$v(x)_{\text{reduced}_i} = \frac{x_{\text{reduced}_i} - x_{\text{complete}}}{x_{\text{complete}}} \tag{102}$$

where the subindex 'reduced$_i$' refers to the i-th reduced model, and 'complete' refers to the complete model. Table 10 shows the fast progression probabilities that have been fitted for the different models and regions. As we can see, the reduced model 3 (homogeneous contacts) tends to fit the most divergent values for the risk of fast latency with respect to the complete model. The different age dependent dynamics of the models are thus compensated to reproduce the same age distribution of TB cases as well as the trends of aggregated TB burden in the fitting window.

From a detailed analysis of figures S11 and S12, along with the probabilities reported in table 10, an interesting feature arises from the comparison between the reduced model 1 and the complete model. Admittedly, variations of diagnosis rates between those models are reduced and not systematic (i.e. the time series crosses, and the reduced model's rate is sometimes higher, sometimes lower that the one from the complete model), similarly to how fast progression probabilities in different age groups and regions behave in these two models (see table 10). More interestingly, variations in scaled infectiousness between those two models are significantly and surprisingly larger in the reduced model than in the complete one. This implies that the increased burden levels observed in EMR, AFRH and AFRL when the demographical evolution is incorporated into the complete model cannot be associated to differences in the fitted parameters, since these differences, if any, would point precisely in the opposite direction. This observation supports our interpretation that the burden forecasts of these regions to emerge are a direct consequence of properly treating the time evolution of the demographic structures, rather than being caused by a technical artifact.

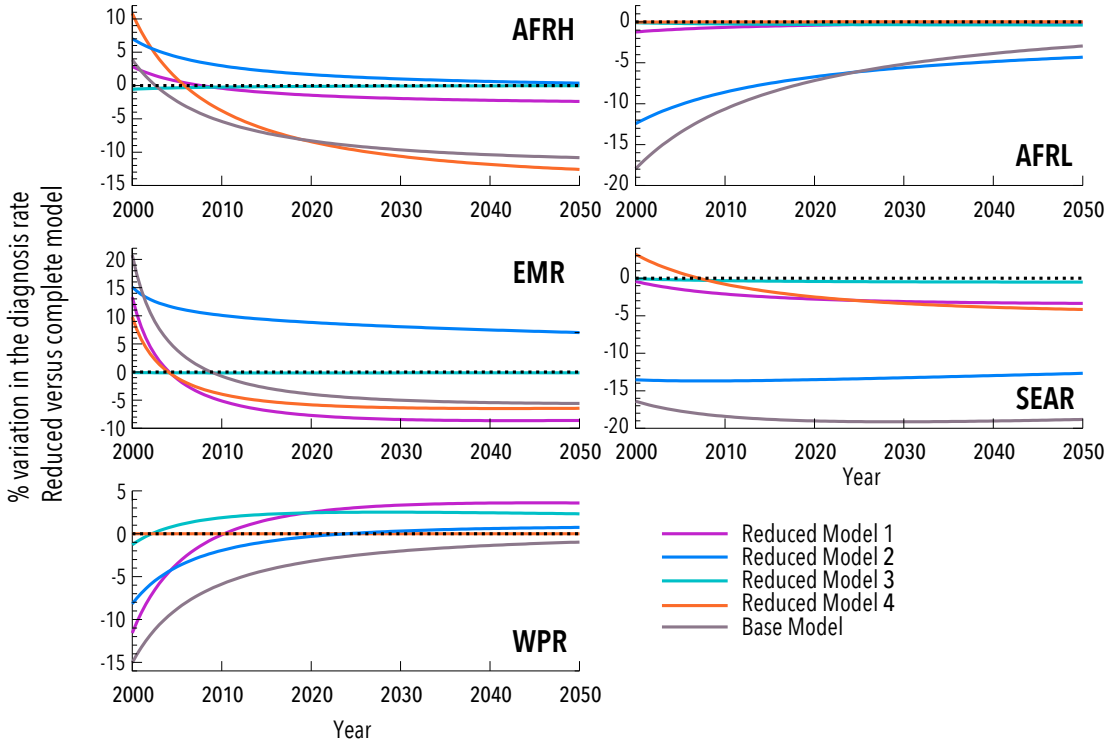

Figure S11: Variability of the diagnosis rate for the different reduced models with respect to the complete model for the different regions, as a function of time.

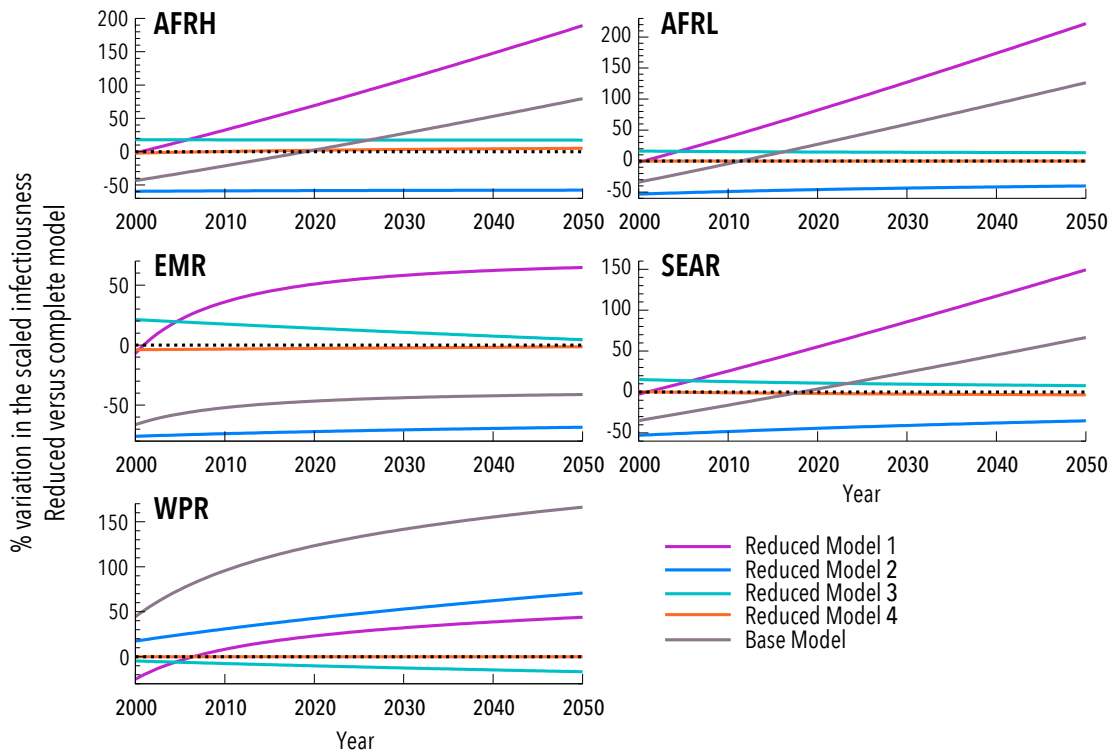

Figure S12: Variability of the scaled infectivity for the different reduced models with respect to the complete model for the different regions, as a function of time.

	Model	0-4	5-14	15-24	25-34	35-44	45-54	+55
AFRH	Complete Model	3.6 (3.0-5.9)	0.6 (0.5-0.8)	3.5 (3.1-4.0)	10.9 (10.2-11.2)	13.5 (12.1-14.3)	12.0 (11.1-12.3)	10.2 (7.7-11.8)
	Reduced model 1	3.4 (2.8,5.5)	0.6 (0.5-0.8)	3.4 (3.1-4.0)	11.5 (10.8-11.8)	13.4 (12.0-14.1)	11.8 (10.9-12.0)	10.0 (7.5-11.6)
	Reduced model 3	2.1 (1.7,3.4)	0.3 (0.3-0.4)	4.3 (3.9-5.0)	13.4 (12.5-13.8)	15.1 (13.6-16.0)	11.3 (10.4-11.5)	7.8 (5.9-9.1)
	Reduced model 4	3.5 (2.9,5.8)	0.6 (0.5-0.7)	3.4 (3.1-3.9)	11.0 (10.3-11.3)	13.7 (12.3-14.5)	12.2 (11.3-12.5)	10.2 (7.7-11.9)
AFRL	Complete model	1.3 (0.6-3.6)	0.9 (0.6-1.5)	3.6 (3.4-3.9)	8.1 (7.6-9.7)	12.6 (9.7-13.2)	12.2 (9.6-12.9)	11.7 (9.0-20.3)
	Reduced model 1	1.2 (0.5-3.2)	0.7 (0.5-1.1)	3.5 (3.2-3.7)	8.5 (8.0-10.2)	12.6 (9.8-13.3)	13.1 (10.4-14.0)	11.7 (9.0-20.3)
	Reduced model 3	0.6 (0.3-1.6)	0.4 (0.3-0.6)	4.1 (3.8-4.6)	10.0 (9.2-11.8)	14.4 (11.2-15.1)	11.9 (9.4-12.7)	9.8 (7.5-17.0)
	Reduced model 4	1.3 (0.6-3.6)	0.9 (0.6-1.5)	3.6 (3.4-3.9)	8.1 (7.6-9.7)	12.6 (9.7-13.2)	12.2 (9.6-12.9)	11.7 (9.0-20.3)
EMR	Complete model	0.5 (0.4-0.8)	1.6 (1.4-2.2)	4.1 (3.7-4.4)	5.3 (5.1-5.4)	4.7 (4.4-5.0)	7.1 (6.5-7.5)	12.0 (10.2-13.0)
	Reduced model 1	0.3 (0.2-0.4)	0.8 (0.7-1.1)	3.9 (3.5-4.3)	6.0 (5.7-6.1)	4.9 (4.6-5.1)	8.4 (7.7-9.0)	12.3 (10.5-13.4)
	Reduced model 3	0.08 (0.07-0.12)	0.6 (0.5-0.8)	5.3 (4.8-5.8)	6.4 (6.2-6.6)	5.1 (4.8-5.3)	6.7 (6.1-7.1)	9.6 (8.2-10.4)
	Reduced model 4	0.5 (0.4-0.8)	1.6 (1.4-2.2)	4.1 (3.7-4.5)	5.3 (5.1-5.4)	4.7 (4.4-4.9)	7.0 (6.4-7.5)	12.0 (10.2-13.0)
SEAR	Complete model	3.4 (2.6-5.3)	1.3 (1.2-1.8)	5.0 (4.8-5.3)	7.3 (7.1-7.9)	8.4 (8.1-8.5)	12.6 (12.0-12.9)	18.2 (17.3-19.0)
	Reduced model 1	2.9 (2.2-4.4)	1.1 (1.0-1.6)	4.6 (4.4-4.9)	7.5 (7.2-8.0)	8.8 (8.5-8.9)	15.3 (14.6-15.7)	20.2 (19.3-21.1)
	Reduced model 3	1.7 (1.3-2.6)	0.6 (0.5-0.8)	6.0 (5.7-6.2)	8.8 (8.6-9.5)	9.3 (9.1-9.5)	12.6 (12.0-12.9)	16.1 (15.4-16.8)
	Reduced model 4	3.4 (2.6-5.3)	1.3 (1.2-1.8)	5.0 (4.8-5.3)	7.3 (7.1-7.9)	8.4 (8.1-8.5)	12.6 (12.1-13.0)	18.3 (17.4-19.1)
WPR	Complete model	0.23 (0.16-0.61)	0.7 (0.6-1.2)	17.3 (12.3-18.1)	16.0 (15.6-20.8)	14.8 (13.7-15.3)	24.3 (22.7-24.6)	43.0 (29.6-45.2)
	Reduced model 1	0.27 (0.19-0.72)	0.02 (0.02-0.04)	22.3 (16.4-23.4)	13.1 (12.8-17.0)	22.4 (20.7-23.2)	32.7 (30.5-33.0)	70.8 (48.8-74.5)
	Reduced model 3	0.00 (0.00-0.00)	0.07 (0.06-0.13)	25.6 (18.9-26.9)	22.1 (21.7-28.9)	19.8 (18.3-20.5)	29.5 (27.4-29.8)	47.3 (32.6-49.7)
	Reduced model 4	0.23 (0.16-0.61)	0.7 (0.6-1.2)	17.3 (12.7-18.1)	16.0 (15.6-20.8)	14.8 (13.7-15.3)	24.3 (22.7-24.6)	43.0 (29.6-45.2)

Table 10: Values of the fitted fast progression probabilities (in %) for the different regions and models. For reduced models 2 and 5 (base model) the probabilities are 5% for children less than 15 years old, 15% for the rest (see table 9).

5.2.4 Effects on age distributions of TB burden

Beyond detailing the effects of the novel model ingredients on forecasting TB's aggregated incidence and mortality rates, it is equally interesting for the purposes of this work to understand how these modeling hypothesis influence the way TB burden is distributed across age strata. In this context, the most relevant ingredient of our model is precisely the detailed calibration step described in section 2.2, aimed at rendering the model compatible with the Case Notifications data about TB incidence distribution across age groups and types of disease. Admittedly, as we can see in figure S8, while the complete model provides a close description of the distribution of TB incident cases across age groups, the reduced model 2, where this detailed calibration of age/TB types cases is not accomplished, is unable to accurately reproduce the observed data. To formally backup this observation, we have compared each fraction produced by each model to the corresponding value from the database (through a simple Z-test as a first approximation). The results of these tests are detailed in tables 11 and 12. As we see there, while the reduced model is essentially incompatible with the observed data; our approach succeeds at reproducing the TB burden per age group that is reported in the Database, with the (marginal) exceptions of children in WPR, where the WHO Database reports lower burden rates when compared to our model.

The detailed calibration of TB distributions across ages and TB types is not the only ingredient that might affect model results per age-strata. To explore the effects of all the novel ingredients on age distributions of burden, we compare in Figure S13 per-age incidence and mortality rates coming from the different reduced models to the corresponding estimates from the complete model. Again, the region chosen for this comparison

Region	0-4	5-14	15-24	25-34	35-44	45-54	55-64
AFRH	$1.6 \cdot 10^{-4}$	$8.2 \cdot 10^{-3}$	$2.0 \cdot 10^{-3}$	$1.6 \cdot 10^{-4}$	$1.2 \cdot 10^{-3}$	$3.3 \cdot 10^{-3}$	$4.1 \cdot 10^{-4}$
AFRL	$2.1 \cdot 10^{-4}$	$5.5 \cdot 10^{-3}$	$4.9 \cdot 10^{-3}$	$4.0 \cdot 10^{-6}$	$3.2 \cdot 10^{-3}$	$1.3 \cdot 10^{-3}$	$2.2 \cdot 10^{-3}$
EMR	$5.8 \cdot 10^{-4}$	$5.9 \cdot 10^{-3}$	$2.0 \cdot 10^{-2}$	$1.5 \cdot 10^{-2}$	$8.0 \cdot 10^{-3}$	$1.2 \cdot 10^{-2}$	$1.1 \cdot 10^{-2}$
SEAR	$1.3 \cdot 10^{-3}$	$3.2 \cdot 10^{-3}$	$8.4 \cdot 10^{-3}$	$7.8 \cdot 10^{-4}$	$1.2 \cdot 10^{-3}$	$3.8 \cdot 10^{-3}$	$7.3 \cdot 10^{-3}$
WPR	$1.1 \cdot 10^{-1}$	$1.2 \cdot 10^{-1}$	$3.5 \cdot 10^{-2}$	$1.0 \cdot 10^{-2}$	$2.1 \cdot 10^{-2}$	$1.8 \cdot 10^{-2}$	$2.7 \cdot 10^{-2}$

Table 11: Null hypothesis p-Values for the comparison of the fraction of cases for age-group obtained by the complete model against the data reported in the case notifications database (Null hypothesis: fractions estimated by the model are different from those reported in the Database). After multiple testing correction (Benjamini-Hochberg method[25]), only two comparisons lye above the 5% false discovery rates: and correspond to the two first age groups in WPR (Fdrs equal to 11% and 12%, respectively).

Region	0-4	5-14	15-24	25-34	35-44	45-54	55-64
AFRH	$\simeq 1$	0.997	$\simeq 1$	$\simeq 1$	$\simeq 1$	$\simeq 1$	0.997
AFRL	0.241	0.859	$\simeq 1$	0.953	$\simeq 1$	$\simeq 1$	$\simeq 1$
EMR	0.997	0.848	$\simeq 1$	0.809	0.847	$\simeq 1$	$\simeq 1$
SEAR	0.990	0.807	$\simeq 1$	0.997	0.999	$\simeq 1$	$\simeq 1$
WPR	$\simeq 1$	$\simeq 1$	0.999	0.729	$\simeq 1$	$\simeq 1$	$\simeq 1$

Table 12: Null hypothesis p-values for the comparison of the fraction of cases for age-group obtained using the reduced model 2 against the data reported in the case notifications database (Null hypothesis: fractions estimated by the model are different from those reported in the Database). The null hypothesis cannot be rejected even under a relaxed nominal p-value threshold of 0.2 for any age/group/region combination, and 28 over 35 tests yields nominal p-values larger than 95% indicating the inability of reduced model 2 to reproduce the observed burden distributions per se.

is EMR, and the rates are shown at three different times –2010, 2030, and 2050.

Beyond the already discussed effects on these results of the description of the demographic structure, contact patterns and the detailed calibration of TB type and age distributions, (discussed in the main text and in the previous paragraphs of this section), from the comparison between the complete model and the reduced model 4, we can conclude that the initial conditions setup procedure has no noticeable effects on model's forecasts regarding the age distribution of TB burden in this region. Furthermore, as we did regarding aggregated burden levels, we observe notable differences when we compare the complete model with the base model, which, unlike the complete model: i) would predict a decreasing TB burden in all regions; and ii) would produce an age distribution of TB burden among adults totally divergent from the case notifications database reports.

Lastly, in order to deepen into the effects of the different new ingredients on the age-dependency of TB burden, we computed other observables, like prevalence fractions of infectious and infected individuals in each age-group, as shown in Figure S14 for EMR. We compute the prevalence of infection for an age group a at a time t, $\mathcal{P}_{\text{infection}}(a,t)$, simply as:

$$\mathcal{P}_{\text{infection}}(a,t) = 1 - S(a,t)/N(a,t) \tag{103}$$

and the prevalence of infectious individuals as follows:

$$\mathcal{P}_{\text{infectious}}(a,t) = \Upsilon(a,t)/N(a,t) \tag{104}$$

where $\Upsilon(a,t)$ is the weighted sum of all infectious individuals in age group a at time t (see equation 6 in section 1.2).

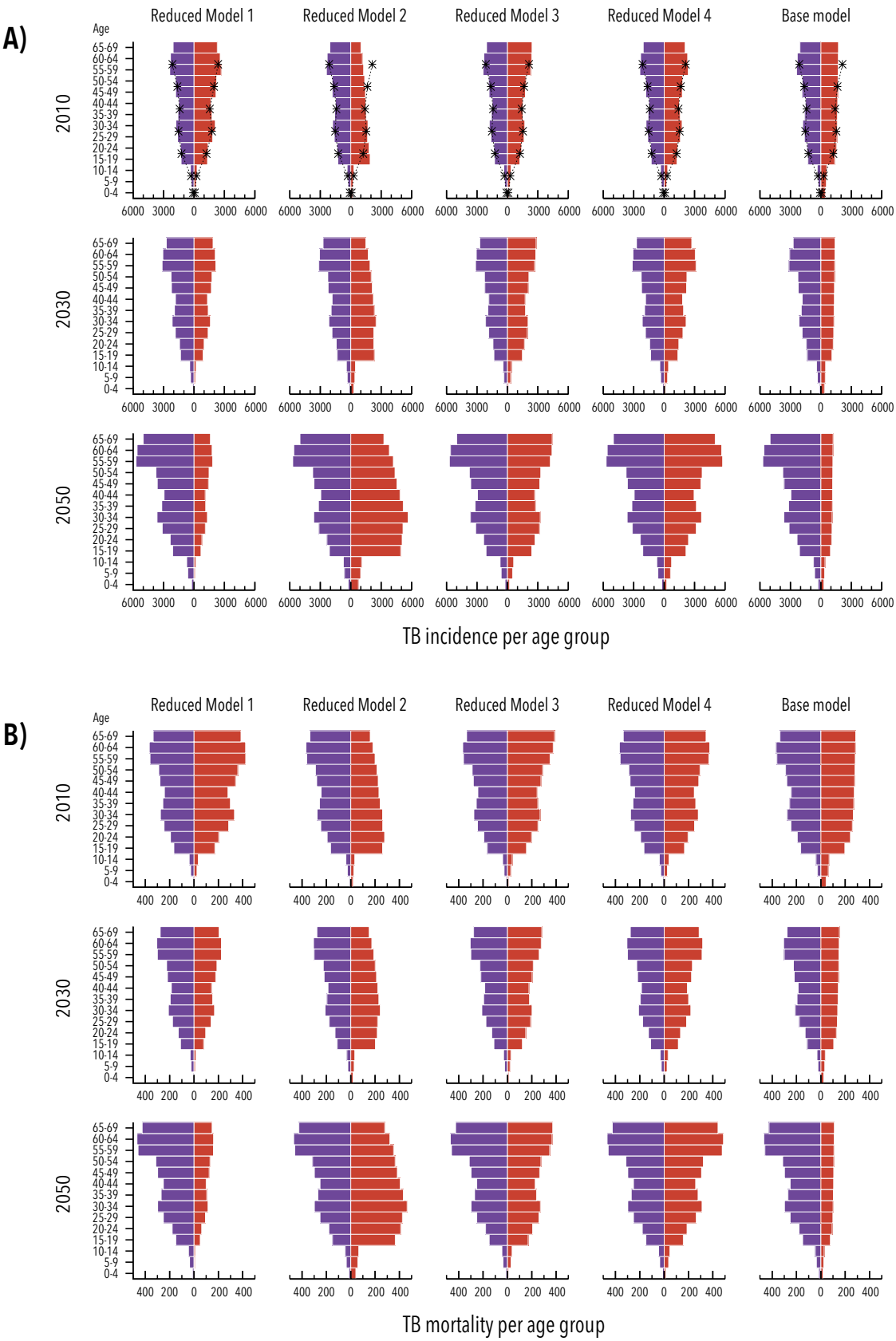

Figure S13: Incidence (panel a) and mortality (panel b) rates in EMR for the different reduced models (red bars) at different time points (2010, 2030 and 2050). Purple bars represent the complete model (column-wise identical, for the sake of comparison), while red bars represent reduced models' estimations.

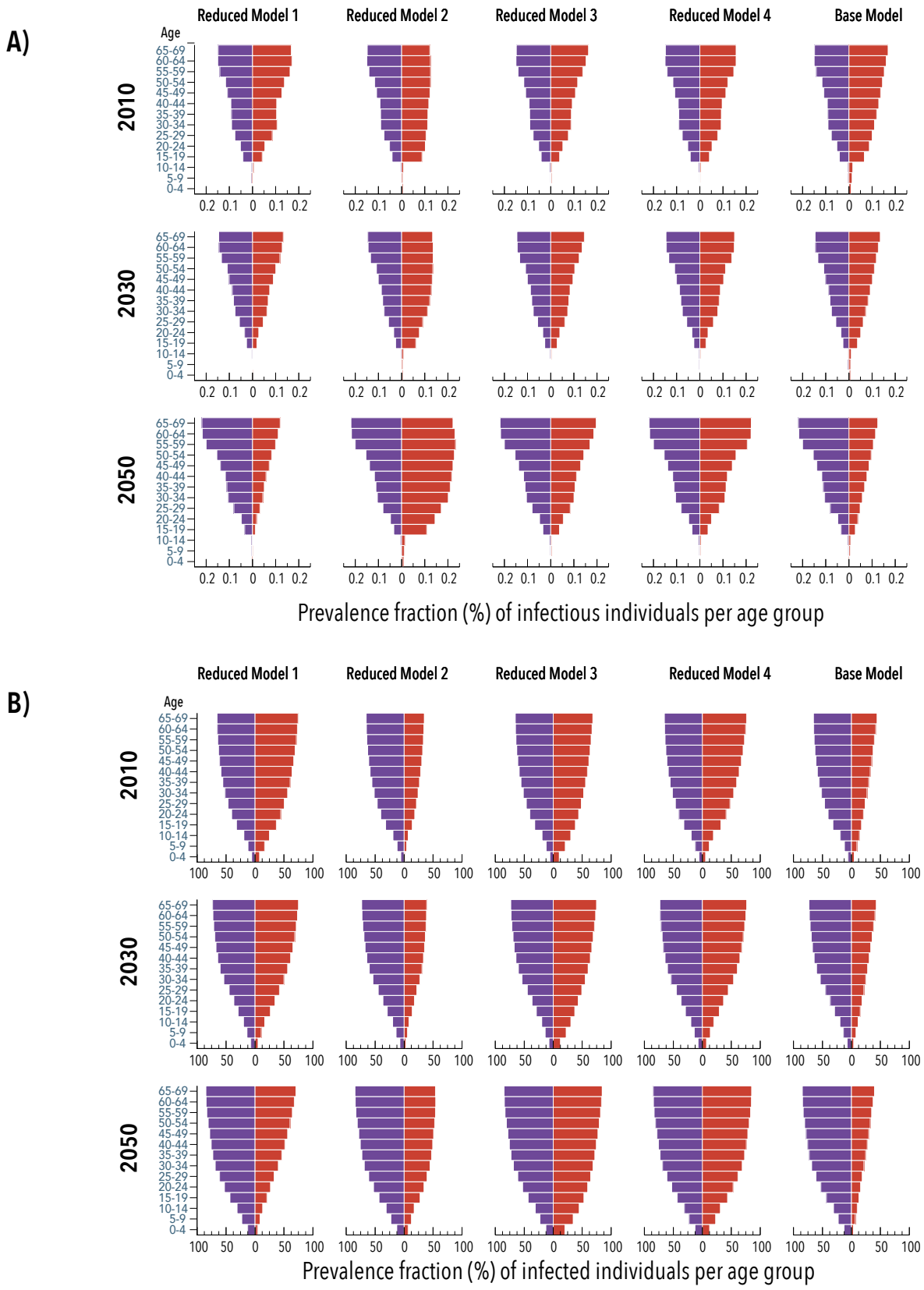

Figure S14: Age distribution of the prevalence fractions of infectious individuals (panel A) and prevalence of infection (panel B) in EMR in 2010, 2030 and 2050, for the different reduced models (red bars) in comparison with the complete model (purple bars).

6 Model forecasts: vaccine impact evaluations

6.1 Outcomes from the complete model

Regarding vaccine impact evaluations, we define the impact of a vaccine $\mathcal{I}$ as the difference of the accumulated incidence between the model with and without vaccine, during the period 2025-2050:

$$\mathcal{I}^{\text{inc}} = \sum_a I(a, 2050) - \sum_a I_{\text{vac}}(a, 2050) \tag{105}$$

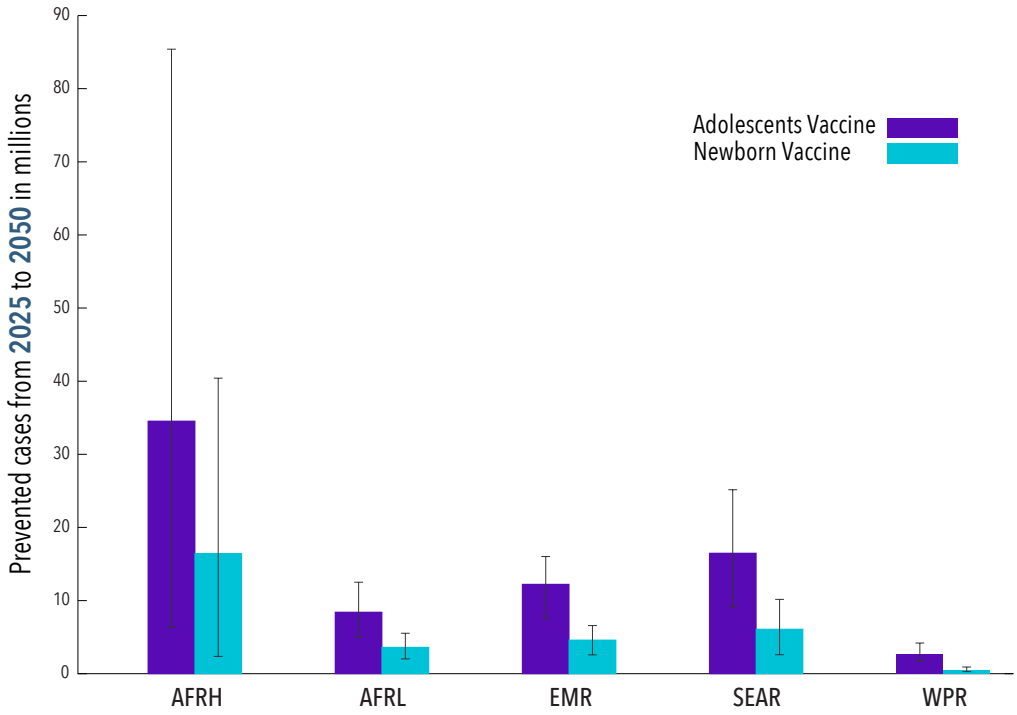

Figure S15: : Vaccine impacts for a newborn vaccine, and an adolescent vaccine.

In figure S15 we show the impacts obtained by following the two vaccination strategies evaluated in this work. As discussed in the main text, the adolescent-focused campaign prevents a larger number of cases than the vaccine inoculated on newborn, but this result is strictly subjected to the condition that both vaccines are equally protective and the coverage fractions they are being applied on are also the same.

6.2 Influence of model features on impact evaluations of novel vaccines.

As a last exploratory set of analyses, we analyze how the new ingredients of our model, defined in section 5.2.1, affect the impacts of the vaccine strategies being compared in this work. To do so, we will compare the impact forecasts provided by our Complete Model to those produced by its reduced versions.

As we see in figure S16, panel A, the booster vaccine, administered in adolescents, provides a greater impact than the substitutive vaccine when the vaccines are equally protective ($\epsilon = 0.2$ in all cases), not only for every region as we have seen before, but also for every reduced model. However, we also observe big differences between models in some regions. This was to be expected as we have already seen significant differences in the projected incidence rates. As expected, the models that yield greater incidences (compare to figure S10), also provide greater impacts, and these differences are larger in the regions where the models compared predict larger differences for basal incidence rates. In order to measure the influence of the different novel ingredients on the measures represented in figure S16, panel A, we compute, in panel B the variation of the different measurements of impacts –impact with the newborn vaccine, impact with the adolescent vaccine and difference between them– of each reduced model with respect to the complete model, as shown in equation 102.

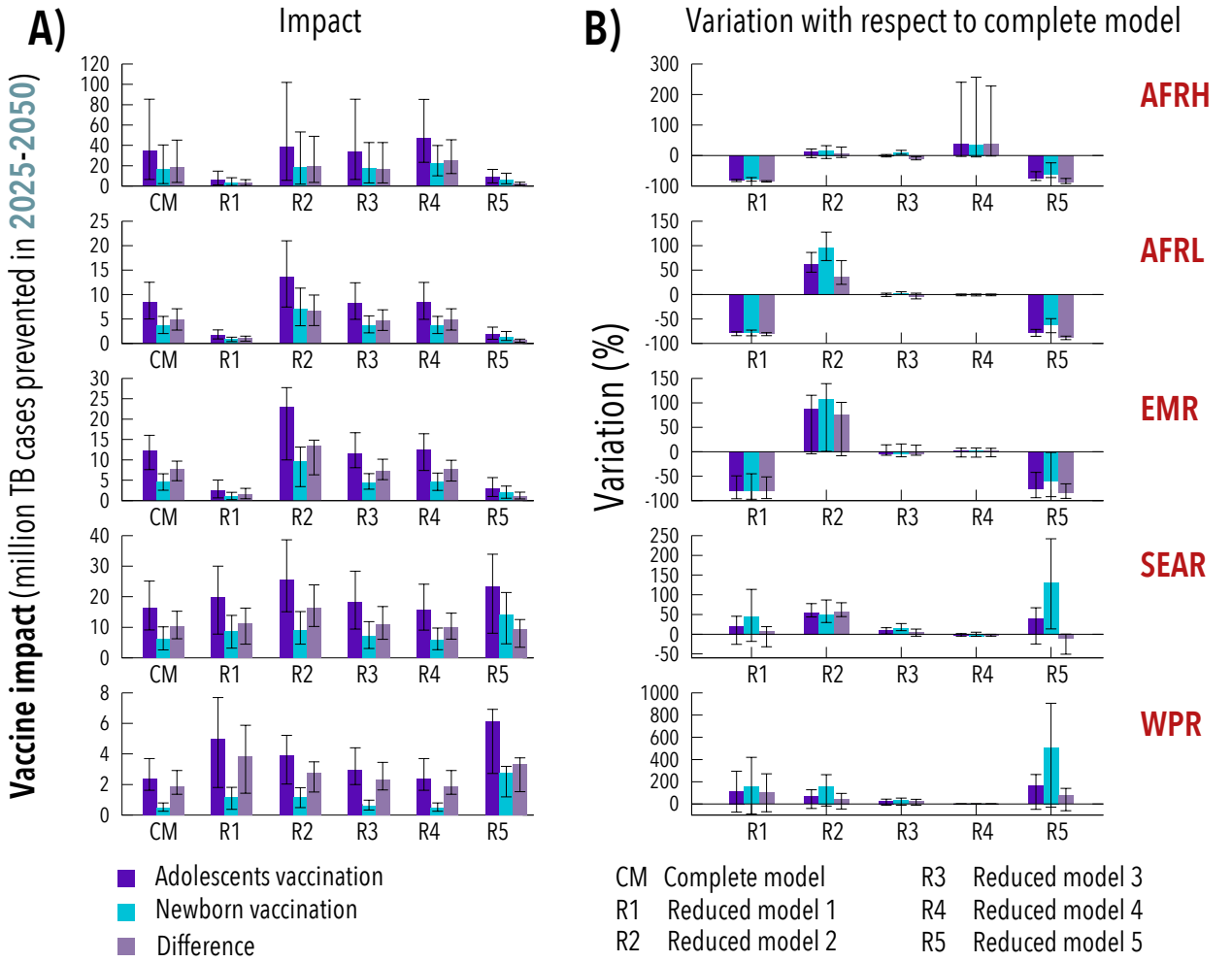

Figure S16: Panel a: Impacts (TB cases avoided from 2025 to 2050) associated to the adolescent-focused vaccine (dark purple, left columns) and the newborn-focused vaccine (blue, center columns), for the different regions and models (x-axis). We also represent the difference on impact between both strategies (light violet, right columns). Panel b: Variation of the measurements shown on the left when comparing the reduced models and the complete model.

Several aspects of figure S16, panel B deserve to be discussed. First, little effects on aggregated impact evaluations are associated to either the initial conditions setup procedure or the modeling approach used to describe contact patterns. In contrast, these impact evaluations are significantly affected by the description of the demographic structure (mostly in Africa and EMR) as well as by the calibration of age/TB types distributions of TB incidence (mainly in AFRL and SEAR).

These evaluations of the effects of each modeling hypothesis on vaccine impact evaluations must be considered only in the context of the specific comparisons here presented (i.e. vaccine impacts and differences between them after 25 years), for the effects observed if, for example, more extended look-up periods are considered, these observations might be different. On the one hand, it is to be expected that the different between vaccinating adolescents and newborns would be less unfavorable to the latter if the period through which impacts are compared is extended. This is due to the fact that, while an adolescent-focused campaign targets faster the age groups that are at the same time more infectious and more hit by the disease, in the long term vaccinating newborns allows to have an entirely vaccinated population, whereas vaccinating teenagers will always leave unprotected (or strictly speaking, only vaccinated with BCG, as they are now) the children younger than fifteen years old. Once again, it is worth mentioning that all these comparisons are based on the idea that the vaccines are equally protective when applied at the different age groups (and implicitly, of course, when compared them to basal protection provided by BCG).

In a similar note, while we have seen that, when impacts are evaluated within 25 after vaccination starts,

contact patterns does not affect the impact evaluations, it is worth remembering that, precisely, the main effect of considering heterogeneities in the contact matrices was observed in the incidence levels of eldest age groups, which are not yet vaccinated during the period simulated for the sake of these comparisons. We hypothesize that more extended impact evaluation periods would grant a more significant role to the contact structure used in the models.

7 Conclusion: model caveats and lines of future work

In this work we have presented a mathematical model for the description of global TB spreading which is specifically designed to evaluate and compare the impacts that novel vaccines might offer when applied on different populations' age-strata. Being the age of vaccination targets such a central question for the problem we address, the distinctive features of our model that differentiate it to previous approaches are introduced in order to pay a special attention to the ways the age structure of the populations is described: how does it evolve and couple to the disease propagation dynamics and burden rates. Regardless the eventual pertinence of these improvements, which has been thoroughly addressed in this study, our model shares a series of conceptual limitations with every single epidemiological model -very specially in the case of TB- that are worth discussing.

As Dowdy, Dye and Cohen thoroughly detail in a recent review,[26] the mathematical modelling of TB spreading is hindered by the inherent complexity of the natural History of the disease; as well as by our limited capabilities to reliably estimate the dynamical parameters that define each step of its infection cycle in different epidemiological settings. In that sense, TB models are almost unavoidably bound to rely on observational studies made on classical cohorts studied far away in time and location to retrieve estimates for the dynamical parameters. These estimations often reflect the features of epidemiological settings whose resemblance to the current TB situation in many parts of the world is, at least, arguable, which ultimately compromises the forecasting abilities of any mathematical model of TB spreading, including ours. As exposed by Dowdy and colleagues,[26] that is the reason why a coordinated research effort for improving our knowledge about TB transmission dynamics in current epidemiological settings should be considered a high priority task in the TB research agenda; and that is the reason why, in the lack of those studies, the choice of proper, data-driven modeling hypothesis and the completion of thorough uncertainty and sensitivity analyses constitute fundamental parts of every modeling work. Such considerations have shaped a number of the modeling hypotheses adopted in this work.

As many previous approaches, our model relies on the fitting of key parameters of the disease's dynamics to reproduce aggregated and -as a distinctive feature- age-distributed burden levels in the populations under study. The choice of the specific parameters to be fitted -infectiousness, diagnosis rates and fast progression probabilities- was done considering the well known fact that precisely these three parameters are those that, while having a deeper impact on disease dynamics, are counted among those whose estimation is more concerned and subject to diverse sources of heterogeneity.[26] In this line, the fact that these are estimated in a region-specific fashion is coherent with the assumption of their intrinsic heterogeneity across epidemiological settings. Notwithstanding that, even if fitting them to reproduce disease burden turns out to be a reasonable choice to produce robust burden forecasts, the estimations provided for the parameters themselves by our calibration procedure become strongly model-dependent and, as such, must be taken with caution. A paradigmatic example of these model-dependences can be seen from the comparison between the fast progression probabilities obtained by our complete model and those obtained upon calibration of the reduced model 3 (i.e. homogeneous contact patterns). Even if those two models produce, upon initial calibration, fairly similar burden forecasts, the sets of fast-progression probabilities fitted in each model are quite different. This is due to the fact that, upon modifying the contact structure used by the models, the calibration procedure of fast-progression probabilities absorbs those variations in order to reproduce the observed incidence rates per age-group.

These aspects have to be considered, on the other hand, on the design of the uncertainty and sensitivity analyses. When a model based on the calibration of some free parameters to reproduce observation has to be tested, the effects of each single parameter and uncertainty source must be evaluated on the entire procedure, including the re-calibration of the model itself upon variation of the input under evaluation. Remarkably enough, when we do so as explained in section 4, model forecasts turn out to be far more sensitive to the uncertainties associated to the initial burden estimates produced by WHO and reported in its data repository[5] than to the uncertainty associated to any dynamical parameter. This fact points out that, for producing more reliable model-based forecasts, along with further research on TB spreading dynamics, developing improved, more accurate epidemiological surveying techniques is likely to be, at least, just as important.

Furthermore, another key question that emerges not just in epidemiology (or even just in Biology), but in any mathematical modeling ground, is the need to find the proper compromise between model simplicity and level of details of the data to be described. In the context of TB compartmental models, the slow, persistent and

complex dynamics of the disease forces us to use Natural History schemes that are far more complex than those customarily used for the description of other diseases dynamics like influenza.[11,12] In this work, we have relied on previous models of TB dynamics to define the basic structure of the disease infection cycle,[1,2] from which our hypotheses depart only to provide, in a data-driven fashion, an improved description of the entanglement between population's age structure and disease dynamics. In this sense, one of our priorities has been addressing the effects of these novel hypotheses on models' forecasts. This set of comparisons, even if not fully formalized, go in the line of the so-called structural uncertainty/sensitivity analysis, (i.e. the quantification of how the different model elements -not just parameter values- impact the overall ability of the model to reproduce current burden levels and its future projections),[26,27] which allowed us to identify the adequate description of the demographic shift processes taking place in some of the regions under analysis as a fundamental piece missing in previous approaches. In this sense, structural sensitivity analysis should be more deeply explored in the context of TB modeling, both to question the presence of ingredients in current TB models whose impact in model forecasts is just as difficult to quantify as their actual relevance in current epidemiological settings, and to study the pertinence of adopting novel hypotheses.

Finally, as previously stated, while the development of a model to evaluate age-focused vaccination campaigns is the main purpose of this work, the specifics of such impact evaluations, as well as the vaccines features that arguably influence such comparisons remained beyond the scope of this work. Aspects like the differential difficulty of attaining equal levels of protection (and coverage) when a vaccine is applied on different age groups or when it is done on areas with different prevalence levels of HIV are aspects still not addressed in this article which are though needed in order to achieve quantitatively meaningful impact evaluations. The exploration of such questions, for which the model here presented represents a first fundamental instrument, will be our main priority in the near future.

References

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

[2] Abu-Raddad LJ, Sabatelli L, Achterberg JT, Sugimoto JD, Longini IM, Dye C, et al. Epidemiological benefits of more-effective tuberculosis vaccines, drugs, and diagnostics. *Prod Nat Acad Sci USA*. 2009;**106**(**33**):13980–5.

[3] Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med*. 2008;**5**(**3**):e74.

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

[5] WHO. Tuberculosis Database. http://wwwwhoint/tb/country/en/indexhtml. (accesed November 2016);.

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

[7] Picon PD, Bassanesi SL, Caramori MLA, Ferreira RLT, Jarczewski CA, Vieira PR. Risk factors for recurrence of tuberculosis. *J Bras Pneum*. 2007;**33**(**5**):572–8.

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

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

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

[11] Miller E, Hoschler K, Hardelid P, Stanford E, Andrews N, Zambon M. Incidence of 2009 pandemic influenza A H1N1 infection in England: a cross-sectional serological study. *Lancet*. 2010;**375**(**9720**):1100–8.

[12] Birrell PJ, Ketsetzis G, Gay NJ, Cooper BS, Presanis AM, Harris RJ, et al. Bayesian modeling to unmask and predict influenza A/H1N1pdm dynamics in London. *Prod Nat Acad Sci USA*. 2011;**108**(**45**):18238–43.

[13] Guzzetta G, Ajelli M, Yang Z, Merler S, Furlanello C, Kirschner D. Modeling socio-demography to capture tuberculosis transmission dynamics in a low burden setting. *J Theor Biol*. 2011;**289**:197–205.

[14] UN. Population Division Database. http://esaunorg/unpd/wpp/indexhtm. (accessed November 2016);.

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

[16] Moré JJ. The Levenberg-Marquardt algorithm: implementation and theory. In: Numerical analysis. Springer; 1978. p. 105–16.

[17] Vynnycky E, Fine P. The natural history of tuberculosis: the implications of age-dependent risks of disease and the role of reinfection. *Epidemiol Infect*. 1997;**119**(**02**):183–201.

[18] Comstock GW. Epidemiology of Tuberculosis 1–3. *Am Rev Respir Dis*. 1982;**125**(**3P2**):8–15.

[19] Murthy VK, Nair S, Gothi G, Chakraborty A. Incidence of tuberculosis among newly infected population and in relation to the duration of infected status. *Ind J Tub*. 1976;**23**:3–7.

[20] Sutherland I, Švandová E, Radhakrishna S. The development of clinical tuberculosis following infection with tubercle bacilli: 1. A theoretical model for the development of clinical tuberculosis following infection, linking from data on the risk of tuberculous infection and the incidence of clinical tuberculosis in the Netherlands. *Tubercle*. 1982;**63**(**4**):255–68.

[21] Marais BJ, Gie RP, Schaaf HS, Hesseling AC, Obihara CC, Nelson LJ, et al. The clinical epidemiology of childhood pulmonary tuberculosis: a critical review of literature from the pre-chemotherapy era [State of the Art]. *Int J Tuberc Lung Dis*. 2004;**8**(**3**):278–85.

[22] Soysal A, Millington KA, Bakir M, Dosanjh D, Aslan Y, Deeks JJ, et al. Effect of BCG vaccination on risk of Mycobacterium tuberculosis infection in children with household tuberculosis contact: a prospective community-based study. *Lancet*. 2005;**366**(**9495**):1443–51.

[23] Roy A, Eisenhut M, Harris R, Rodrigues L, Sridhar S, Habermann S, et al. Effect of BCG vaccination against Mycobacterium tuberculosis infection in children: systematic review and meta-analysis. *BMJ*. 2014;**349**:g4643.

[24] Arregui S, Sanz J, Marinova D, Martín C, Moreno Y. On the impact of masking and blocking hypotheses for measuring the efficacy of new tuberculosis vaccines. *PeerJ*. 2016;**4**:e1513.

[25] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B*. 1995;p. 289–300.

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

[27] Foss AM, Vickerman PT, Chalabi Z, Mayaud P, Alary M, Watts CH. Dynamic modeling of herpes simplex virus type-2 (HSV-2) transmission: issues in structural uncertainty. *Bull Math Biol*. 2009;**71**(**3**):720–49.