TESIS DE LA UNIVERSIDAD DE ZARAGOZA

2019 11

Sergio Arregui Remón

# Modeling Tuberculosis spreading for the evaluation of new vaccines

Departamento
Física Teórica

Director/es
MORENO VEGA, YAMIR

http://zaguan.unizar.es/collection/Tesis

ISSN 2254-7606

Prensas de la Universidad
Universidad Zaragoza

Reconocimiento – NoComercial – SinObraDerivada (by-nc-nd): No se permite un uso comercial de la obra original ni la generación de obras derivadas.

© Universidad de Zaragoza
Servicio de Publicaciones

ISSN 2254-7606

Universidad Zaragoza
1542

Tesis Doctoral

# MODELING TUBERCULOSIS SPREADING FOR THE EVALUATION OF NEW VACCINES


Autor

Sergio Arregui Remón

Director/es

MORENO VEGA, YAMIR

**UNIVERSIDAD DE ZARAGOZA**

Física Teórica


2018


**Repositorio de la Universidad de Zaragoza – Zaguan http://zaguan.unizar.es**

# Modeling Tuberculosis spreading for the evaluation of new vaccines

Sergio Arregui

Director: Yamir Moreno

Universidad de Zaragoza
Facultad de Ciencias, Departamento de Física Teórica
Instituto de Biocomputación y Física de Sistemas Complejos


# Contents

# List of Figures

# List of Tables

# Resumen

La Tuberculosis (TB) es una enfermedad infecciosa que causa más de 10 millones de nuevos casos y 1.5 millones de muertes al año. La actual vacuna, BCG, no es capaz de proporcionar una eficacia consistente, y por ello existe la imperiosa necesidad de desarrollar nuevas vacunas. En este contexto, la modelización matemática puede jugar un papel clave en la evaluación y comparación de estas nuevas vacunas con el propósito final de asistir en la elaboración de políticas y optimización de las estrategias de vacunación.

El objetivo de esta tesis es la creación de un modelo apropiado para la evaluación del impacto de estas nuevas vacunas. Para ello centramos nuestros esfuerzos en dos vertientes distintas: la modelización de la propagación de la Tuberculosis per se, y la parametrización de estas nuevas vacunas para su evaluación con estos nuevos modelos. En lo que se refiere a la modelización de la propagación de la enfermedad, los principales avances propuestos en esta tesis están relacionados con la estructura de edad de las poblaciones. Específicamente implementaremos, por primera vez en un modelo de propagación de TB, contactos dependientes de la edad y la evolución temporal de las pirámides demográficas.

Así, comenzamos la tesis estudiando el problema teórico de implementar patrones de contacto empíricos dependientes de la edad en distintas esructuras demográficas. Es una tendencia actual en epidemiología utilizar estos patrones de contacto por edades heterogéneos, superando así la asunción clásica de mezcla homogénea. Sin embargo, estos patrones de contacto han sido medidos en poco más de una decena de localizaciones diferentes, y queda pendiente la cuestión de hasta que punto unos patrones de contacto que corresponden a una población específica son transferibles a otra localización diferente. En esta tesis estudiamos distintos métodos para proyectar matrices de contacto de una población a otra con distinta estructura demográfica, y analizaremos las difierencias que existen en los patrones de contacto de distintos países. Este estudio es fundamental para la construcción de nuestro

modelo en el que pretendemos acoplar patrones de contacto por edades con una evolución temporal de la estructura por edades de la población, de forma que deberemos adaptar esas matrices de contacto por edades en cada paso temporal.

En el siguiente capítulo desarrollamos un modelo de propagación de la Tuberculosis en el que integramos una gran cantidad de datos sobre una Historia Natural para la enfermedad con 19 estados diferentes (incluyendo dos estados de latencia, tres tipos distintos de enfermedad con distinta infecciosidad, y distintos resultados del tratamiento). Así, nuestro modelo utilizará como input, datos de incidencia y mortalidad específicos de cada país, parámetros epidemiológicos con dependencia de la edad obtenidos de diferentes estudios, y, como ya hemos avanzado, incorporamos por primera vez en el campo proyecciones demográficas y matrices de contacto por edades. En este trabajo identificamos sesgos substanciales arraigados en una descripción inadequada de estos aspectos, a nivel tanto de incidencia y mortalidad agregadas como en su distribución por edades.

Una vez que la base del modelo de propagación de Tuberculosis está establecida, el siguiente paso es el estudio de la parametrización de los efectos de la vacuna en el contexto del modelo introducido. Aunque nuestro objetivo último es estudiar el impacto final que tendrán hipotéticas nuevas vacunas, es fundamental obtener toda la información posible de la actual vacuna BCG, ya que muchos de los efectos y problemas que tiene esta vacuna podrían darse también de forma inevitable en las nuevas vacunas. En concreto, sobre BCG estudiaremos la serie de ensayos clínicos BCG-REVAC, diseñados para intentar discernir qué mecanismo producido por exposición previa a micobacterias (enmascaramiento y/o bloqueo) está detrás de la variabilidad en la eficacia de BCG medida en distintos lugares. Aunque esta discusión ya había sido realizada cualitativamente, en esta tesis proponemos varios modelos matemáticos (con bloqueo, con enmascaramiento y con los dos efectos), comprobamos cúal de ellos ajusta mejor a los datos obtenidos por esta serie de ensayos clínicos y cuantificamos estos efectos.

A continuación, estudiamos el diseño de ensayos clínicos que se implementarán sobre las nuevas vacunas y que nos proveerán de toda la información posible para su evaluación con un modelo de propagación, ya que debido a la falta de correlaciones de protección de Tuberculosis, los ensayos clínicos son la única forma de determinar la eficacia de una vacuna.. La formulación clásica de estos ensayos clínicos ofrece una parametrización muy limitada de la vacuna. En concreto, ofrece un único dato de eficacia contra enfermedad,

cuando en realidad existen múltiples mecanismos con los que una vacuna puede interrumpir el ciclo del patógeno, y que permanecen indistinguibles en un ensayo clínico lo que provoca grandes incertidumbre en la posterior evaluación de impacto. Estudiaremos un nuevo diseño para estos ensayos clínicos, capaz de ofrecer una parametrización más completa de la vacuna.

Finalmente, una vez que ya hemos desarrollado un nuevo modelo de propagación de Tuberculosis y hemos estudiado en detalle la descripción de las vacunas en este contexto, evaluamos diferentes vacunas hipotéticas. Nos centraremos en el debate actual sobre la edad óptima de vacunación.

# Summary

Tuberculosis (TB) is an infectious disease that causes more than 10 million new cases and 1.5 million deaths every year. The current vaccine, Bacille Calmette-Guérin (BCG), is unable to provide a consistent efficacy, and thus there is an urgent nedd to develop new vaccines. In this context, mathematical modelization could play a key role in evaluating and comparing these new vaccines with the ultimate purpose of assist policy making and optimize vaccination strategies.

The goal of this thesis is to create the proper modelling framework for the impact evaluation of these new vaccines. For this we focus our efforts in two different pathways: the modeling of TB spreading, and the parameterization of these new vaccines for their evaluation with new models. Regarding the modeling of TB spreading, the main advances proposed in this thesis are related with the age structure of the populations. Specifically, we will implement in a model of TB spreading, for the first time, age-dependent contact patterns and the temporal evolution of demographic pyramids.

Thus, we start the thesis studying the theoretical problem of implementing empirical age-dependent contact patterns in different demographical structures. Nowadays is a common trend in epidemiology the use of these heterogeneous age-dependent contact patterns, thus surpassing the classical assumption of homogeneous mixing. However, these contact patterns have been measured in just a few different settings, and the question remains to what extent contact patterns that correspond to a specific population are transferable to a different location. In this thesis we study different methods of projecting contact matrices from one population to another with a different demographic structure, and we analyze the differences that exist between the contact patterns of different countries. This study is fundamental for the construction of our model in which we intend to couple age-dependent contact patterns with the temporal evolution of the age-structure of the population, in a way that oblige us to update contact matrices at each temporal step.

In the next chapter, we develop a TB spreading model in which we integrate an important amount of different data sources over a Natural History of the disease with 19 different states (including two latency states, three different types of disease with different infectiousness, and different treatment outcomes). Thus, our model will use as an input, data of incidence and mortality specific for each country, age-dependent epidemiological parameters obtained from different studies, and as we said before, we incorporate for the first time in the field demographic projections and age-dependent contact matrices. In this work, we identify substantial biased rooted in an inadequate description of these aspects, at the level of both aggregated incidence and mortality rates and their distribution across age strata.

Once the base modeling framework for TB spreading is set, the next step is to study the parametrization of vaccine effects in the context of the model introduced. Even though our ultimate goal is studying the final impact that hypothetical new vaccines will have, it is fundamental to obtain all possible information from current BCG vaccine, since many of the effects and problems that this vaccine has, could be inevitable by new vaccines. Specifically, about BCG we will study the series of clinical trials known as BCG-REVAC, designed to disentangle which mechanism, produced by previous exposition to mycobacteria (masking and/or blocking), is behind the variability in BCG efficacy when measured in different locations. Even though this discussion had already been made qualitatively, in this thesis we propose several mathematical models (with blocking, with masking, and with both), we check which of them fit better to the data obtained from these clinical trials and we quantify these effects.

After this, we study the clinical trials that will be implemented over new vaccines and will provide us all possible information for their evaluation through a spreading model, since due to the lack of TB correlates clinical trials are the only way of determining the efficacy of a vaccine. The classical formulation of these clinical trials offer a very limited parameterization of the vaccine. Specifically, it offers a single value of efficacy against disease, when actually there are multiple mechanisms with which a vaccine can disrupt the cycle of the pathogen, and they remain indistinguishable in a clinical trial, which provoke enormous sources of uncertainty in the posterior evaluation of impact. We will study a new design for clinical trials, that is able to offer a more complete parameterization of the vaccine.

Finally, once we have developed a new TB spreading model and we have

studied in detail the description of vaccines in this context, we evaluate different hypothetical vaccines. We focus on the current debate about the optimal vaccination age.

# Chapter 1

# Introduction

## 1.1 Mathematical Epidemiology: Models to fight infectious diseases

Since before the dawn of civilization, humanity has faced many deadly epidemics that have decimated its population. The most well known example probably is the Black Death, responsible for the death of 30-60% of Europe's total population between 1347 and 1351.[1] At that time, people did not have the tools or the knowledge to understand why and how that mantle of death was falling on them. They knew nothing about viruses, bacteria or any other agents of disease. The most pious among them could only conclude that it was God's punishment for their sins. The few physicians, learners of classical Greek medicine, attributed the disease to miasma,[2] a kind of pollution emanating from rotting organic matter.[2,3] Scholars moved later to an astrological theory, when the Medical Faculty of the University of Paris blamed the conjunction of Saturn, Jupiter and Mars for the disastrous epidemic.[2,4] Academics wanted answers, Kings needed solutions and despair grew among common folk, as the corpses of family and friends piled up in the streets. Yet, the first successful attempt at understanding, or at least somehow controlling, the spreading of an infectious disease was still centuries away.

Forgetting the irrational explanations based on gods or celestial bodies, the first step in the right direction was taken by Daniel Bernoulli in the mid-18th century. His works on inoculation against smallpox,[5–7] in which he exposed the benefits of vaccination, included what is often considered the first model in mathematical epidemiology.[8] Despite the achievements of these early works, the field was still very far from what it is today. Specifically, it was lacking a dynamical principle behind the spreading of an infectious

disease. The idea of living agents that cause illness can be traced back to the work of Aristotle (384-322 B.C.).[8] The demostration of the existence of these microorganisms is due to van Leeuwenhoek (1632-1723), and the germ theory of diseases was then developed by outstanding names such as Jacob Henle (1809-1885), Robert Koch (1843-1910), Joseph Lister (1827-1912), and Louis Pasteur (1822-1875).[8] Thus, diseases were no longer caused by the wrath of God, the capricious journeys of planets and stars or some undetermined effluvium that impregnates the air, but by living microorganisms that could pass from one host to another.

The English physician and one of the fathers of modern epidemiology, John Snow, became one of the first challengers of the widely accepted miasmatic theory, and embraced the principles of a, still in development, germ theory. In 1855, almost one century after Bernouilli's work, John Snow was able to pinpoint the source of a cholera epidemic in London by the study of the spatial and temporal pattern of the reported cases.[9–11] Specifically, he identified a water pump in Broad Street as the source of the outbreak. Even though the local authorities followed Snow's recommendations and disabled the pump, germ theory was not accepted until years later. Inspired on John Snow's discovery, William Budd identified Bristol's water supply as a source of typhoid fever in 1873.[12,13]

The birth of modern *compartmental models* still had to wait until the 20th century. In 1906, W. H. Hamer proposed that the spread of an infection should depend on the number of both susceptible and infective individuals and a mass action law for the rate of new infections.[8,14] These ideas still remain at the core of the field of mathematical epidemiology, and became the foundation of *compartmental models* in epidemiology. In 1911, Ross used a simple compartmental model of malaria, including humans and mosquitoes, that showed that the complete eradication of the mosquito population is not needed for the eradication of the disease.[15] This work introduced for the first time the concept of *basic reproduction number* (the average number of secondary infections produced by an infected individual in an otherwise susceptible host population) although it was not given a name until 1957.[8,16] The pioneer work by Ross not only brought many theoretical advances, it was also supported by field trials leading to a huge success in malaria control.[8]

Kermack and McKendrick developed the basic compartmental models, that are still used nowadays, in a set of papers between 1927 and 1933.[17–19] In these works, they introduced the SIR model that divides the population into three categories: Susceptible (S), Infectious (I) and Recovered (R). According

to this very simple model, when susceptible individuals maintain a contact with infectious individuals, they acquire the infection with probability $\lambda$:

$$S + I \xrightarrow{\lambda} I + I \tag{1.1}$$

Simultaniously, at each time step, infected individuals might recover to the disease with a probability $\mu$:

$$I \xrightarrow{\mu} R \tag{1.2}$$

The model uses only two parameters: the infectiousness $\lambda$ and the recovery rate $\mu$. All individuals are considered dynamically equivalent, which is known as the Homogeneous Mixing Approximation. Therefore, all susceptible individuals have the same probability of getting infected, which is proportional to the number of infected individuals. Furthermore, all individuals are considered to be potentially in contact with each other (well-mixed populations) and the number of infections per unit of time that takes place is $\lambda\beta SI$, where $\beta$ is the number of contacts hold by an individual per unit of time, and $S$ and $I$ are the population of the susceptible and infectious classes (and $R$ is the number of recovered individuals). The dynamics can be reduced to an ordinary differential equations system:

$$\dot{S}(t) = -\lambda\beta\frac{S(t)I(t)}{N} \tag{1.3}$$

$$\dot{I}(t) = \lambda\beta\frac{S(t)I(t)}{N} - \mu I(t) \tag{1.4}$$

$$\dot{R}(t) = \mu I(t) \tag{1.5}$$

When a small infectious seed is introduced in a susceptible population, the SIR model describes a situation where the population of infected individuals begins growing until it reaches a maximum and then starts to decay as the recovery class (which does not get the infection again: recovered individuals are inmune) gains weight in the population. Thus, this model is useful to describe rapid outbreaks as those produced by influenza-like diseases. However, the SIR model, as it has been described here, inevitable leads to the extinction of the disease (as the population of the susceptible state can only diminish), a situation that unfortunately does not apply to every case.

If the disease does not provide immunity to future re-infections, then we can consider that infected individuals, when they recover, go back to the

susceptible state. This is known as the SIS model, that can be described according to the following system:

$$\dot{S}(t) = -\lambda\beta\frac{S(t)I(t)}{N} + \mu I(t) \tag{1.6}$$

$$\dot{I}(t) = \lambda\beta\frac{S(t)I(t)}{N} - \mu I(t) \tag{1.7}$$

The dynamical behaviour of the SIS model differs greatly of the SIR, as it allows a stable *endemic* level for the disease. If we rewrite the previous equation system as a one-dimensional system (using the condition $S = N - I$, where $N$ is the total population that remains constant) we obtain:

$$\dot{I}(t) = \left(\lambda\beta\frac{N - I(t)}{N} - \mu\right) I(t) \tag{1.8}$$

The model has two solutions: $I = 0$ which is the trivial disease-free fixed point and another endemic solution given by:

$$\frac{I^*}{N} = 1 - \frac{\mu}{\lambda\beta} \tag{1.9}$$

which gives a population of infected individuals greater than zero when:

$$\lambda > \lambda_c = \frac{\mu}{\beta} \tag{1.10}$$

This critical value $\lambda_c$ is called *epidemic threshold* and gives the minimum infectiousness $\lambda$ necesary for the disease to survive in an endemic equilibrium. $\lambda_c$ has also a meaning in the context of the SIR model; $\lambda > \lambda_c$ assures the existence of an outbreak that affects a macroscopic fraction of the population.

These models, despite their simplicity, give us useful insights about the qualitative behaviour of an epidemic. Terms as the *basic reproductive number* and the *epidemic threshold* are part of the daily vocabulary of epidemiologists. However, these models are impregnated with simplified assumptions, in many ocassions unrealistic, that often introduce heavy biases in the results obtained. The description of the disease itself, for example, with only 3 or 2 different states falls short for many diseases. The incorporation of a non-infectious exposed state ($E$) is needed in many cases, obtaining the SEIR model.[20] Different manifestations of the same disease, multiple strains,[21,22] or other complex phenomenologies can also be considered.

Population dynamics is also over-simplified in these early models. For simplicity, they consider the total population $N$ as constant, which could be a good approximation for short-cycle diseases but results unrealistic for persistent diseases, for which simulations that span longer periods of time, even longer than the typical life expectancy of individuals, are needed. These models often rely upon the well-mixing hypothesis, according to which every individual of our population is equally capable of interacting with any other individual. This assumption has been abandoned progressively by more sophisticated approaches as the heterogeneous mean field,[23,24] the implementation of age-dependent contact matrices,[25] or the use of networks of contacts with different topologies in the case of individual-based models.[26–30] Currently, many models include even the mobility of individuals between different locations.[31–36]

In mathematical modeling there is always a trade-off between simple models, deprived of almost every detail, that are used to study qualitative behaviour; and more detailed models, designed for policy making, used for specific situations and including quantitative predictions. In sharp contrast to the SIR or SIS models, we have nowadays highly advanced modeling frameworks integrating worldwide high-resolution demographic and mobility data as well as a wide set of disease's compartmental descriptions.[37] Now, it is even possible to predict the peak of the influenza seasons, weeks in advance.[38,39] However, despite the growing confidence of epidemiologists, new threats such as the recent Ebola outbreak[40] or the Zika virus[41] keep challenging the field.

And while the focus of epidemiological modeling often remains at these fast outbreaks, slow killers such as Tuberculosis (TB) stay hidden. It is precisely the goal of this Thesis to develop a more realistic model of TB spreading for the evaluation of the impact of new vaccines. An enterprise that is as difficult as it is necesary.

## 1.2 The strange life of *M. tuberculosis*

The disease of Tuberculosis is caused by the pathogen *Mycobacterium tuberculosis* (MTB), that was first identified as a pathogen by Robert Koch in 1882.[42,43] The life cycle of MTB inside the host body and the interactions between host and pathogen are extremely complex, and present a series of particular traits that we have to take into account for a proper modeling of TB spreading dynamics.

TB is spread through the air when people with active TB in their lungs cough or sneeze, releasing tiny droplets with the pathogen in it that are later inhaled by susceptible individuals. Thus, TB does not diverge of the germ theory that forms the basis of most compartmental models. The first nuance we have to consider is that, although MTB attacks to the lung 70% of the cases,[44] it can also disseminate to other organs such as lymph nodes, bone and meninges, causing what is called extra-pulmonary TB.[45] The distinction between different types of TB plays a major role in the correct assessment of the spreading of the bacterium for two reasons. On the one hand, as only individuals with the pathogen hosted at the lungs can infect other susceptible people, non pulmonary TB is considered as mainly not infectious. On the other hand, diagnosis times for both forms of disease are very different. Non pulmonary TB, with its wide range of possible manifestations, is in general more difficult to detect than the pulmonary TB, which poses less of a diagnostic challenge for health practitioners.[46]

Even more important is the fact that most infections never progress to active disease and remain asympomatic (what is known as Latent TB Infection, LTBI).[47,48] It is estimated that one third of the world population suffers from LTBI,[48,49] but only about 5–10% of infected individuals will develop active TB disease in their lifetimes,[50,51] a risk that is much higher among HIV infected individuals.[52] The existence of two different time scales in the process from infection to active disease brings another challenge for spreading modeling. Thus, while some individuals (fast-progressors) develop active disease right after infection, those who develop LTBI form a reservoir of potential transitions to active disease that, even if they occur at a much slower rate, can not be ignored due to the large fraction of individuals that suffers from LTBI.

Furthermore, these different immunogenic mechanisms depend on the age of the host that suffers the infection.[53] Fast-progressors are more common among newborns and adults than among school-aged children.[54] Also, adults tend to develop pulmonary TB (thus being more infectious), while children usually develop non-pulmonary TB.[55]

The typical SIR or SEIR models with only 3 or 4 compartments can not be adapted to the nuanced nature of TB disease. We need to add different latency states (with different time scales), different forms of active TB (with different infectiousness) and an age-structure in our population to differentiate all those processes that depend on the age of the host. And that only covers the basic needs of any good TB spreading model, without accounting

for other specific situations such as the appearence of drug-resistant strains or co-infection with HIV. Definitely, we need more complex models.

## 1.3 A brief look at the epidemiology of TB

Up to this point, one should consider whether it is worth the enormous effort that supposes the modelization of the spreading of TB, a disease that seems more typical of romanticism than of the contemporary era. TB is in fact a very old disease, whose origin can be traced back at least to the migrations of humans out of Africa during the Neolithic period about 70,000 years ago.[56] TB has been without a doubt one of the deadliest diseases that has persecuted humanity across history, causing the death of one in five adults in Europe and North America between the seventeenth and nineteenth centuries.[57] Fortunately, this picture improved during the last century, and an increasing availability of antibiotics and diagnosis tools led to a sustained decrease in the incidence levels of the disease,[58] and even complete eradication did not seem to be out of reach.[59]

This optimistic panorama changed in the past few decades, and control of TB is facing new challenges. The emerging of HIV, with its ability to compromise the immune system of affected individuals, favors the proliferation of other epidemics including TB.[60–66] The appearance of multi-drug and extensively drug resistant strains of MTB (i.e. strains that are resistant to antibiotic treatment) also imposes a major threat.[67–70] TB has become a global emergency, causing 10 million new cases and almost 2 million deaths every year.[71]

Nowadays the most important tool that public health authorities have to fight TB is the BCG (Bacille Calmette-Guerin) vaccine, but the reach of its immunogenic action is in doubt.[72] While it appears to be effective in protecting newborns from the non-pulmonary forms of meningeal and milliary TB, its effect in adults against the most infectious manifestations of TB is almost null.[73–76] Besides the dependence on age, efficacy of BCG also presents a dependency with latitude, hinting a possible interference with environmental mycobateria.[76–78]

Thus, the development of a new vaccine aimed at either substituting BCG or enhancing its immunogenic power is considered essential for the control and eventual eradication of TB.[79,80] Several vaccines are right now under development,[81] hoping to complete all different stages within the next

few years.[82–85] But resources are limited, and not all different approaches to TB vaccination can be funded. In this context, having a highly detailed modeling framework is essential to guarantee a correct resource allocation and to optimize new interventions against TB.

## 1.4 TB Modeling: what is done, what is left to do

The field of TB spreading modeling has not remained isolated from this situation of public-health emergency, and in the past decades has experienced an important growth. In a recent review by Rebecca Harris et al.[86], the authors identify 23 different articles that used modelization to evaluate TB interventions, 20 of them published after 2000.

We find many different approaches between these articles. The most common is the use of deterministic compartmental models using difference or differential equations, in the same spirit of the early works by Kermack and McKendrick. Up to 18 articles used this methodology.[55,87–103] Other possible methodologies include Markov decision trees,[79,104,105] a simpler model with a fixed number of transmissions per case,[106] and a statistical model.[107]

The range of complexity of these works, in what regards the description of the disease, comes from the most generic SIR or SEIR models, to models that address some of the most important treats of TB that we discussed before. Thus, we can find some models that incorporate two diferent rates of progression from infection to disease, different forms of active disease with different infectiousness, and an age-structured population with an age dependent on some of the dynamical parameters that describe the disease.[55,101,108] Some of these models also address more specific scenarios. The work by Cohen et al.[92] include co-infection of two strains, and the model developed by Knight et al.[101] differentiates the population according to the HIV status (infected or not).

However, there are still some aspects that would make these models more realistic. We have discussed the importance that the age of the host has in the development of the disease. Nevetheless, we find two important oversimplifying hypotheses regarding the age structure of the population in these models. On the one hand, we have that these models consider constant demographic pyramids even when the simulations cover very long periods (decades). Thus,

it remains unexplored how a realistic evolution of the demographic structure of the population would compromise epidemiological outcomes. On the other hand, these models assume homogeneous random mixing between individuals, meaning that individuals mix among them randomly, without considering the age of the individuals. We know from empirical studies[25] that this assumption is not true. Individuals tend to maintain contacts with people of the same age, with the appearance of some inter-generational contacts, typically connecting members of different generations within the same family. Although the inclusion of heterogeneous age-dependent contact patterns has been widely implemented recently for short-cycle influenza-like diseases,[26,109–113] for the case of TB the implications of heterogeneity remain vastly unexplored (up to our knowledge there is only one model of TB that incorporates realistic contacts in a low burden setting[114]). Besides, the implementation of heterogeneous contacts in recent works has always been done in bounded settings (both geographically and temporally), leaving unsolved theoretical and technical challenges.

Ultimately, the goal of these sophisticated models is to serve as a tool for policy making, allowing to evaluate new interventions, vaccines and treatments. The way we introduce these interventions in our models is also full of nuances. Specially complex is the situation of TB with tens of new vaccines under development, very different in nature, around an immunology that is not completely understood.[115] The possibilities of new vaccines comprise a wide range of effects.[116] The challenge for mathematical modeling could not be greater.

The differences among the wide range of possible vaccines currently under development might affect the age groups on which these are administered.[115,116] While some of the vaccines are thought as a substitute of BCG, and therefore to be applied in newborns,[117] others are being designed as a booster of BCG with an optimal age of deployment around adolescence where BCG loses its efficacy.[79] Therefore, the importance of achieving a detailed description of the coupling between age and TB transmission gains even more weight, and an oversimplification of demographic evolution and social mixing with unrealistic assumptions is even more likely to introduce substantial biases to model outcomes. Even though, previous models have compared the impact and cost-effectiveness of newborn and adolescent vaccination, concluding that adolescent vaccination generates greater and faster impact.[101] One of the main goals of this work is that of quantifying the effects on the impact derived from adolescents versus newborn targetted vaccination campaigns that can be tracked back to over-simplified description of the coupling

between TB transmission and populations' age structure.

Another layer of phenomenology has to do with the specific mechanism with which the vaccine interrupts the pathogen's life cycle. During many years it was believed that BCG's efficacy came from an interrumption in the progression from infection to disease (what is known as Prevention of Disease or PoD). Recently meta-analyses of clinical trials showed that it also offers some protection against the infection itself (Prevention of Infection, PoI).[75] But other possible mechanisms remain unexplored: PoD can be achieved by reducing the rate of progression or by avoiding fast-progression directly and we do not know the exact effect that takes place, the effect over LTBI is also unknown, etc. And what is still unknown for BCG, a vaccine that was discovered 100 years ago, will hardly be unravelled for new vaccines. This forces modelers to work in a large parameter space and to extract as much information as possible from clinical trials whose designs are affected by a series of relevant economic, logistic and conceptual constraints. Clinical trials are the only way to determine the efficacy of a vaccine, as there are no correlates of protection for TB[118–120], a piece of information that is essential for its posterior evaluation of impact and cost-effectiveness. Unfortunately the knowledge that we can extract from these trials is very limited, specially in the complex scenario of TB with multiple possible interruptions of the pathogen's cycle that can be triggered by the vaccine. Thus there exists a urge need of improving the assessment of efficacy provided by these trials.

A final effect that should be taken into account is the sensitization with environmental non-tuberculous mycobacteria (NTM). Previous sensitization to NTM can interfere in the assessment of efficacy of a vaccine in two ways: by blocking its immunogenic power or by offering a protection that masks the protection given by the vaccine.[121] One or both of these effects are thought to be behind the very dispair efficacies shown by BCG in different settings.[121] The implications for modeling is twofold. On the one hand it brings another age-dependent effect: newborns will not suffer from previous exposure to NTM. On the other hand, the levels of NTM varies greatly across different countries, difficulting to transfer results and simulations from one setting to another.

The BCG-REVAC trials tried to unravel the mechanisms behind this interference by NTM in the efficacy assessment of BCG.[77] By vaccinating cohorts at different ages in two different cities (Manaus and Salvador, supposed to have different levels of NTM), researchers reached the conclussion that blocking was the most plausible effect, but they could not quantify its

impact.

## 1.5 Structure of the thesis

The challenges faced by the field of TB modeling are numerous. It is the goal of this thesis to tackle some of them and open new pathways in the field. After the introduction in Chapter 1, the thesis is organized as follows.

In Chapter 2 we deal with the theoretical challenge that supposes the implementation of heterogeneous age-dependent contact patterns in the scenario of evolving demographic structures. Although we are more interested in the modeling of TB spreading, the framework developed in this chapter is transversal and of interest for the modeling of any disease.

In Chapter 3 we developed a model for TB spreading that, for the first time, incorporates empiric age-dependent contact rates and realistic evolution of demographic structure (thus using the framework developed in Chapter 2). We will show in this chapter that the incorporation of these new features significantly changes epidemiological forecasts.

In Chapter 4 we will study the mechanisms of masking and blocking produced by previous sensitization with environmental non-tuberculous mycobacteria over the inoculation of BCG. We will analyze the BCG-REVAC trials in order to parameterize these two effects. Our results show that blocking has a greater effect than masking, and therefore we should take this into account when designing new vaccination campaigns.

In Chapter 5 we will focus on the desing of efficacy clinical trials. We propose new analyses for these trials that will allow for a more complete parameterization of the effect of a vaccine, which will improve the accuracy of posterior impact evaluations.

Finally, in Chapter 6 we will use the model developed in Chapter 3 to evaluate hypothetical new vaccines. For the implementations of these new vaccines the lessons learnt in Chapters 4 and 5 will also be of use. We will focus specifically in the controversial debate of stablishing which vaccination strategy is better: a replacement of BCG with a better vaccine, or a booster of BCG applied during the adolescence.

The results contained in this Thesis have been published in the following

publications and pre-prints.

- S. Arregui, J. Sanz, D. Marinova, C. Martin and Y. Moreno. On the impact of masking and blocking hypotheses for measuring the efficacy of new tuberculosis vaccines. PeerJ 4, e1513
- S. Arregui, M.J. Iglesias, S. Samper, D. Marinova, C. Martin, J. Sanz and Y. Moreno. Data-driven model for the assessment of Mycobacterium tuberculosis transmission in evolving demographic structures. Proceedings of the National Academy of Sciences, 201720606
- S. Arregui, D. Marinova, C. Martin, J. Sanz and Y. Moreno. Design principles for TB vaccines' clinical trials based on spreading dynamics. bioRxiv, 249847
- S. Arregui, A. Aleta, J. Sanz and Y. Moreno. Projecting social contact matrices to different demographic structures. bioRxiv, 343491

# Chapter 2

# Projecting age-dependent contact patterns into different demographic structures

## 2.1 Introduction

During recent years, models on disease transmission have improved in complexity and depth, integrating high-resolution data on demography, mobility and social behavior.[37,122] Specifically, the topology of social contacts plays a major role in state-of-the-art modeling.[26,109–113] The complete knowledge of the network of contacts through which an epidemic spreads is usually unreachable or impossible to implement, and for modeling purposes it is useful to remain at the coarse level of age-groups. Under this view, the population under study is divided into different groups, according to its age distribution, and different contact rates are assumed among these groups. Age-dependent contact patterns give powerful insights on the transmission of diseases where epidemiological risk is correlated to age, either as a result of behavioral or physiological factors. Relevant examples are influenza-like diseases,[111–113,123,124] pertussis,[125] and varicella.[126] Furthermore, they are instrumental for modeling and implementing more efficient interventions.[127,128]

Age-dependent patterns will also be relevant for the assessment of TB spreading,[114,129] which is the reason why this chapter is relevant for this thesis. As we have explained throughout the introduction, the effects caused by the bacteria depend strongly on the age of the host. Not all age groups will have the same ability to transmit the disease, and therefore it will be essential to implement realistic social mixing in a model of TB, specially considering

that this has been a question overly ignored in previous literature.[55,101]

Notwithstanding that, and given the utmost importance of contact heterogeneities, the study of age-dependent social mixing has become a priority in epidemiology. In 2008, Mossong et al.[25] published a seminal work with the measurements of age-dependent contact rates in eight European countries (Belgium, Finland, Germany, Great Britain, Italy, Luxembourg, Netherlands and Poland) via contact diaries. Other authors have replicated this work in other countries such as China,[130] France,[131] Hong-Kong,[132] Japan,[133] Kenya,[134] Russia,[135] Uganda,[136] and Zimbabwe,[137] thus expanding significantly the available data on social mixing in the last few years. In these studies, participants are asked how many contacts they have during a day and with whom. This allows to obtain the (average) number of contacts that an individual of a particular age $a$ has with individuals of age-group $a'$. The resulting matrix is not symmetric due to the different number of individuals in each age-group. However, it is precisely the demographic structure what imposes constraints in the entries of this matrix, as reciprocity of contacts should be fulfilled at any time (i.e., the total number of contacts reported by age-group $a$ with age-group $a'$ should be ideally equal in the opposite direction). Therefore, an empirical contact matrix, that has been measured on a specific population, should not be used directly, without further considerations, in another population with a different demographic structure.

This issue has important consequences in the field of disease modeling. As contact matrices play a key role in disease forecast, it is essential to assure that the matrices implemented are adapted to the demographic structure of the population considered in order to avoid biased estimations. For some short-cycle diseases like influenza, the time scale of the epidemic is much shorter than the typical times needed for a demographic structure to evolve. That means that, typically, the demographic structure can be safely considered constant,[124] and the eventual evolution of the contact matrix can be neglected throughout the simulation of an outbreak. For these diseases, the problems might arise when modelers use contact matrices that are not up to date. For instance, one might wonder whether the patterns reported in Mossong et al.[25] in 2008 can be used nowadays, a decade later, during which all the European countries analyzed in that study aged significantly. The same issue appears when a contact matrix measured in a given location (e.g., a specific country) is directly used to simulate disease spreading in another region or country with a different population structure.

The previous considerations are even more troublesome for the case of

persistent diseases that need long-term simulations (as will be the case with TB), for which the hypothesis of constant demographic structures does not hold anymore.[129] In those cases, contact matrices should continuously evolve during the simulation to reflect the effect that an evolving demography should exert on contact structures. As the goal of this thesis is precisely to develop a model for TB spreading that includes realistic demographic projections and empiric contact patterns, we should solve this problem before.

Furthermore, it remains to be known up to what extent the variations between contact matrices coming from different countries are due to differences in the demographic structures, divergent cultural traits and/or methodological differences between studies. For instance, elderly people exhibit higher contact rates with children in African countries than in Europe.[137] This could be explained by the different demographic structures: one might expect to observe higher contact rates toward the younger age strata in Africa than in Europe because their populations have a higher density of young individuals. However, it is not clear yet whether the demographic structure is the only driver of geographical heterogeneity between empirical contact matrices.

The main focus of this chapter is to study how contact matrices among age-groups, originally obtained for a specific setting (country and year), can be adapted to different demographic structures ,i.e., to another (location and/or time) setting. To this end, we first study the magnitude of the reciprocity error incurred when the adaptation of empirical social contacts to different age structures is ignored, thus justifying the need of studying possible projections that solve this problem. Next, we analyze different methods to perform these adaptations, highlighting the differences induced in the contact patterns by the use of these methods. We also compare empirical contact matrices of 16 countries in different areas worldwide filtering the influence of the demographic structure. This allows to isolate what are the differences that are caused by other factors such as cultural traits. Finally, we implement a SEIR model to study the differences in prospected incidences that arise when applying the methods analyzed to project social contact matrices. The considerations made in this chapter will be crutial for the posterior development of a TB spreading model.

## 2.2 Materials and Methods

### 2.2.1 Data used

For this work we have gathered 16 different contact matrices coming from several countries: 8 from the POLYMOD project[25] (Belgium, Finland, Germany, Great Britain, Italy, Luxembourg, Netherlands and Poland), China,[130] France,[131] Hong-Kong,[132] Japan,[133] Kenya,[134] Russia,[135] Uganda,[136] and Zimbabwe.[137]

Data regarding the time evolution of demographic structures, either observed in the past or projected until 2050, has been retrieved from the UN population division database.[138]

### 2.2.2 Different Scales to define Contact Patterns

The contact matrices are often described with the matrix $M_{a,a'}$ that is the number of contacts that an individual in age group $a$ maintains (on average) with individual in age group $a'$ during a certain period of time. We will refer to this quantity as the intensive scale. However, sometimes two other scales are used:

- Extensive scale: $C_{a,a'}$ that is the total number of contacts between age-groups $a$ and $a'$.
$$C_{a,a'} = M_{a,a'} N_a \tag{2.1}$$
- Density scale: $F_{a,a'}$ that is the fraction of contacts between age-groups $a$ and $a'$ among all possible contacts that actually take place.
$$F_{a,a'} = \frac{C_{a,a'}}{N_a N_{a'}} = \frac{M_{a,a'}}{N_{a'}} \tag{2.2}$$

where $N_a$ is the population of age group $a$.

These two scales are symmetric by construction. Therefore they might seem a better option to work with, as they directly avoid the problem of non-reciprocity. However, unlike $M_{a,a'}$ they do not remain invariant when the total number of individuals increases maintaining the shape of the demographic structure, so a correction for the total population will always be needed. In this work, we describe magnitudes in the intensive scale mostly, but the translation to other scales is always possible through equations 2.1 and 2.2. It is also possible to work with any of these scales as long as the force of infection is properly defined.

### 2.2.3 Treatment of empirical survey matrices

We need to perform some transformations before we can compare the different empirical studies:

- Most of the studies report the matrix in the intensive scale ($M_{a,a'}$). Some of these studies however (specifically France,[131] Japan,[133] and China[130]) report their results using the density scale (see equation 2.2) except for a global factor. Thus, we transform them so they operate in the scale of $M_{a,a'}$, assuming the demographic structure as reported by the United Nations (UN) Population Database[138] for the specific year of the survey.

- The studies use different age granularities. We adapt all surveys to the division of age groups used by the POLYMOD project, i.e., 15 age groups: 0-5, 5-10, 10-15,..., 65-70, 70+. For this adaptation sometimes we have to aggregate age-groups (some studies have more divisions for young children) but mostly we have to divide age-groups. In order to do so, we assume homogeneity inside the broader original age-groups. The studies with less age-groups (China that has 4, and Kenya that has 6) are therefore more sensitive to this process.

In Table 2.1 we summarize the information about the 16 empirical contact matrices collected for this work.

| Country | Year | Age Groups | Participants | Reference |
|---|---|---|---|---|
| Belgium | May 2005 - September 2006 (2005) | 15: 0-5, 5-10,...,65-70,70+ | 750 | Mossong et al. (POLYMOD)[25] |
| Finland | May 2005 - September 2006 (2005) | 15: 0-5, 5-10,...,65-70,70+ | 1006 | Mossong et al. (POLYMOD)[25] |
| Germany | May 2005 - September 2006 (2005) | 15: 0-5, 5-10,...,65-70,70+ | 1341 | Mossong et al. (POLYMOD)[25] |
| Great Britain | May 2005 - September 2006 (2005) | 15: 0-5, 5-10,...,65-70,70+ | 1012 | Mossong et al. (POLYMOD)[25] |
| Italy | May 2005 - September 2006 (2005) | 15: 0-5, 5-10,...,65-70,70+ | 849 | Mossong et al. (POLYMOD)[25] |
| Luxembourg | May 2005 - September 2006 (2005) | 15: 0-5, 5-10,...,65-70,70+ | 1051 | Mossong et al. (POLYMOD)[25] |
| Netherlands | May 2005 - September 2006 (2005) | 15: 0-5, 5-10,...,65-70,70+ | 269 | Mossong et al. (POLYMOD)[25] |
| Poland | May 2005 - September 2006 (2005) | 15: 0-5, 5-10,...,65-70,70+ | 1012 | Mossong et al. (POLYMOD)[25] |
| France | February - May 2012 | 15: 0-5,5-10,...,65-70,+70 | 2033 | Béraud et al.[131] |
| Russia | January - February 2016 | 12: 0-5,5-10,...,50-55,+55 | 505 | Ajelli and Litvinova[135] |
| China | 2009 - 2010 (2009) | 4: 0-5,6-19,20-64,+65 | 1821 | Read et al.[130] |
| Japan | Spring 2011 | 12: 0-2,3-5,6-11,12-14,15-19,20-29,...,70-79,+80 | 3146 | Ibuka et al.[133] |
| Hong Kong | 2015-2016 (2015) | 15: 0-5,5-10,...,65-70,+70 | 1149 | Leung et al.[132] |
| Kenya | August 2011 - January 2012(2011) | 6: 0-1, 1-5, 6-15, 16-19, 20-49, +50 | 1080 | Kiti et al.[134] |
| Uganda | January - March 2014 | 10: 0-2, 2-4,5-9, 10-14, 15-24,...,55-64,+65 | 568 | Le Polain de Waroux et al.[136] |
| Zimbabwe | March - August 2013 | 16: 0-5,5-10,...,70-75,75-80 | 2490 | Melegaro et al.[137] |

Table 2.1: **Information about the 16 contact matrices collected.** When the period of the study comprises more than one year, we specify in parenthesis the year taken as reference for posterior analysis.

From each one of these studies, and after performing the transformations previously specified (when needed) we obtain a contact matrix $M^{(surv.)}_{a,a'}$. However, due to limitations in the survey and errors in the reporting process,

these matrices will not fulfil reciprocity perfectly. Thus, we perform an initial correction for reciprocity, that consists of averaging the number of contacts measured in one direction ($a$ to $a'$) and the reciprocal ($a'$ to $a$), which for the $C$ matrix means:

$$C_{a,a'}^{(surv.)} = \frac{M_{a,a'}^{(surv.)} N_a n_a + M_{a',a}^{(surv.)} N_{a'} n_{a'}}{n_a + n_{a'}} \tag{2.3}$$

and, when taken back to the $M$ scale:

$$M_{a,a'}^{(rec.)} = \left( M_{a,a'}^{(surv.)} n_a + M_{a',a}^{(surv.)} \frac{N_{a'}}{N_a} n_{a'} \right) \frac{1}{n_a + n_{a'}} \tag{2.4}$$

where $n_a$ is the number of participants in age-group $a$ and $N_a$ is the population of age-group $a$ in the corresponding country during the specific year of the survey as extracted from the UN database.[138] Thus, we are applying a weighted average that needs the age-distribution of the participants. When that information is not available, we perform a regular average. This process can also induce some bias since for some cases, the survey of contacts has been performed in some specific settings (rural or semi-urban) that might not be representative of the whole country. This correction of reciprocity has been used before, either to present empirical data[137] or to ensure reciprocity in a simulation by adapting the contact patterns to the specific demographic structure considered.[139,140] The possibility to use this pair-wise correction, which we will label later as Method 1 (M1) to solve the problem of reciprocity will be explored in this chapter.

In equation 2.4, we have named the resultant matrix $M_{a,a'}^{(rec.)}$, where the super-index indicates that the matrix fulfils reciprocity. We still apply one more transformation, that consists of normalizing the matrices so that the mean connectivity of the population is equal to 1. The mean connectivity follows the formula:

$$\langle k \rangle = \frac{\sum\limits_{a,a'} M_{a,a'}^{(rec.)} N_a}{N} \tag{2.5}$$

where $N = \sum\limits_a N_a$. Thus, we obtain:

$$M_{a,a'}^{(rec.+norm.)} = \frac{M_{a,a'}^{(rec.)}}{\langle k \rangle} = \frac{M_{a,a'}^{(rec.)} N}{\sum\limits_{a,a'} M_{a,a'}^{(rec.)} N_a} \tag{2.6}$$

Hence, we are not addressing the average intensity of contacts (and how it differs between countries) and we focus exclusively on the relative differ-

ences between age-groups. Although geographical differences in global contact rates is a problem of interest, we have not studied them for three main reasons. First, for France, Japan and China, the studies report a matrix that is proportional to $F_{a,a'}$ and we do not have direct information on the intensity of contacts. Second, it depends on the definition of contact itself and therefore methodological biases could arise. Most studies define a contact as any interaction that includes physical touching and/or face-to-face conversation for a minimal time period (that varies between studies), and even report data on physical contact separatedly, but some only consider one type of contact (only physical contacts in Kenya,[134] and only conversation in Russia[135] and Uganda[136]). And third, for many problems, the scaling of this contact matrix is performed independently and therefore not having the overall intensity of contacts is irrelevant.[113,129]

After this process we end up with the matrix $M_{a,a'}^{(rec.+norm.)}$ (that we will name simply as $M_{a,a'}$ from now on) for each one of the 16 countries for which we have contact data. We assume that, except for the limitations in the survey, these matrices are valid in their respective countries during the year when they were measured.

### 2.2.4 Projections of a Contact Matrix

The basic problem explored in this work is: how can we transform the (empirical) contact matrix $M_{a,a'}$, that has been measured for a specific demographic structure $N_a$, into a different contact matrix $M'_{a,a'}$ that is compatible with a different demographic structure $N'_a$? This could mean to adapt data obtained in one specific country to another different region that has a different demography. But the problem can appear even if we remain in the same geographical setting, as a contact matrix measured at a specific time $\tau$, could not be valid for an arbitrary time $t$ if the demographic structure of that population has changed. In the following sections, we formulate the problem of non-reciprocity and we present and discuss different methods of using contact matrices in an arbitrary demographic structure.

#### 2.2.4.1 Method 0 (M0): Unadapted Contact Matrix. The problem of non-reciprocity

The number of contacts must fulfil reciprocity, i.e., there is the same number of total contacts from age-group $a$ to age-group $a'$ than from $a'$ to $a$. This imposes the following closure relation for the contact matrix:

$$M_{a,a'}N_a = M_{a',a}N_{a'} \Longrightarrow \frac{M_{a,a'}}{M_{a',a}} = \frac{N_{a'}}{N_a} \tag{2.7}$$

where $N_a$ is the number of individuals of age-group $a$.

Therefore, in the case of an evolving demographic structure for which the ratio $\frac{N_a}{N_{a'}}$ is not constant; the contact matrix $M_{a,a'}$ must change with time. Otherwise we will have non-reciprocal contacts (contacts that inconsistently appear in one direction but not in the other). When comparing different methods for correcting for reciprocity we will usually also compare with the case in which this problem is completely ignored, and the matrix $M_{a,a'}$ is taken directly from the survey without any further consideration. We will refer to this case as Method 0 (M0).

The following methods correct this problem, introducing different transformations of the original contact matrix $M_{a,a'}$, that was measured in a demographic structure $N_a$, into a new contact matrix $M'_{a,a'}$ that is well adapted to a new demographic structure $N'_a$ (at least avoiding the problem of no reciprocity).

#### 2.2.4.2 Method 1 (M1): Pair-wise correction

The basic problem that we want to avoid is to have a different number of contacts measured from $a$ to $a'$ than from $a'$ to $a$. Thus, an immediate correction would be to simply average those numbers, so the excess of contacts measured in one direction is transferred to the reciprocal direction. This correction can be formulated as:

$$M'_{a,a'} = \frac{1}{N'_a}\frac{1}{2}\left(M_{a,a'}N'_a + M_{a',a}N'_{a'}\right) = M_{a,a'}\frac{1}{2}\left(1 + \frac{N_a N'_{a'}}{N_{a'}N'_a}\right) \tag{2.8}$$

Previous works that have used this approach include Apolloni et al. [112] and Riolo et al. [141]

#### 2.2.4.3 Method 2 (M2): Density correction

An alternative approach is to adapt contact patterns to different demographic structures correcting by the density of available contactees, which we formalize with the following equation:

$$M'_{a,a'} = \Gamma_{a,a'}\frac{N'_{a'}}{N'} \tag{2.9}$$

Thus, we interpret that the matrix $M_{a,a'}$ is the product of two factors:

- The intrinsic connectivity matrix: $\Gamma_{a,a'}$
- The fraction of individuals in $a'$: $\frac{N'_{a'}}{N'}$

Thus, we are assuming that an individual has an intrinsic preference over certain age-groups depending on its age, captured by $\Gamma_{a,a'}$, and the final contact rate is modified according to the density of available contactees.

The matrix $\Gamma_{a,a'}$ corresponds, except for a global factor, to the contact pattern in a "rectangular" demography (a population structure where all age groups have the same density). We can obtain these matrices $\Gamma_{a,a'}$, that are country-specific, from survey data using equation 2.9:

$$\Gamma_{a,a'} = M_{a,a'} \frac{N}{N_{a'}} \tag{2.10}$$

which allows to rewrite equation 2.9 as a function of the original matrix $M_{a,a'}$:

$$M'_{a,a'} = M_{a,a'} \frac{N N'_{a'}}{N_{a'} N'} \tag{2.11}$$

This methodology for adapting contact patterns has already been used by De Luca and collaborators, introducing the matrix $\Gamma_{a,a'}$ directly in the force of infection.[113] Also a similar correction is used in Prem et al.[142] to adapt European contact matrices to other countries (although this work integrates more data beyond demographic structures).

#### 2.2.4.4 Method 3 (M3): Density correction + Normalization

A cardinal feature of M2 is that it does not preserve the mean connectivity of the entire network of contacts. As a result, depending on the initial contact matrix and the dynamics of the demography, the evolution of the contact structure can produce average connectivities that depart strongly from its initial value. For the sake of disease modeling, this situation is essentially irrelevant if the contact rate of the outbreak to model can be calibrated at its early stages (i.e. its reproductive number). In that case, any global scaling factor multiplying the contact matrix is absorbed by the estimation of a larger or smaller infectiousness $\beta$. However, if that is not the case and epidemiological parameters measured in the past (i.e. a pathogen's infectiousness) are used to generate forecasts of independent outbreaks that might occur later in time, the overall scaling factor of the contact networks become

extremely relevant. In such scenario, to couple an a-priori characterization of a pathogen's infectiousness on top of contact networks with different mean connectivities will artificially inflate or shrink the size of modeled epidemic events as a function of time. Although considering an evolution of the mean connectivity as demography changes might be reasonable, the inability of M2 of producing contact matrices of stable mean connectivities might suppose a liability in some scenarios. This is the approach we will take in the model of TB spreading, as will be motivated later.

Taking that potential issue into consideration, we have proposed an alternative approach that, in addition of correcting for the densities of contactees, preserves the mean connectivity of the overall system across time. Thus, an evolution of the mean connectivity could always be forced by adding a global factor in a controlled way.

To do so, we begin by defining $\tilde{M}_{a,a'}$ as the connectivity matrix from M2:

$$\tilde{M}_{a,a'} = \Gamma_{a,a'} \frac{N'_{a'}}{N'} \tag{2.12}$$

and then we divide it by its connectivity:

$$M'_{a,a'} = \frac{\tilde{M}_{a,a'}}{< \tilde{k} >} \tag{2.13}$$

Thus:

$$M'_{a,a'} = \frac{\Gamma_{a,a'} N'_{a'} N'}{\sum\limits_{a,a'} \Gamma_{a,a'} N'_a N'_{a'}} = M_{a,a'} \frac{N'_{a'}}{N_{a'}} \frac{N'}{\sum\limits_{a,a'} M_{a,a'} \frac{N'_a N'_{a'}}{N_{a'}}} \tag{2.14}$$

Notice that all methods trivially coincide in the year in which the data was obtained (i.e. when the survey was done). Also the definition of $\Gamma_{a,a'}$ does not change between M2 and M3 in these cases, as the initial $M_{a,a'}$ has been normalized to have a mean degree of 1, and we extract it with the same equation as before (eq. 2.10).

### 2.2.5 Properties of projection methods

In this section we will study analytically some properties of the different methods we have just introduced. We will quantify these properties in section 2.3

The different methods of transforming a matrix $M_{a,a'}$ into a matrix $M'_{a,a'}$ when moving from a demography $N_a$ to a demography $N'_a$ that we have studied can be formulated as:

- M0: Unadapted Contact Patterns.

$$M'_{a,a'} = M_{a,a'}. \tag{2.15}$$

- M1: Pair-wise Correction

$$M'_{a,a'} = \frac{1}{N'_a}\frac{1}{2}\left(M_{a,a'}N'_a + M_{a',a}N'_{a'}\right) = M_{a,a'}\frac{1}{2}\left(1 + \frac{N_a N'_{a'}}{N_{a'}N'_a}\right). \tag{2.16}$$

- M2: Density Correction

$$M'_{a,a'} = M_{a,a'}\frac{NN'_{a'}}{N_{a'}N'}. \tag{2.17}$$

- M3: Density Correction + Normalization

$$M'_{a,a'} = M_{a,a'}\frac{N'_{a'}}{N_{a'}}\frac{N'}{\sum\limits_{a,a'} M_{a,a'}\frac{N'_a N'_{a'}}{N_{a'}}}. \tag{2.18}$$

#### 2.2.5.1 Reciprocity

The main property that a Contact Matrix should fulfill is reciprocity in the number of contacts:

$$M_{a,a'}N_a = M_{a',a}N_{a'} \Longrightarrow \frac{M_{a,a'}}{M_{a',a}} = \frac{N_{a'}}{N_a}. \tag{2.19}$$

Thus, when moving to a different demographic structure (where, in general, the ratio $\frac{N_{a'}}{N_a}$ will change), M0 does not preserve reciprocity:

$$\frac{M'_{a,a'}}{M'_{a',a}} = \frac{M_{a,a'}}{M_{a',a}} = \frac{N_{a'}}{N_a} \neq \frac{N'_{a'}}{N'_a}. \tag{2.20}$$

However, it is easy to see that the rest of the methods do preserve reciprocity:

- Reciprocity in M1:

$$\begin{aligned}\frac{M'_{a,a'}}{M'_{a',a}} &= \frac{M_{a,a'}}{M_{a',a}}\frac{1+\frac{N_a N'_{a'}}{N_{a'}N'_a}}{1+\frac{N_{a'}N'_a}{N_a N'_{a'}}} = \frac{N_{a'}}{N_a}\frac{\frac{N_{a'}N'_a+N_a N'_{a'}}{N_{a'}N'_a}}{\frac{N_a N'_{a'}+N_{a'}N'_a}{N_a N'_{a'}}} \\ &= \frac{N'_{a'}}{N'_a}\frac{N_{a'}N'_a + N_a N'_{a'}}{N_a N'_{a'} + N_{a'}N'_a} = \frac{N'_{a'}}{N'_a}.\end{aligned} \tag{2.21}$$

- Reciprocity in M2:

$$\frac{M'_{a,a'}}{M'_{a',a}} = \frac{M_{a,a'}}{M_{a',a}} \frac{NN'_{a'}}{N'N_{a'}} \frac{N'N_a}{NN'_a} = \frac{N_{a'}}{N_a} \frac{N'_{a'}}{N_{a'}} \frac{N_a}{N'_a} = \frac{N'_{a'}}{N'_a}. \tag{2.22}$$

- Reciprocity in M3:

$$\frac{M'_{a,a'}}{M'_{a',a}} = \frac{M_{a,a'}}{M_{a',a}} \frac{N'_{a'}}{N'_a} \frac{N_a}{N_{a'}} = \frac{N_{a'}}{N_a} \frac{N'_{a'}}{N'_a} \frac{N_a}{N_{a'}} = \frac{N'_{a'}}{N'_a}. \tag{2.23}$$

  Notice that in this case the factor $\frac{N'}{\sum_{a,a'} M_{a,a'} \frac{N'_a N'_{a'}}{N_{a'}}}$ is the same for $M'_{a,a'}$ and for $M'_{a',a}$.

As a global measure of the reciprocity error ($E \in [0,1]$) we use the total fraction of non-reciprocal contacts:

$$E = \frac{\sum_{a,a'>a} |C_{a,a'} - C_{j,i}|}{\frac{1}{2}\sum_{a,a'} C_{a,a'}} = \frac{\sum_{a,a'>a} |M_{a,a'}N_a - M_{a',a}N_{a'}|}{\frac{1}{2}\sum_{a,a'} M_{a,a'}N_a}. \tag{2.24}$$

It will be zero for methods M1, M2 and M3. The evolution of the reciprocity error $E$ for the 16 countries considered (as their own demographic structure evolves in time) is shown in Figure 2.1 (panel C).

#### 2.2.5.2 Intrinsic Connectivity

Let us define the matrix $H_{a,a'}$ as the contact pattern resulting after assuming homogeneous mixing (also known as proportional mixing) embedded in a population structure $N_a$ (normalized to have a connectivity of 1):

$$H_{a,a'} = \frac{N_{a'}}{N}, \tag{2.25}$$

where $N = \sum_i N_a$.

Thus, we can define the intrinsic connectivity $\Gamma_{a,a'}$ as the ratio between the contact rate and what we would have in the case of homogeneous mixing ($\Gamma_{a,a'} = M_{a,a'}/H_{a,a'} = M_{a,a'}\frac{N}{N_{a'}}$). The entries of this matrix will be larger than 1 if the connectivity between age-groups $a$ and $a'$ is more than what would be expected under the homogeneous mixing scenario, and viceversa.

We study the evolution of this matrix (i.e., $\frac{\Gamma'_{a,a'}}{\Gamma_{a,a'}} = \frac{M'_{a,a'}}{M_{a,a'}} \frac{N_{a'}}{N'_{a'}} \frac{N'}{N}$) with the four methods:

- Intrinsic Connectivity with M0:

$$\frac{\Gamma'_{a,a'}}{\Gamma_{a,a'}} = \frac{M'_{a,a'}}{M_{a,a'}} \frac{N_{a'}}{N'_{a'}} \frac{N'}{N} = \frac{N_{a'}}{N'_{a'}} \frac{N'}{N}. \tag{2.26}$$

  It is not constant.

- Intrinsic Connectivity with M1:

$$\begin{aligned} \frac{\Gamma'_{a,a'}}{\Gamma_{a,a'}} &= \frac{M'_{a,a'}}{M_{a,a'}} \frac{N_{a'}}{N'_{a'}} \frac{N'}{N} = \frac{1}{2}\left(1 + \frac{N_a N'_{a'}}{N_{a'} N'_a}\right) \frac{N_{a'}}{N'_{a'}} \frac{N'}{N} \\ &= \frac{1}{2}\left(\frac{N_{a'}}{N'_{a'}} + \frac{N_a}{N'_a}\right) \frac{N'}{N}. \end{aligned} \tag{2.27}$$

  It is not constant.

- Intrinsic Connectivity with M2:
  It is constant by definition:

$$M_{a,a'} = \Gamma_{a,a'} \frac{N_{a'}}{N} = \Gamma_{a,a'} H_{a,a'} \Longrightarrow \Gamma_{a,a'} = \frac{M_{a,a'}}{H_{a,a'}} \text{ is constant.} \tag{2.28}$$

- Intrinsic Connectivity with M3:

$$\frac{\Gamma'_{a,a'}}{\Gamma_{a,a'}} = \frac{M'_{a,a'}}{M_{a,a'}} \frac{N_{a'}}{N'_{a'}} \frac{N'}{N} = \frac{N'_{a'}}{N_{a'}} \frac{N_{a'}}{N'_{a'}} \frac{N'}{\sum_{a,a'} M_{a,a'} \frac{N'_a N'_{a'}}{N_{a'}}} \frac{N'}{N} = \frac{N'}{\sum_{a,a'} M_{a,a'} \frac{N'_a N'_{a'}}{N_{a'}}} \frac{N'}{N}. \tag{2.29}$$

  It only changes in a global factor $\frac{N'}{\sum_{a,a'} M_{a,a'} \frac{N'_a N'_{a'}}{N_{a'}}} \frac{N'}{N}$.

Only M2 (and M3 except for a global factor) preserves the intrinsic connectivity. M0 and M1 however, change the tendency of mixing between different age-groups. In order to understand the importance of this, let us suppose that we have an initial matrix that follows the assumption of homogeneity (i.e., $M_{a,a'} = H_{a,a'} \Longrightarrow \Gamma_{a,a'} = 1$). Then, we would have that from an initial situation of homogeneous mixing we end up with heterogeneities artificially produced by the change on the shape of the demographic pyramid if we follow methods M0 or M1, while M2 will still provide homogeneous mixing regardless of the demographic structure. The error in Intrinsic Connectivity is shown for Poland and Zimbabwe at different years in Figure 2.1 (panels D and E).

Also notice that there exists a relation between $\Gamma_{a,a'}$ and the matrix $F_{a,a'} = \frac{M_{a,a'}}{N_{a'}}$ that we described in equation 2.2. The method M2 can also be seen as the one that preserve the density scale applying a correction for the total population size $F'_{a,a'} = F_{a,a'}\frac{N}{N'}$.

#### 2.2.5.3 Mean connectivity

The mean connectivity $\langle k \rangle$, i.e, the average number of contacts per individual, is given by:

$$\langle k \rangle = \frac{\sum\limits_{a,a'} M_{a,a'} N_a}{N} \tag{2.30}$$

We have normalized the empirical contact matrices to have a mean connectivity of 1. When transferring these contact matrices to different demographies we have that the mean connectivity might change depending on the method:

- Mean connectivity in M0:

$$\langle k' \rangle = \frac{\sum\limits_{a,a'} M_{a,a'} N'_a}{N'}. \tag{2.31}$$

- Mean connectivity in M1:

$$\langle k' \rangle = \frac{\sum\limits_{a,a'} M_{a,a'} \frac{1}{2}\left(1 + \frac{N_a N'_{a'}}{N_{a'} N'_a}\right) N'_a}{N'} = \frac{\sum\limits_{a,a'} M_{a,a'} N'_a}{N'}. \tag{2.32}$$

  As M1 consists of a re-arrengment of the contacts that already exist in M0 to correct reciprocity, it can be shown that they actually share the same mean connectivity.

- Mean connectivity in M2:

$$\langle k' \rangle = \frac{N}{(N')^2} \sum_{a,a'} M_{a,a'} \frac{N'_{a'} N'_a}{N_{a'}}. \tag{2.33}$$

- Mean connectivity in M3:

$$\langle k' \rangle = \frac{\sum\limits_{a,a'} M_{a,a'} \frac{N'_{a'}}{N_{a'}} \frac{N'}{\sum\limits_{a,a'} M_{a,a'} \frac{N'_a N'_{a'}}{N_{a'}}}}{N'} = \frac{\sum\limits_{a,a'} M_{a,a'} \frac{N'_a N'_{a'}}{N_{a'}}}{\sum\limits_{a,a'} M_{a,a'} \frac{N'_a N'_{a'}}{N_{a'}}} = 1. \tag{2.34}$$

M3 consists of a normalization of M2, so it trivially preserves the mean connectivity.

## 2.3 Results

### 2.3.1 Reciprocity error

In order to study the error incurred when using M0, we transform the contact matrices obtained from empirical studies in different countries to new matrices that correspond to the same location but at different years (that could be past records or future projections). As the population changes over time, the new matrices incorporate the population demographies of the same countries across time. We define the reciprocity error as the coefficient of variation of the number of contacts measured in both directions, which gives us a matrix that we will call non-reciprocity matrix ($NR_{a,a'}$):

$$NR_{a,a'} = \frac{M_{a,a'}N_a - M_{a',a}N_{a'}}{\frac{1}{2}\left(M_{a,a'}N_a + M_{a',a}N_{a'}\right)} \tag{2.35}$$

It is an antisymmetric matrix, in which a positive value of the entry $(a, a')$ means that there are more contacts from $a$ to $a'$ than in the opposite direction, and viceversa. A value of 0 would mean that the contacts between $a$ and $a'$ are well balanced.

In Figure 2.1 we represent the demographic structures of Poland (panel A) and Zimbabwe (panel B) for different years alongside the corresponding non-reciprocity matrices. In the case of European countries (Poland in panel A as an example), demographic structures have suffered from an aging process during the last decades, which is predicted to continue in the future. This aging tends to provoke negative values under the diagonal for the matrices $NR_{a,a'}$ when we assumed past demographic structures, while the opposite will occur in the future. The behaviour for African countries (Zimbabwe in panel B) is slightly different, as their demographies have been more stable for the last decades, and only now they are beginning to age faster. In brief, when we use directly a contact pattern in a demographic structure that is younger than when it was measured, it will lead to an overestimation of the contact rate of (and the force of infection corresponding to) the youngest age-groups. The opposite will occur when we use contact patterns in an older population.

In figure 2.1C we represent the evolution of the proportion of non-reciprocal contacts for all 16 countries considered (as defined in equation 2.24). This

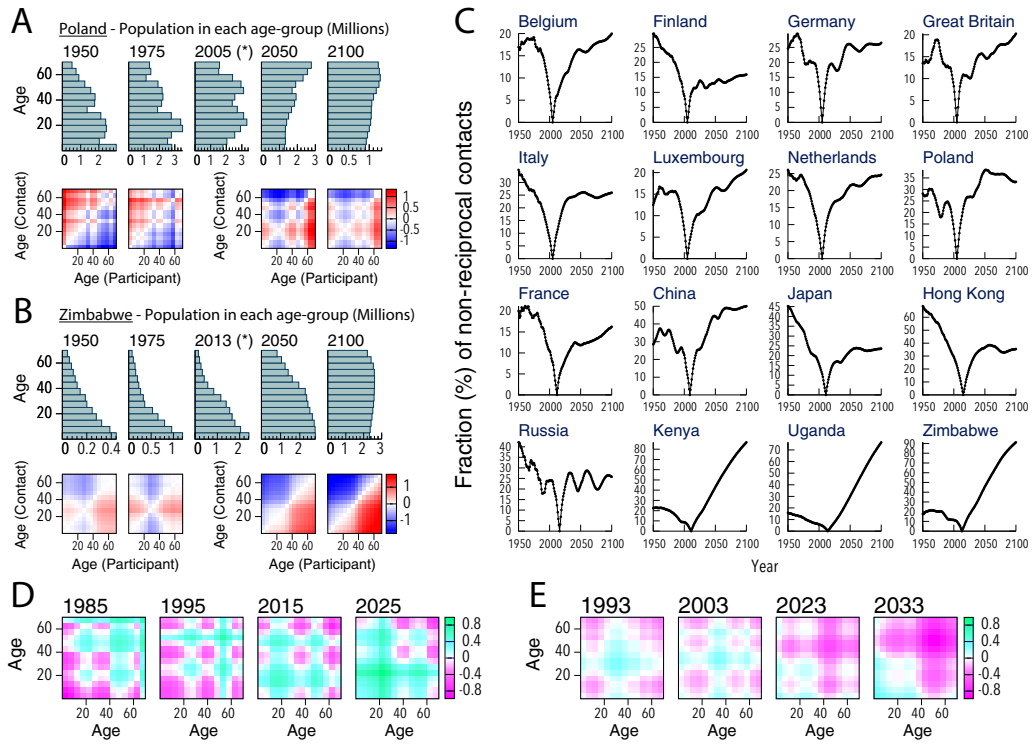


Figure 2.1: **Analysis of methods M0 and M1.** A-B: Demographic structures for different years and the respective non-reciprocal matrices $NR'_{a,a'}$ for Poland and Zimbabwe respective using M0. C: Evolution of the total fraction of non-reciprocal contacts for M0 in the 16 countries analyzed in this study. D-E: $\log 2(\frac{\Gamma'_{a,a'}}{\Gamma_{a,a'}})$ for Poland and Zimbabwe respectively, in four different years (10/20 years before/after the measurement of the contact patterns) for M1. The original data corresponds to 2005 for Poland and 2013 for Zimbabwe.

magnitude is equal to zero in the year when the contact matrix was measured, as we have applied a correction for the empirical matrices to fulfill reciprocity at the reference setting. However, it dramatically increases as we move far from the year of the survey. In the examples shown here, only two years before/after the survey time, the fraction of non-reciprocal contacts already reaches 5%. Note that methods M1, M2 and M3 are well balanced by construction, thus $NR_{a,a'} = 0$ for every $(a, a')$ when using any of them.

### 2.3.2 Intrinsic Connectivity error

We next study the evolution of the ratio between the age-dependent contact rates and an homogeneous mixing scenario. This ratio gives us the

matrix $\Gamma_{a,a'}$, defined as the intrinsic connectivity in equation 2.10. As explained in section 2.2.5.2, the entries of $\Gamma_{a,a'}$ are bigger than 1 when the interactions between age-groups $a$ and $a'$ surpasses what it is expected from the case of homogeneous mixing, and smaller than 1 in the opposite case.

In Figure 2.1D and 2.1E we show 4 snapshots of the ratio of the intrinsic connectivity and the original survey ($\Gamma'_{a,a'}/\Gamma_{a,a'}$) obtained using M1 for Poland and Zimbabwe respectively. Each panel corresponds to an adaptation of the contact matrix to the population demography of the countries 10 and 20 years before and after the survey (i.e., the 4 matrices correspond to $t = \tau - 20y$, $t = \tau - 10y$, $t = \tau + 10y$ and $t = \tau + 20y$, where $\tau$ corresponds to the time when the survey that determined the contact patterns took place). We can see that, even if M1 corrects the appearance of non-reciprocity, this method changes the tendency of some age-groups to mix with respect to others. Specifically, we can see that M1 will over-represent contacts between young individuals (and under-represent contacts between old individuals) as the population gets older.

Furthermore, the previous results are quantitatively important. For instance, if we were to use the contact matrices that we have from Poland (measured in 2005) today (2018), we would have that the ratio $\Gamma'_{a,a'}/\Gamma_{a,a'}$ surpasses 1.5 for some specific age-group pairs, while it goes down to almost 0.5 in others, or, in other words, the usage of M1, which does not take into account the changes in the fractions of individuals in each age-strata that occurred between 2005 and 2018, causes a bias of more than 50% in the contact densities projected between certain age groups. Consequently we say that M1 does not preserve intrinsic connectivity. The density correction (M2) avoids this problem, as it explicitly considers a fixed intrinsic connectivity matrix ($\Gamma_{a,a'}$ as defined in section 2.2.4) that is modified according to the density of each age-group (see equation 2.9).

### 2.3.3 Evolution of mean connectivity

In Figure 2.2A-B we represent the contact patterns obtained with M2 and M3 for Poland and Zimbabwe, respectively, in different years. We see how, specially in the case for Zimbabwe, as the population gets older, the values of the matrix below the diagonal (contacts toward young individuals) fade in favor of contacts toward older individuals as those age-groups gain more representation.

As for the mean connectivity (Figure 2.2C), considering the evolution of

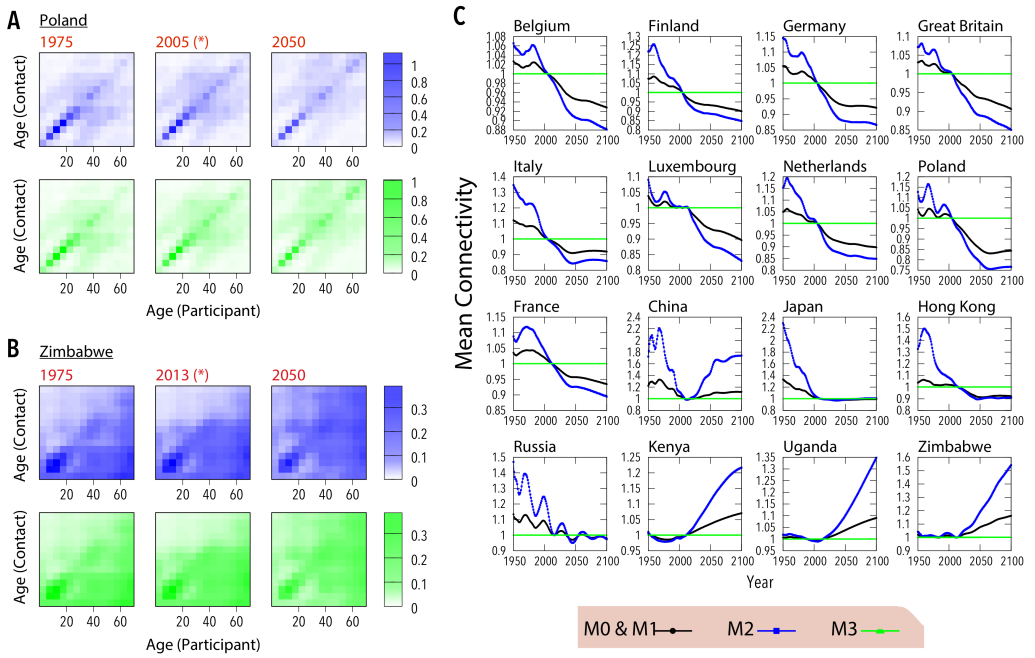


Figure 2.2: **Analysis of methods M2 and M3.** A-B: Contact patterns $M_{a,a'}(t)$ for five different years with methods M2 (blue) and M3 (green) for Poland and Zimbabwe, respectively. C: Evolution of Mean Connectivity for M2 (blue), M3 (green) and M0 and M1 (black, both methods give the same mean connectivity).

contact patterns in M2 or considering them constant (M0) leads to the same qualitatively behaviour, although variances are higher with M2. These trends are decreasing in Europe and increasing in Africa. M0 and M1 have the same mean connectivity, as M1 consists basically of a rewiring of those connections that exist in M0 in order to correct for reciprocity. M3 is a normalization of M2 so the connectivity is constant in this case.

### 2.3.4 Overview of different methods

We have shown up to four different methods of use heterogeneous contact patterns when demography evolves in time (being the first one of them to simply use them without any further consideration regarding the demographic structure). In table 2.2 we summarize their properties.

| Method | Reciprocity? | Preserves Intrinsic Connectivity? | Constant average connectivity? |
|---|---|---|---|
| M0: Unadapted contact patterns | No | No | No |
| M1: Pair-wise correction | Yes | No | No |
| M2: Density correction | Yes | Yes | No |
| M3: Density correction + Normalization | Yes | Yes (with a global factor) | Yes |

Table 2.2: **Overview of different methods.** Summary of the different methods to deal with contact patterns and their properties.

### 2.3.5 Geographical Comparisons

The intrinsic connectivity matrices $\Gamma_{a,a'}$ that we obtain for every country allow us to compare the contact patterns of different settings once the influence of demography has been accounted for, and removed. In Figure 2.3A we represent these matrices for the 16 countries analyzed in this work. Just by visual inspection we can identify some distinctive features: European matrices are more assortative and present higher interaction intensities among young individuals than African ones. To formalize this observation, in Figure 2.3B, we place the different matrices in a two dimensional plot defined by the proportion of overall connectivity produced by young individuals ($Y$) and the assortativity coefficient ($r$)[143] as defined by:

$$Y = \frac{\sum\limits_{a<20y.,a'} \Gamma_{a,a'}}{\sum\limits_{a,a'} \Gamma_{a,a'}} = \frac{\sum\limits_{a<20y.} k(a)}{\sum\limits_{a} k(a)}, \tag{2.36}$$

$$r = \frac{\sum\limits_{a,a'} \left( \Gamma_{a,a'} - \frac{k(a)k(a')}{\sum\limits_{a,a'} \Gamma_{a,a'}} \right) E(a)E(a')}{\sum\limits_{a,a'} \left( k(a)\delta_{a,a'} - \frac{k(a)k(a')}{\sum\limits_{a,a'} \Gamma_{a,a'}} \right) E(a)E(a')}, \tag{2.37}$$

African and European countries cluster around different values of these two magnitudes: specifically, in African countries we found less assortativity and the contacts are less dominated by young individuals than in the European countries. As for the Asia region we see that Japan and China have significantly higher assortativity and fraction of contacts among young individuals than either African or European countries. In turn, Hong Kong, with its particular geographic idiosyncrasy –a small country, predominantly urban, with one of the highest population densities in the world–, presents an intrinsic connectivity matrix that is more similar to one from a European country than from China or Japan.

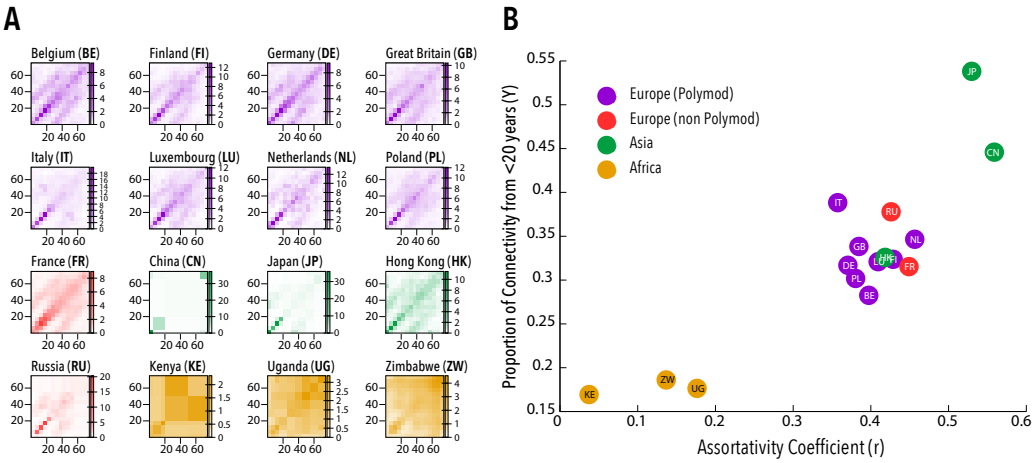


Figure 2.3: **Geographical comparison of empirical contact matrices.** A: $\Gamma_{a,a'}$ matrices for the 16 countries considered in this work. B: Proportion of the overall connectivity that comes from individual with less than 20 years ($Y$) vs the assortativity coefficient ($r$) for the 16 countries.

### 2.3.6 Short cycle SEIR dynamics

Up to now, we have shown that there are several ways to deal with demographic change and evolving populations regarding the structure of the contact patterns for a given population. We next address how these different methods impact disease modeling. To this end, we implement a Short cycle SEIR model. We choose a short-cycle disease so that we can assume that the population structure is constant during each simulation. To parameterize the model we will use the values of an influenza outbreak that took place in Belgium in the season 2008/2009.[113]

The disease is described using a SEIR model. Susceptible individuals can catch the disease with a transmissibility rate $\beta$ per-contact with an infective individual. Once infected, individuals remain on a latency state for $\epsilon^{-1} = 1.1$ days on average. Then, they become infectious for $\mu^{-1} = 3$ days on average, period when they can transmit the infection to susceptible individuals. After that, they recover and become immune to the disease. We use a discrete and stochastic model where the population is divided into 15 age classes. Social mixing is quantified by the contact matrices extracted from the data, normalized accordingly. Time is discretized with a time step of $\Delta t = 1$ day and the results are averaged over $10^4$ runs.

Finally, we consider the reproductive number to be $R = 2.12$, the same at all times and countries. With this value, we obtain the appropiate $\beta$ for each country and each year so that the largest eigenvalue of the matrix

$$K_{a,a'} = \frac{\beta}{\mu} M_{a,a'} \frac{N_a}{N_{a'}} \tag{2.38}$$

in the case of the methods M0, M1 and M3 and

$$K_{a,a'} = \frac{\beta}{\mu} \Gamma_{a,a'} \frac{I_a}{N} \tag{2.39}$$

in the case of the method M2 is equal to the reproductive number, $R$. Note that as we are adjusting $\beta$ in each case to obtain the desired value of the reproductive number, the mean connectivity of the matrix will not play a role. Thus, we would obtain exactly the same results using M2 and M3, as their only difference is the global contact rate and it will be incorporated in $\beta$.

These simulations allow us to study a situation where a short-cycle, influenza-like pathogen appears in a given location, at different possible times, associated to the same reproductive numbers. Under this hypothetical scenario, we would like to know how different would be the forecasted size of the epidemic as a result of considering different contact matrices coming from the different projection methods proposed in this work. In particular, this scenario is instrumental to distinguish the outcomes from models M0, M1 and M2. However, the requirement of the outbreaks to have the same reproductive numbers implies the assumption that the infectiousness $\beta$ can be estimated independently in each event. As a consequence, since the matrices derived from M2 and M3 only differ by a global scaling factor, this operation absorb the differences between M2 and M3, making them indistinguisable.

The results of this exercise are presented in Figure 2.4. In Figure 2.4A we can see that, while methods M0 and M1 predict lower age-aggregated incidences in European countries in 2050 with respect to 2000, M2 reduces these differences and the incidences are comparable for both years or even positive. A different situation occurs in Africa, where M0 and M1 predict an increase in incidence in the future while using M2 would lead to a decrease, though differences remain small (less than 5% of variation).

In panel 2.4B we represent, for two examples of Europe and Africa (Poland in blue and Zimbabwe in orange), the temporal evolution of the incidence observed with the different methods. Furthermore, we represent the age-specific incidence for both countries in three different years: 2010, 2030 and

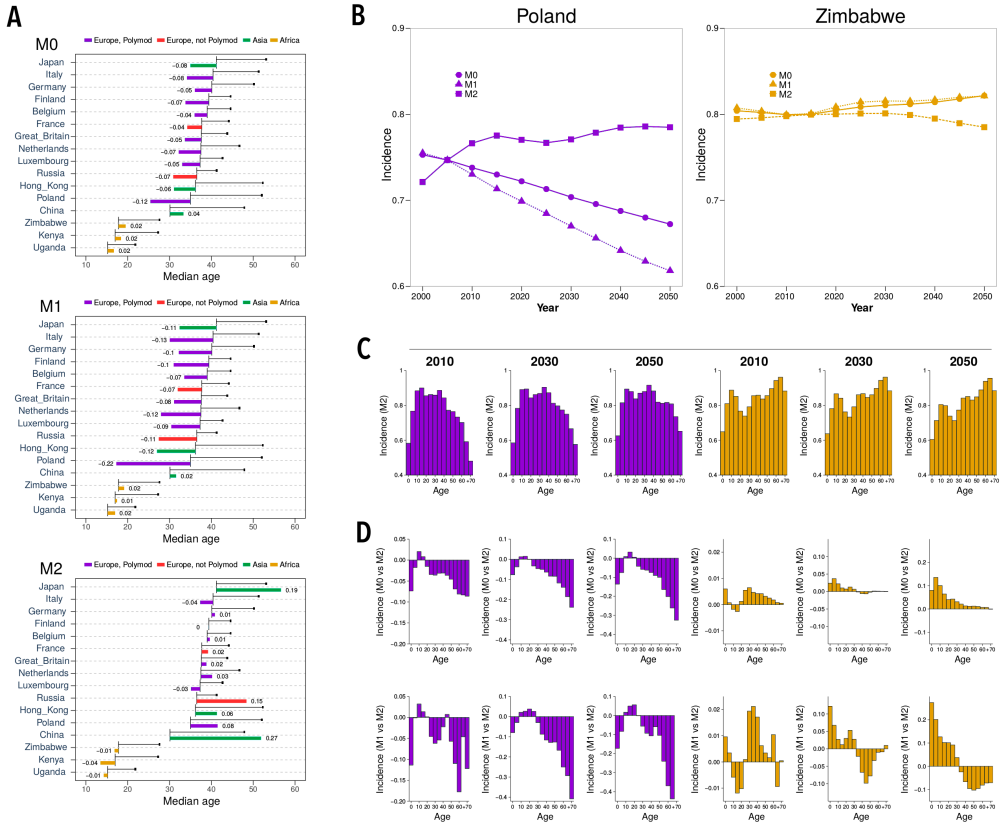


Figure 2.4: **SEIR dynamics with different methods.** A: Median age at 2000 and 2050 (black line, beginning with the value at 2000 and ending with a bullet point with the value at 2050) for the 16 countries considered and relative variation in incidence over the same period (colored bars), for M0, M1 and M2. B: Incidence (over all ages) vs Year for Poland (blue) and Zimbabwe (orange) using M0, M1 and M2. C: Incidence by age group for Poland and Zimbabwe in 2010, 2030 and 2050 using M2. D: Relative differences of the incidence by age group of M0 and M1 with respect to M2 ($\frac{Inc(M0)-Inc(M2)}{Inc(M2)}$ and $\frac{Inc(M1)-Inc(M2)}{Inc(M2)}$).

2050 (Panel 2.4C). The age-distribution of the incidence evidences the differences in connectivity patterns between Poland and Zimbabwe. While the incidence in elderly people drops in Poland (as the contact rates for those age-groups also drop), it remains high in Zimbabwe for the same age-groups.

The different methods of implementing contact rates also affect the age-specific incidence. In panel 2.4D we represent the relative variation in age-

specific incidence obtained with methods M0 and M1 with respect to M2 for Poland and Zimbabwe. In Poland we see that M0 and M1 tend to underestimate the incidence specially among the elder age-groups. In Zimbabwe M0 tends to overestimate the incidence among young individuals, while with M1 we encounter both effects: and overestimation among the youngest and a underrepresentation among the eldest.

The reshaping of the age-specific incidence between models is coherent with the changes in topology already studied. For the case of M0, i.e., maintaining the contact patterns constant in time, we have that in the future, as the demographic structure shifts to older populations, contacts toward children will be overrepresented and contacts toward adults will be underrepresented. At first order we can obviate the contacts that are far from the diagonal, and establish that M0 mainly underrepresents contacts between adults and overrepresents contacts between young individuals (in the context of aging populations). Thus, we will obtain an underrepresentation of the incidence in adults, and the opposite in children. However, as the eldest age-groups increase their population in Europe, they dominate the dynamics and cause and underestimation of the global incidence that eventually affects all age-groups. In African countries, where the contact patterns are less assortative than European countries, this effect is smaller. Besides, as African populations are still young even in 2050, the overestimation of young contacts dominates the dynamics, and the differences in incidence are mainly positive. The situation is similar for M1. As represented in Figure 2.1D-E, for M1 we also have an underrepresentation of contacts between adults and an overestimation between young individuals, yielding to similar results than M0.

Altogether, these results illustrate how an ill adaptation of the contact patterns observed in the past in a given country to a later time point can translate into epidemiological forecasts that are highly biased. Regarding the dynamic equivalence of methods M2 and M3, we have to emphasize that it emanates only from the assumption that reproductive numbers can be measured at the early stages of any of the epidemics being simulated in each year, which is a conservative -often optimistic- assumption. Alternatively, we could think of an scenario where the reproductive number of a given pathogen was estimated in a given year, and that information used to infer the probability of transmission per contact (the infectiousness $\beta$) of the pathogen, with the aim of producing a-priori forecasts for posterior re-apparences of the same pathogen. In such alternative scenario, the usage of different contact matrices projections would be even more relevant, for it would impact directly

the reproductive number of the forecasted outbreaks, now characterized by a common $\beta$. In such an scenario, (which is conceptually similar to the task of producing long term forecasts of persistent diseases,[129] based on epidemiological parameters calibrated on an initial time-window, as we will do in the next chapter), the dynamic equivalence of models M2 and M3 is broken, since running M2 or M3 with the same infectiousness parameter and different average connectivities yields different reproductive numbers.

### 2.3.7 Comparison with Prem et al.

In a recent work by Prem et al.,[142] the contact pattern from the Polymod project[25] (averaged over the 8 European countries) is exported to different countries. However, the approach proposed in Prem et al.[142] to complete that task implies the integration of a series of data of disparate nature: household structures, pupil-to-teacher ratios, school enrollment rates, and also age distributions. In this sense, although this work also allows to perform a similar analysis and project an averaged Polymod matrix (for example) to other non-Polymod countries, in this case, this is done using demographic data on populations' age distributions alone without including further data. As a means to evaluate the accuracy of the more economic approach introduced here compared to that proposed in Prem et al.[142], we have focused on the eight non-Polymod countries for which we have independent empiric data (France, Russia, China, Hong Kong, Japan, Kenya, Uganda and Zimbabwe), and compared, for each of them, three different contact matrices: A) the empiric ones, B) the projections obtained starting from the Polymod data, and applying the transformations proposed in Prem et al.[142] and C) an equivalent projection of the Polymod pattern using either M2 or M3 as proposed in this work.

In order to obtain the projections based on this method, we first need to produce an average of the Polymod contact matrices. To do so, we first select the Polymod matrices, correct for reciprocity in each of them (assuming demographies of the year 2016) and normalize. Then we build an average Polymod matrix as:

$$M_{a,a'}^{(Polymod)} = \frac{\sum_c M_{a,a'}^{(c)} n_a^{(c)}}{\sum_c n_a^{(c)}} \tag{2.40}$$

where the superindex $c$ indicates the specific country of the Polymod study, and $n_a^{(c)}$ is the number of participants of age $a$ in country $c$. Then, we

extract $\Gamma_{a,a'}^{(Polymod)}$ assuming as demography the sum over all Polymod countries. Using this matrix, we obtain projections for the other non-Polymod countries (using either M2 or M3, which are equivalent for the sake of these analyses).

Once we obtained the trans-national projections of the Polymod data, we can compare them against the matrices proposed in Prem et al. [142] as well as against the empirical data reported in the different surveys conducted in the eight countries listed in table 2.3.

| Conuntry | This work vs. Prem et al. | Prem et al. vs. Surveys | This work vs. Surveys |
|---|---|---|---|
| France | 0.932 | 0.827 | 0.849 |
| Russia | 0.878 | 0.800 | 0.909 |
| China | 0.962 | 0.235 | 0.315 |
| Japan | 0.933 | 0.675 | 0.770 |
| Hong Kong | 0.909 | 0.661 | 0.856 |
| Kenya | 0.971 | 0.566 | 0.563 |
| Uganda | 0.948 | 0.649 | 0.670 |
| Zimbabwe | 0.964 | 0.573 | 0.609 |

Table 2.3: **Comparison with the method proposed by Prem et al.** Correlation index of the comparison of different contact patterns. For the 8 non-Polymod countries considered in this work, we make three comparisons: (1) projections of the Polymod matrix based on this work vs. the projections made by Prem et al., (2) the projections made by Prem et al. vs the empirical data measured in those countries and (3) projection based on this work vs the same empirical data.

Importantly, the projections based on this method, which are based on the age-distributions of the countries analyzed alone, correlate very strongly with the projections proposed in Prem et al. [142] ($r>0.9$ everywhere but in Russia), even though these precise a considerably larger amount of detailed data to be built. Even more surprisingly, we see how our projections correlate slightly better to the actual contact patterns observed in empirical studies conducted in these different countries, pointing to the fact that, among all sources of heterogeneity in social mixing, age is arguably more important than other considerations that seemingly play secondary roles. However, correlations between projections and empiric data are not always consistent, but, perhaps unsurprisingly, only high enough when Polymod data is projected into other European settings (Russia and France). This warns once again against the indiscriminate extrapolation of contact structures measured in specific

geographical areas to settings belonging to other continents or characterized by profound cultural or demographic differences.

## 2.4 Conclussion

Summarizing, empirical contact patterns belong to a specific time and place. If we want to integrate the heterogeneity of social mixing into more realistic models, we need to address how (and in what cases) to export contact patterns from empirical studies to the populations we want to study. In this work, we have studied and quantified the significant bias incurred when a specific contact pattern is blindly extrapolated to the future (or the past), even if we remained inside the same country where those contacts were measured. In fact, only a couple of years after the measurement of these contact patterns, the changes in the age structure of the population make them vary significantly. Thus, for any meaningful epidemic forecast based on a model containing age-mixing contact matrices, we would need to adapt them taking into account the evolution of the demographic structures. Moreover, as we have shown, even for cases that do not expand into long periods of time and a constant demography could be assumed, it is necessary to make an initial adaptation of whatever empirical contact structure we want to implement, into the specific demographic structure of our system. We have also seen how these relevant differences in the topology of contacts yield to significant consequences for the spreading of a disease. Applying different methods to deal with contact patterns leads to important differences not only in the global incidence for a SEIR model, but also on age-specific incidences. Having such an important impact for the spreading of a disease, the insights provided by this work should be taken into consideration by modelers and also by public health decision-makers.

In a similar way, we have explored the differences between the contact patterns of different countries. Thus, we have found the existence of some specific characteristics beyond the underlying demographic pyramid, which warns against exporting contact patterns across different geographic areas (i.e. continents). As there exists different intrinsic connectivity patterns (i.e., once demography effects have been subtracted) between countries, it is also likely that there exists a time-evolution of the intrinsic connectivity inside the same setting. Although it is impossible to predict how society will change in the future, we should always take this into account as a limitation in any forecast for which the heterogeneity in social mixing is a key element.

Finally, we note that there are some limitations that could affect quantitatively the results shown in this work. First of all, we have derived the contact patterns of the different studies according to the demographic structures of the specific country for the year the survey took place. Thus, we are implicitly assuming that the setting where the different surveys were performed are comparable with the national data in terms of their demographic pyramids. In other words, we assume that the surveys are representative of the population at large. This is likely true for most of the countries analyzed, but there are certain cases in which this might not be the case, either because of small study size or putatively biased recruitment of participants. Besides, as we have already discussed in section 2.2.3, the different granularity (i.e., definition of the age-groups) used throughout the bibliography studied also imposes some limitations when comparing the data. It is also worth pointing out that, although in this work we have focused on age-structured systems (which has had its relevance in recent history of epidemiology), the problem studied here can be extrapolated to other models that might categorize their individuals based on other different traits that determine their social behavior.

The results reported here and their implications open several paths for future research. One is related to the social mixing patterns themselves. In order to predict the large-scale spreading of a disease, multiple scales need to be integrated and coupled together. This implies that when integrating different spatial scales, we need to deal with different contact matrices and local demographies. For instance, in developed countries, it is known that the structure of the population is not the same in the most central or most populated cities as compared to smaller ones or the countryside. Thus, nation-wide demographies and surveys to infer contact matrices might need to be disaggregated. What is the right spatial scale to measure both quantities is an interesting and unsolved question. In this sense, here we have limited the simulated disease scenario to the case of isolated populations −a country−, but it remains to be seen what are the effects over a metapopulation framework, in which we have mobility between subpopulations of potentially very different demographic structures.

The insights provided by this work will be essential for the development of a TB spreading model with heterogenous age-dependent contact patterns and an evolving demographic structure, as related in the next chapter. We have already seen the biases that different methods of projecting contact matrices have over a simple SEIR model. Moving to a model with a more

accused age dependency and a continuously evolving demography can be a risky action if the warnings that we have given in this chapter go unnoticed.

# Chapter 3

# Data-driven model for the assessment of Tuberculosis transmission

## 3.1 Introduction

Throughout the next chapter we will develop a model for the assessment of TB transmission. As the ultimate goal is the evaluation and comparison of age-focused interventions, such as newborns or adolescents-focused vaccination campaigns, that could help policy making, we are specially interested in achieving a proper description of the multiple ways whereby TB dynamics couples with populations' age structure, a limitation that has been pointed out as one of the most critical.[67,144] For example, patients' age is strongly correlated to the type of disease that they tend to develop more often, as well as to the probability of developing active TB immediately after infection (usually called "fast progression"[55]). This way, while a larger fraction of children younger than 15 years of age develop non-infectious forms of extrapulmonary TB with respect to adults (25% vs 10%[55,93,145]), the risk of fast progression is larger in infants (50% in the first year of life), then decays (20-30% for ages 1-2; 5% for 2-5 and 2% for 5-10), and raises up again in adults (10-20% for individuals older than 10).[53] Additionally, transmission routes of TB, being this a paradigmatic air-borne disease, are expected to show significant variations in intensity across age.[25,146] The empirical characterization of these contact structures constitutes an intense focus of research in data-driven epidemiology of air-borne diseases,[123] and their influence on the transmission dynamics of diseases like influenza has been recently explored with relevant implications.[147,148]

Thus, if subjects' age modifies the disease-associated risks at the level of single individuals, it is likely to expect that changes in the demographic age-structure at the population level will impact TB burden projections. This is mainly due to the slow dynamics that is characteristic to TB, which forces modelers to describe the evolution of the disease during long periods of time, typically spanning several decades. These time-scales are rather incompatible with the assumption of constant demographic structures, at least nowadays, since world-wide human populations are presumed to age from the current median of 30 years old to 37 in 2050.[138] And yet, achieving a sensible description of TB transmission able to capture the effects of time-evolving demographic structures remains an elusive goal in TB modeling. Demographic dynamics are traditionally neglected in TB transmission models, the same way that contact structures are assumed to be homogeneous across age-groups.[55,101,108]

In this model we incorporate empirical data on demographic dynamics and contact patterns around classical formulations of TB spreading models, thus unlocking less biased descriptions of the spreading dynamics of the disease. To this end, we present a TB spreading model (Figure 3.1A) whereby we provide a data-driven description of TB transmission that presents two main novel ingredients with respect to previous approaches. First, the model incorporates demographic forecasts by the United Nations (UN) population division[138] (Figure 3.1B) in order to describe the coupling between demographic evolution and TB dynamics. Secondly, the model integrates region-wise empirical data about age-dependent mixing patterns adapted from survey-based studies conducted in Africa and Asia[130,133,134,136,137] (Figure 3.1C), following the methodology developed in the previous chapter of this thesis, instead of assuming that all the individuals in a population interact homogeneously, as traditionally considered in the literature.[55,101,108]

Upon model calibration in some of the countries most affected by the disease in 2015 and subsequent simulation of TB transmission dynamics up to 2050 (Figure 3.1D), we scrutinize the implications derived from integrating these pieces of empirical data within the model, and discuss their impact on the forecasts produced, both at the level of aggregated incidence and mortality rates and on their distributions across age-strata. Specifically, we quantify the effects of populations' aging on prospected incidence rates until 2050, as well as the impact on the age distribution of the disease burden that emanate from introducing empiric contact data in the models. Furthermore, we quantify the sensitivity of these effects to the different model inputs, and

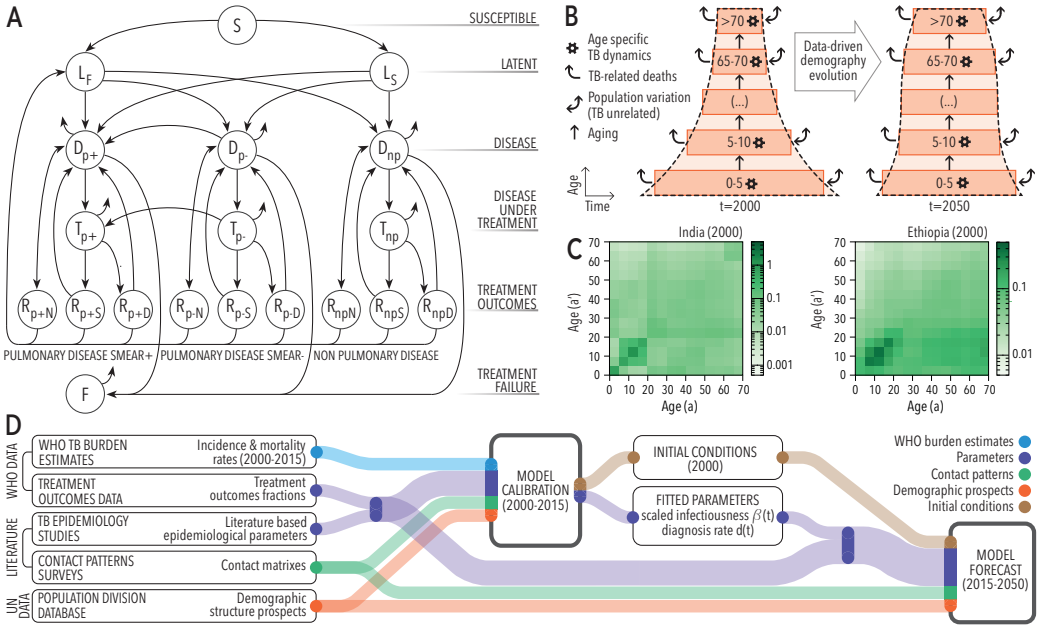


Figure 3.1: **Model description.** (A): Natural History scheme of the TB spreading model. Description of the different classes is given through section 3.2.1 (B): Scheme of the coupling between TB dynamics and demographic evolution. The transmission model summarized in panel A describes the evolution of the disease in each age group, including the removal of individuals due to TB-mortality (curved arrows). The evolution of the total volume of each age-strata is corrected (bi-directional arrows: TB unrelated population variations) to make the demographic pyramid evolve according UN prospects. (C): Empirical contact patterns used for African and Asian countries. (D): Data flow scheme. Epidemiological parameters, contact matrices, and demographic prospects are used to calibrate the model, with the goal of reproducing observed TB incidence and mortality trends during the period 2000-2015. As a result of model calibration, scaled infectiousness, diagnosis rates and initial conditions of the system in 2000 are inferred. These elements are then used (along with epidemiological parameters, contacts and demographic data) to extend model forecasts up to 2050.

assess their statistical significance and robustness under a series of alternative modeling scenarios. To do so, alongside the construction of a full model that incorporates demographic evolution and empiric contact patterns, we will also build two reduced models that will serve as baseline comparison: reduced model 1 (RM1) that presents constant demographic structure (but has empiric contact patterns) and reduced model 2 (RM2) that assumes

homogeneity of contacts (but includes demographic evolution).

## 3.2 Model

### 3.2.1 Natural history of the disease

The model of Tuberculosis (TB) spreading is essentially based on previous models by C. Dye and colleagues,[55,108] on which new ingredients –heterogeneous contact patterns[25] and demographical evolution– have been incorporated. The natural history scheme has also been refined so as to render it more suited to the definitions by the World Health Organization (WHO), mostly in what regards to treatment outcomes.

Summarizing, we deal with an ordinary differential-equations based, age structured model of TB in which we consider a class of unexposed individuals –susceptible–, two different latency paths to disease –fast and slow– and six different kinds of disease, depending on its aetiology: -non pulmonary, pulmonary smear positive and pulmonary smear negative–, and depending on whether it is untreated or treated. After the disease phase, we explicitly consider the main treatment outcomes contemplated by the WHO data schemes: treatment completion (or success), default, failure and death.[71,149] We also consider natural recovery without treatment, smear progression (from smear negative to smear positive) and mother-child infection.

The model is structured in 15 age groups, 14 of them covering 5 years of age up to 70 years old, and a last group including all individuals older than 70 years old. When specifying the population of a certain state at a specific age-group and time we will use the notation $X(a,t)$, where $X$ is the concrete state (see list of disease states in table 3.6), $t$ represents the time, and $a \in [0,14]$ is the index representing any of the fifteen age groups.

In the following, we detail the natural history ingredients and transitions between states that we have considered to build up this model; whose natural history is schematized in figure 3.1A.

#### 3.2.1.1 Primary Tuberculosis infection

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

of age group $a$ of being infected, then the total number of susceptible individuals getting infected per unit time will be approximated by $\lambda(a,t)S(a,t)$. We will address the explicit form of $\lambda(a,t)$ in section 3.2.2.

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

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

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

Most of the disease parameters that regulate the fluxes of the Natural History are taken from different bibliographic sources. Most of them are taken directly from the works by Dye et al.[108] and Abu-Raddad et al.[55], that form the base of this work. A list with the values of the different disease parameters can be found in section 3.2.9. However, the values of $p(a)$ used in this model deserve further explanation.

The risk of fast progression to disease after infection is highly variable, and strongly depends on the age of the individual. In the most recent and complete work on this matter, by Marais et al.[53], a complex outlook regarding this parameter is reported, according to which fast-progression risk is higher in newborns, then decreases in children and increases again during adolescence. This pattern, which is largely accepted in the current literature,[150–154] is summarized in Table 3.1.

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

Table 3.1: **Values for $\boldsymbol{p(a)}$.** $p(a)$ values obtained after adapting the values on the risk of developing disease after infection from Marais et al.[53]

The values of Table 3.1 have been obtained from Marais et al.[53] after summing up pulmonary and non-pulmonary disease risks, and propagating their uncertainties assuming independence. We assign these values as the parameters $p(a)$. For the case of $a = 0$, which corresponds to ages from 0 to 5 years, we have three different values. We construct $p(0)$ as a weighted average:

$$p(0) = \alpha_{<1} p_{<1} + \alpha_{1-2} p_{1-2} + \alpha_{2-5} p_{2-5} \tag{3.1}$$

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

#### 3.2.1.2 Progression from latency to (untreated) disease

Either from fast or slow latency, infected individuals can fall sick, progressing to one of the three different active forms of the disease. In the first of these forms, the non-pulmonary disease $D_{np}$, the pathogen can grow in disparate parts of the host body, including the nervous system, bones, kidneys and other organs foreign to lungs. The main characteristic of this kind of TB is that, since the bacilli can not reach the respiratory tract, the individuals are considered, for the purposes of this 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 $\omega_f$ and $\omega_s$ represent the rates at which fast and slow progression occur, and $\rho_{p+}(a)$,$\rho_{p-}(a)$ and $\rho_{np}(a)$ represent the probability to develop each of the three different forms of TB previously described: pulmonary smear-positive, pulmonary smear-negative and non-pulmonary, respectively. We are using the closure relation $\rho_{p+}(a) + \rho_{p-}(a) + \rho_{np}(a) = 1$, so we have only two independent parameters –i.e., $\rho_{p+}(a)$ and $\rho_{np}(a)$–. These three probabilities, as the fast progression probability, are age-dependent –children are known to develop more often non-pulmonary forms of TB–.[155]

#### 3.2.1.3 Tuberculosis related deaths

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

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

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

Finally, individuals who were treated in the past that did not respond to treatment will ultimately die of the disease, at the larger rate $\mu_{p+}$ regardless of their initial type of disease. Class $F$ is detailed later on in this section, in the subsection regarding treatment outcomes.

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

#### 3.2.1.4 TB diagnosis and treatment

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

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

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

positive types of disease.

The estimations for the parameter $\eta$ are based upon the case detection ratios $\chi$ for each type of disease ($D_{p+}$, $D_{p-}$ and $D_{np}$) reported by Abu-Raddad and colleagues.[55] 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.[55], estimations for the case detection ratios are provided for each type of disease and WHO region: $\chi_{p+}$, $\chi_{p-}$ and $\chi_{np}$; 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 $\eta$ for each region:

$$\eta = \frac{\chi_{p-}}{\chi_{p+}} \left( \simeq \frac{\chi_{np}}{\chi_{p+}} \right) \tag{3.2}$$

The errors have been estimated by considering a 15% as the typical uncertainty of both $\chi_{p+}$ and $\chi_{p-}$, as was done in Abu-Raddad et al.[55] for several parameters of the Natural History. We obtain the Confidence Interval for $\eta$ by propagating errors.

In the table 3.2 the values of $\eta$ calculated are listed for the different regions defined by the WHO. We have studied 5 countries from the AFRH region (Nigeria, South Africa, Democratic Republic of the Congo, Ethiopia and Tanzania), 1 from EMR (Pakistan), 4 from SEAR (India, Indonesia, Bangladesh and Myanmar) and 2 from WPR (China and Philippines).

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

Table 3.2: **Values of $\eta$.** Values of $\chi_{p+}$ and $\chi_{p-}$ considered in Abu-Raddad et al.[55] and the values of $\eta$ for each region.

The diagnosis rate is allowed to vary in time, as it has been done in other previous models (see section 3.2.8 for details).

#### 3.2.1.5 Treatment outcomes

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

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

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

Therefore, let us denote as $f_S^{p+}$, $f_D^{p+}$, $f_F^{p+}$ and $f_\mu^{p+}$, the fraction of pulmonary, smear positive TB sick individuals who finish their treatments belonging respectively to success, default, failure and death groups, as they are available in the WHO database.[149] We will have the closure relationship $f_S^{p+} + f_D^{p+} + f_F^{p+} + f_\mu^{p+} = 1$ that allows us to substitute $f_S^{p+} = 1 - (f_D^{p+} + f_F^{p+} + f_\mu^{p+})$ so as to work just with these three fractions of unsuccessful treatment outcomes. For pulmonary smear negative and non pulmonary TB cases, the WHO database does not differentiate the fractions of treatment outcomes,[149] and so we have $f_S^{p-}$, $f_D^{p-}$, $f_F^{p-}$ and $f_\mu^{p-}$ 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 3.9, where confidence intervals correspond to two typical deviations of a multinomial distribution.

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

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

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

- Death during treatment of smear positive TB:
  $\Psi f_\mu^{p+} T_{p+}(a,t)$ individuals/unit time.

- Successful treatment completion of smear positive TB: transition from $T_{p+}(a,t)$ to $R_{p+S}(a,t)$:
  $\Psi(1 - f_D^{p+} - f_F^{p+} - f_\mu^{p+}) T_{p+}(a,t)$ individuals/unit time.

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

- Failed treatment completion of smear negative TB: transition from $T_{p-}(a,t)$ to $F(a,t)$:
  $\Psi f_F^{p-} T_{p-}(a,t)$ individuals/unit time.

- Death during treatment of smear negative TB:
  $\Psi f_\mu^{p-} T_{p-}(a,t)$ individuals/unit time.

- Successful treatment completion of smear negative TB: transition from $T_{p-}(a,t)$ to $R_{p-S}(a,t)$:
  $\Psi(1 - f_D^{p-} - f_F^{p-} - f_\mu^{p-}) T_{p-}(a,t)$ individuals/unit time.

- Early treatment abandon (default) of non pulmonary TB: transition from $T_{np}(a,t)$ to $R_{npD}(a,t)$:
  $\Psi f_D^{p-} T_{np}(a,t)$ individuals/unit time.

- Failed treatment completion of non pulmonary TB: transition from $T_{np}(a,t)$ to $F(a,t)$:
  $\Psi f_F^{p-} T_{np}(a,t)$ individuals/unit time.

- Death during treatment of non pulmonary TB:
  $\Psi f_\mu^{p-} T_{np}(a,t)$ individuals/unit time.

- Successful treatment completion of non pulmonary TB: transition from $T_{np}(a,t)$ to $R_{npS}(a,t)$:
  $\Psi(1-f_D^{p-}-f_F^{p-}-f_\mu^{p-})T_{np}(a,t)$ individuals/unit time.

where the different $R_{xy}$ variables stand for the groups of individuals that have completed their treatment for disease of type $x$ (pulmonary smear positive $p+$ or negative, $p-$, or non-pulmonary $np$) with an outcome denoted by $y$ (Success, $S$, default $D$ and fail $F$). We have also used the subindex $\mu$ 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–.

#### 3.2.1.6 Natural recovery

In certain occasions, natural recovery from TB is possible without medical intervention or treatment.[55] This is modeled by introducing three new classes of naturally recovered individuals in the first branch: $R_{npN}(a,t)$, $R_{p-N}(a,t)$ and $R_{p+N}(a,t)$. Undiagnosed and sick individuals of each type of TB join these new classes after natural recovery as follows:

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

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

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

where $\nu$ is the rate of natural recovery.

#### 3.2.1.7 Endogenous reactivations after treatment or natural recovery

Nonetheless, naturally recovered individuals may experience an endogenous reactivation of the disease, since generally speaking disease recovery does not suppose the total elimination of the bacilli from the host organism.[62] If we denote by $r_N$ the endogenous relapse rate of naturally recovered individuals we have:

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

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

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

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

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

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

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

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

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

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

where $r_S$ is the endogenous relapse rate after successful treatment completion.

In what regards its estimation, there exist many epidemiological studies based on the surveillance of cohorts of TB patients after treatment completion during defined follow-up periods, which are aimed at determining the relapse rates, as well as the main risk factors associated to its increment. In the exhaustive meta-analysis by Korenromp and colleagues,[62] an ensemble of such studies is considered. In that work, it is reported that, in all the works re-analized, an average of 4.2% (3.1−5.3 c.i.) of HIV uninfected subjects have a TB relapse episode during the follow-up period of the study, of which, 77% ($63-91$ C.I.) is due to endogenous reactivation. This means that the fraction of population that do not develop a relapse is the 96.77% (95.73−97.80).

Another relevant result of the meta-analysis is the finding that the risk for TB relapse after treatment decreases with time. This can be seen from the fact that the relapse rates calculated in the different studies considered tend to be lower as the follow-up period of the trials is higher. This would imply that most patients that experiment a relapse after treatment, do it within the first years after the initial episode. This second result motivates the assumption that the risk of developing a relapse during the follow-up period of an epidemic surveillance study ($100 - 96.77 = 3.23\%$ of the population) can be associated to the total risk of developing such relapse during the entire life of an individual. Hence, the task is to calculate an annual risk of relapse such that, when applied over the whole period of life expectancy of a recovered individual, it yields the same 3.23% of relapse cases. To this end, we estimate that the average life expectancy of individuals within classes $R$ is equal to 35 years, estimation that follows from assuming, as a first order approximation, that infection and further recovery are events that occur uniformly in all ages.

Therefore, and since we are assuming that the relapse rate is constant in age and time, we have an exponential decay describing the relapse of $R_{xS}$ individuals ($R_{p+S}$, $R_{p-S}$ or $R_{npS}$) of the form: $R_{xS}(t) \sim e^{-r_S t}$. Thus, after a period $t = 35$ years assimilable to the average life expectancy of an individual

that has already entered into class $R$, from an initial fraction of $R_{xS} = 1$, there remains $R_{xS}(t = 35) \sim e^{-r_S 35} = 0.9677$. This calculation yields the actual value of $r_S$ used in this work, $r_S = 9.4 \cdot 10^{-4}$ 1/year ($6.4 \cdot 10^{-4} - 1.3 \cdot 10^{-3}$). The confidence interval of $r_S$ has been obtained after the propagation of the fraction of not relapsing population as the main source of uncertainty.

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

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

$r_D$ stands for the endogenous relapse rate after treatment default, which has been calculated as the product of $r_S$ and the relative risk factor for endogenous relapse related to treatment non-compliance, 4.02 (1.79-9.01 c.i.), taken from Picon et al.[156], which yields the final value of $r_D = 3.8 \cdot 10^{-3}$ 1/year ($1.4 \cdot 10^{-3} - 8.6 \cdot 10^{-3}$).

#### 3.2.1.8 Exogenous reinfection of infected individuals

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

- Exogenous re-infection of $L_s(a,t)$ individuals yielding fast progression: from $L_s(a,t)$ to $L_f(a,t)$:
  $p(a)q\lambda(a,t)L_s(a,t)$ individuals/unit time.

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

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

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

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

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

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

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

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

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

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

On the other hand, if fast progression after the secondary infection does not take place, even if the initial state of the individual is one of the possible $R$ states, the rule is that no transition must be considered from these states to $L_s$, because an endogenous reactivation from those initial states to disease

is always more likely than from $L_s$, as either $r_N$, $r_S$ or $r_D$ are greater than the rate of transition from slow latency to disease $\omega_s$. However, we will count these potential transitions when accounting for the number of infections that occurred in the system.

In summary, re-infection has no effect on the 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.[157] it is described how the progression to TB disease increases with the number of exposures.

#### 3.2.1.9 Smear progression

In certain cases, it is documented that patients of smear negative pulmonary TB progress to smear positive ,[108] even after being treated. In order to describe this phenomenon, we introduce the smear progression by considering the following two transitions:

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

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

where $\theta$ stands for the smear progression rate.

#### 3.2.1.10 Mother-child infection transmission

We model the possibility of mother-child transmission right after birth (peri-natal infection). While most of the newborns that enter the system at each time step $\Delta_N(a=0,t)$ (see section 3.2.5 to see how we estimate it) will enter the system as susceptible, a fraction of them will do it directly to the latency classes. This reflects the known fact that a fraction $m_c$ of sick women who are pregnant transmits the disease to the children within the first weeks of their lives.[158] The density of infected newborns depends then on the fraction of mothers who have the disease and are able to transmit it at time step $t$, which leads to the question of what is the relative risk of transmitting the pathogen to the offspring for women in each of the infectious classes included in the model. In this work we considered that the total number of newborn infections is proportional to $m_d(t)$:

$$m_d(t) = \frac{\sum_{a=3}^{a=7} D_{p+}(a,t) + D_{p-}(a,t) + D_{np}(a,t)}{\sum_{a=3}^{a=7} N(a,t)} + \frac{\sum_{a=3}^{a=7} R_{p+D}(a,t) + R_{p-D}(a,t) + R_{npD}(a,t) + F(a,t)}{\sum_{a=3}^{a=7} N(a,t)} \quad (3.3)$$

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

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

### 3.2.2 Force of infection

The force of infection $\lambda(a,t)$ is, as it has been said before in section 3.2.1.1, the rate at which infection occurs at time step $t$ for a susceptible individual in age-group $a$. This magnitude is calculated according to the following expression:

$$\lambda(a,t) = \beta(t) \sum_{a'} M^c_{a,a'}(t) \Upsilon(a',t) \quad (3.4)$$

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

$$\Upsilon(a',t) = \frac{D_{p+}(a',t) + F(a',t) + \phi_{p-} D_{p-}(a',t)}{N(a',t)} + \frac{\phi_D R_{p+D}(a',t) + \phi_{p-}\phi_D R_{p-D}(a',t)}{N(a',t)} \quad (3.5)$$

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

On the other hand, $M^c_{a,a'}(t)$ represents the relative contact frequency that an individual of age $a$ has with individuals of age $a'$ at time $t$, with respect to the overall average of contacts that an individual has per unit time with anyone else.

### 3.2.3 Contact patterns

One of the two principal novelties introduced in this work is the use of empirical age-dependent matrices. It has been previously shown that abandoning the hypothesis of homogeneous age-mixing in favor of data-driven approaches based on empiric data[25,146] improves the descriptive capabilities of epidemiological models of influenza-like diseases.[147,148] Similarly, in a study by Guzzetta and collaborators,[114] the importance of considering heterogeneous contacts in the modeling of TB was assessed in the context of Individual Based Modeling (IBM). Despite these first attempts, in mathematical TB modeling at the level of broad populations (i.e. countries or international macro-regions) the assumption of homogeneous mixing across age-strata still constitutes one of the most pervasive simplifying hypothesis.

We have discussed thourughly in chapter 2 the implications of implementing an empirical contact matrix in demographic settings that are different of that where it was measured. In what concerns the development of this model, there are two warnings we must consider: we should correct contact matrices for reciprocity, and we shouldn't export contact matrices across different geographical areas if possible. With this in mind, we take two decisions: 1) we will use method 3 (M3) (see equation 2.12) to adapt contact matrices to the specific demographic structure at each time-step and 2) we will build continental intrinsic connectivities that we then implement for each country according to its geographical location.

For the computation of the contact matrices used here, we have collected different survey studies from several countries: Kenya[134], Zimbabwe[137], Uganda[136], China[130], Japan[133] and Europe (8 countries)[25]. With the first three studies

we build a contact matrix that will be adapted to be used in African countries (specifically for this work: Nigeria, South Africa, Democratic Republic of the Congo, Ethiopia and Tanzania), while the next two are used to build an Asian Contact Matrix (to be adapted to India, Indonesia, China, Pakistan, Bangladesh, Philippines and Myanmar). Finally, the surveys performed in the Polymod project, that corresponds to 8 European countries will be aggregated into an European Contact Matrix that we will use to check the robustness of the results against different contact structures (used only in sections 3.3.5.5 and 3.3.7.1).

In order to construct continent-wise contact patterns, we begin by applying the same initial treatment described in chapter 2, section 2.2.3. Once we have the normalized and reciprocal matrices for each study, we perform a weighted sum per continent (using number of participants in each study as weights) to obtain the correspondent regional matrices, that we will call $M^{\text{reg}}_{a,a'}$ (for Africa, Asia and Europe, respectively, see figure 3.2, left column). We then use the dispersion between studies to define the uncertainty associated to the contact patterns (which we will be propagating to the final matrices and finally, to model outcomes).

These regional matrices $M^{\text{reg}}_{a,a'}$ only fulfill the symmetry of contacts in the setting of reference where they have been obtained, defined as the union of the countries being averaged in each case. In order to apply M3 to fulfill symmetry at every demographic setting (and at each time step), we extract the correspondent intrinsic connectivity matrix ($\Gamma^{\text{reg}}_{a,a'}$) by using equation 2.10. These regional connectivity matrix will be used to build at each time step the contact patterns of the system using M3 (eq. 2.12)

We have decided to use M3 (that ensures a constant mean connectivity) instead of the most natural M2 for two reasons. On the one hand, the main objective that we pursue in this work regarding contact patterns is to quantify the influence that they exert on model forecasts, by comparing the outcomes of this model to a case where contacts are considered to be homogeneous. Since, in the latter case, the mean contact intensity is trivially constant over time, to use heterogeneous contact matrices that share this same property constitutes a conservative choice that makes the comparison between the full and the reduced model easier to interpret.

On the other hand, this procedure implies that the average contact intensity in both cases is not just time-invariant, but equal to 1 in both cases. This makes the scaled infectiousness parameter $\beta(t)$ to be comparable be-

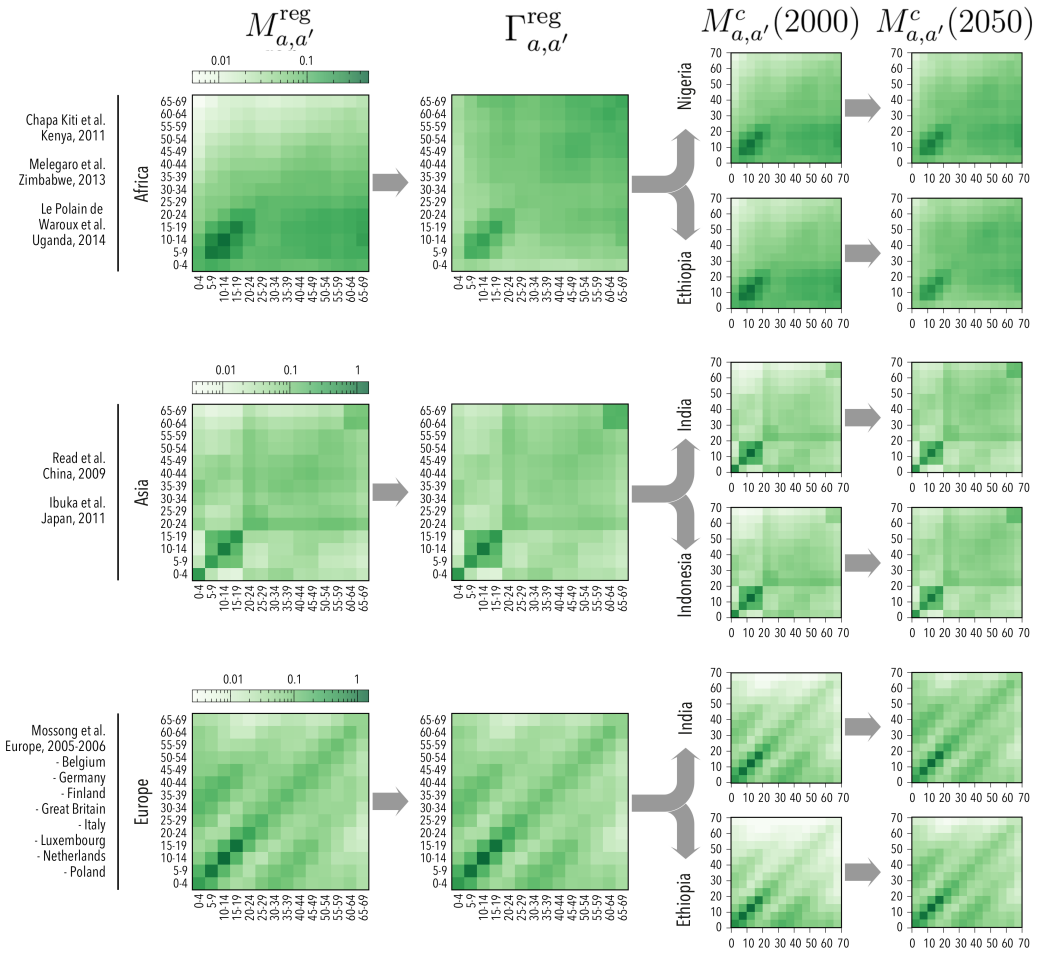


Figure 3.2: **Continent-wise empiric contact patterns.** Empiric contact patterns in evolving demographies, as number of contacts that an individual in age-group $a$ (x-axis) has with individuals in age group $a'$ (y-axis). (Left) Region-wise contact matrices $M^{\mathrm{reg}}_{a,a'}$ derived from weighted averages from studies completed in each continent. (Middle) Intrinsic connectivity matrices $\Gamma^{\mathrm{reg}}_{a,a'}$, (Normalized, in order to display $\langle k \rangle = 1$) (Right): Country-wise, time dependent contact matrices used in this study for the four countries that have been studied in most depth. European contacts are showed when applied to Ethiopia and India, to illustrate that the approach can be used to test the effect of alien contact structures obtained in any setting (these are only used in the analyses presented in figures 3.16 and 3.13)

tween the full and the reduced model, thus providing an overall scale for the average capability of an infected individual to propagate the disease in both

cases.

Besides, any possible evolution in the mean connectivity will be absorbed by the infectious parameter $\beta(t)$. Formally, this can be interpreted considering that $\beta$ is the product of two indistinguishable nuisance parameters: the average connectivity of the network of contacts (i.e. the average number of epidemiologically relevant interactions that any individual has per unit time) and the "intrinsic" infectiousness, (i.e. the probability of a contagion to occur upon one of those contacts, if established between an infectious and a susceptible individual). Among the alternative scenarios that we have studied to assess the robustness of the results, we include different temporal evolutions of $\beta(t)$.

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

$$\frac{M^c_{a,a'}(t_1)}{M^c_{a,a''}(t_1)} \Big/ \frac{M^c_{a,a'}(t_0)}{M^c_{a,a''}(t_0)} = \frac{N(a',t_1)}{N(a'',t_1)} \Big/ \frac{N(a',t_0)}{N(a'',t_0)} \tag{3.6}$$

In the figure 3.2, we summarize this process, showing the contact patterns obtained at each step, normalized to a common scale in each region. As it can be observed, the corrections explained above designed to capture the influence of the different demographic structures across countries and time introduce slight variations when compared to the differences observed between the three geographical areas.

We will compare the results with a reduced model in which we maintain the classical hypothesis of contact homogeneity (reduced model 2, RM2). This consists of assuming that the probability for two individuals to interact is the same, which means that the contact frequency that an individual in group $a$ has with people in group $a'$ only depends on the frequency of the target group in the population:

$$M^{c,\mathrm{RM2}}_{a,a'}(t) = N(a',t)/N(t) \tag{3.7}$$

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

model.

In the tables 3.3, 3.4 and 3.5, we specify the values of the intrinsic connectivity matrices $\Gamma^{\text{reg}}_{a,a'}$ for the three macro-regions considered (Africa, Asia and Europe). Notice that these are not directly the contact patterns considered, but they have to be corrected by the demography of the individual setting (equation 2.12).

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

Table 3.3: **Intrinsic Connectivity Matrix $\Gamma^{\text{reg}}_{a,a'}$ for Africa**

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

Table 3.4: **Intrinsic Connectivity Matrix $\Gamma^{\text{reg}}_{a,a'}$ for Asia**

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

Table 3.5: **Intrinsic Connectivity Matrix $\Gamma^{\mathbf{reg}}_{a,a'}$ for Europe**

### 3.2.4 Aging

The model considers 15 different age groups. Each of these groups comprises an age interval of $\Delta_t = 5$ years, except the last one that corresponds to individuals older that 70 years old. The relevance of such an age structured description of the system comes from the fact that some of the most relevant dynamical parameters take distinct values for each age group. To account for the aging of individuals as time goes by -and thus the evolution of their age $a$- we introduce on the system of equations the following aging transitions, that stands for the promotion of individuals within whatever dynamical class of the model $X(a,t)$ to the next age class $X(a+1,t)$.

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

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

### 3.2.5 Demographic evolution

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

$$\begin{aligned} N(a,t) = S(a,t) + L_f(a,t) + L_s(a,t) + D_{p+}(a,t) + D_{p-}(a,t) \\ +D_{np}(a,t) + F(a,t) + T_{p+}(a,t) + T_{p-}(a,t) + T_{np}(a,t) \\ +R_{p+N}(a,t) + R_{p-N}(a,t) + R_{npN}(a,t) + R_{p+S}(a,t) + R_{p-S}(a,t) \\ +R_{npS}(a,t) + R_{p+D}(a,t) + R_{p-D}(a,t) + R_{npD}(a,t) \end{aligned} \tag{3.8}$$

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 on 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 $\mathcal{N}(t) = \mathcal{N}(t=0)\forall(t)$.[55] A more sophisticated alternative, adopted in Dye et al.[108], is based on imposing that the system preserves the initial age structure of the population by making that, in each age group: $N(a,t) = N(a,t=t_o)\forall(t)$. This 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: $N(a,t) = N_{UN}(a,t)$.[138] Population aging is a common feature in virtually all the countries under study.

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

**Reduced model 1: Constant demographic structure**

A first approach consists on 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,[108] where the dynamical states indicate densities rather than numbers of individuals. In this case, the force of infection is calculated as an average of the densities of sick individuals in each age group, weighted by the number of individuals within each age class of the demographic structure. In this way, Dye et al.[108] 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 the model –where states represent number of individuals rather than densities–, we start by calculating the variation of population due to TB and aging in each age group:

$$\begin{aligned}\dot{N}_o(a,t) = ((1-\delta(a))N(a-1,t) - N(a,t))/\Delta_t \\ -\mu_{p+}(D_{p+}(a,t) + F(a,t)) - \mu_{p-}D_{p-}(a,t) - \mu_{np}D_{np}(a,t) \\ -\Psi f_\mu^{p+} T_{p+}(a,t) - \Psi f_\mu^{p-}(T_{p-}(a,t) + T_{np}(a,t))\end{aligned} \tag{3.9}$$

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

$$\dot{N}(a,t) = \dot{N}_o(a,t) + \Delta_N(a,t) = 0 \ \forall(a,t) \tag{3.10}$$

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:

$$\Delta_X(a,t) = \Delta_N(a,t)\frac{X(a,t)}{N(a,t)} \tag{3.11}$$

and obviously:

$$\Delta_N(a,t) = \sum_X \Delta_X(a,t) \tag{3.12}$$

This scheme has the advantage, with respect to consider (as in Abu-Raddad et al.[55]), 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.[108] Additionally, the main problem with this approach comes from the fact that no variation of the age structure of the population can be considered by proceeding this way, which might introduce significant biases from current demographic

forecasts, specially when studying populations subjected to strong processes of demographic aging.

**Full model: Evolving demography according to an external constraint**

Starting from the last scheme for modeling the demographic evolution, it is easy to obtain a final approach that explicitly considers not only the influence of the age structure into the spreading, but also the population growth and the variation of the age structure itself. To this end, it is necessary to know the actual –or projected– evolution of the demographic structure of the population during the period under analysis. In this 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.[138] Thus, from the UN database we obtain the actual annual population series expected by the UN for the populations at each age group, which can be trivially fitted to a continuous function $N_{UN}(a,t)$, from which we can derive an analytical derivative $\dot{N}_{UN}(a,t)$ at any moment. For the purpose of this work, a polynomial of degree 10 is more than enough for building the continuous function $N_{UN}(a,t)$ from the annual data series from UN Database during the period under study.[138]

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 3.9, we can also introduce a term $\Delta_N(a,t)$, aimed, this time, not at balancing $\dot{N}_o(a,t)$, but at forcing the total temporal evolution of $N(a,t)$ to follow precisely the function $N_{UN}(a,t)$. This is achieved by defining, at each time step:

$$\Delta_N(a,t) = \dot{N}_{UN}(a,t) - \dot{N}_o(a,t) \tag{3.13}$$

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 $a > 0$:

$$\Delta_X(a,t) = \Delta_N(a,t)\frac{X(a,t)}{N(a,t)}\ \forall a > 0 \tag{3.14}$$

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

The assumption that migration occurs independently of the disease state is arguable, and, in principle, is hard to anticipate whether TB patients (or latent TB carriers) are more or less prune to migrate than susceptible individuals. However, without specific data that could motivate an informed alternative, we decided to take as the null hypothesis that all individuals are equally prune to migrate regardless of their TB status. Nonetheless, migratory fluxes do not represent the major cause of population variation in any of the twelve countries studied, where the migratory balance (immigrants-emigrants) supposes less than 25% of the vegetative growth (less than 10% in 8 over 12 countries, all but South Africa, China, Myanmar and Bangladesh) during the period under study. Thus, if the actual reality is more complex than this assumptions, and migrants and not migrants do present different TB prevalence levels, the effects of these differences should be bound by the reduced role of migration in the total variation of the populations under study.

The situation is different for the first age class $a = 0$. In the first age group, the birth of new individuals is the main cause of population variation. For these reasons, and once observed that $\Delta_N(a = 0, t) > 0 \; \forall t$ in all countries under consideration, for simplicity $\Delta_N(a = 0, t)$ is directly associated to the number of newborns and introduced in the $S$, $L_s$ and $L_f$ states, as described in section 3.2.1.10.

The uncertainty of UN demographic projections is also reported at UN Database,[138] which allows us to reconstruct the demographic structures at the extremes of the confidence interval (95%) $N_{UN}^{\text{low}}(a,t)$ and $N_{UN}^{\text{high}}(a,t)$. Therefore, its influence on the model forecasts is also measurable, as we will discuss in the section devoted to uncertainty and sensitivity analysis (section 3.2.10).

### 3.2.6 Ordinary differential equations system

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

$$\begin{aligned}\dot{S}(a,t) &= -\lambda(a,t)S(a,t) - (1-\delta(a-14))S(a,t)/\Delta_t \\ &+ (1-\delta(a))S(a-1,t)/\Delta_t + \delta(a)(1-m_c m_d(t))\Delta_N(a,t) \\ &+ (1-\delta(a))\Delta_N(a,t)S(a,t)/N(a,t)\end{aligned} \tag{3.15}$$

$$\begin{aligned}\dot{L}_s(a,t) &= (1-p(a))\lambda(a,t)S(a,t) - p(a)q\lambda(a,t)L_s(a,t) - \omega_s L_s(a,t) \\ &+ \delta(a)m_c m_d(t)(1-p(0))\Delta_N(a,t) - (1-\delta(a-14))L_s(a,t)/\Delta_t \\ &+ (1-\delta(a))L_s(a-1,t)/\Delta_t + (1-\delta(a))\Delta_N(a,t)L_s(a,t)/N(a,t)\end{aligned} \tag{3.16}$$

$$\begin{aligned}\dot{L}_f(a,t) &= p(a)\lambda(a,t)S(a,t) + p(a)q\lambda(a,t)(L_s(a,t) + R_{p+N}(a,t)) \\ &+ p(a)q\lambda(a,t)(R_{p-N}(a,t) + R_{npN}(a,t) + R_{p+S}(a,t) + R_{p-S}(a,t)) \\ &+ p(a)q\lambda(a,t)(R_{npS}(a,t) + R_{p+D}(a,t) + R_{p-D}(a,t) + R_{npD}(a,t)) \\ &- \omega_f L_f(a,t) - ((1-\delta(a-14))L_f(a,t) - (1-\delta(a))L_f(a-1,t))/\Delta_t \\ &+ \delta(a)m_c m_d(t)p(0)\Delta_N(a,t) + (1-\delta(a))\Delta_N(a,t)L_f(a,t)/N(a,t)\end{aligned} \tag{3.17}$$

$$\begin{aligned}\dot{D}_{p+}(a,t) &= \omega_f\rho_{p+}(a)L_f(a,t) + \omega_s\rho_{p+}(a)L_s(a,t) - \mu_{p+}D_{p+}(a,t) \\ &- d(t)D_{p+}(a,t) - \nu D_{p+}(a,t) + r_N R_{p+N}(a,t) + r_S R_{p+S}(a,t) \\ &+ r_D R_{p+D}(a,t) + \theta D_{p-}(a,t) - (1-\delta(a-14))D_{p+}(a,t)/\Delta_t \\ &+ (1-\delta(a))D_{p+}(a-1,t)/\Delta_t + (1-\delta(a))\Delta_N(a,t)D_{p+}(a,t)/N(a,t)\end{aligned} \tag{3.18}$$

$$\begin{aligned}\dot{D}_{p-}(a,t) &= \omega_f(1-\rho_{p+}(a)-\rho_{np}(a))L_f(a,t) - \rho_{np}(a))L_s(a,t) \\ &+ \omega_s(1-\rho_{p+}(a) - \eta d(t)D_{p-}(a,t) + r_N R_{p-N}(a,t) \\ &+ r_S R_{p-S}(a,t) + r_D R_{p-D}(a,t) - (1-\delta(a-14))D_{p-}(a,t)/\Delta_t \\ &+ (1-\delta(a))D_{p-}(a-1,t)/\Delta_t + (1-\delta(a))\Delta_N(a,t)D_{p-}(a,t)/N(a,t) \\ &- \mu_{p-}D_{p-}(a,t) - \nu D_{p-}(a,t) - \theta D_{p-}(a,t)\end{aligned} \tag{3.19}$$

$$\begin{aligned}\dot{D}_{np}(a,t) &= \omega_f\rho_{np}(a)L_f(a,t) + \omega_s\rho_{np}(a)L_s(a,t) \\ &- \mu_{np}D_{np}(a,t) - \eta d(t)D_{np}(a,t) - \nu D_{np}(a,t) \\ &+ r_N R_{npN}(a,t) + r_S R_{npS}(a,t) + r_D R_{npD}(a,t) \\ &- ((1-\delta(a-14))D_{np}(a,t) - (1-\delta(a))D_{np}(a-1,t))/\Delta_t \\ &+ (1-\delta(a))\Delta_N(a,t)D_{np}(a,t)/N(a,t)\end{aligned} \tag{3.20}$$

$$\begin{aligned}\dot{T}_{p+}(a,t) &= d(t)D_{p+}(a,t) - \Psi T_{p+}(a,t) + \theta T_{p-}(a,t) \\ &- ((1-\delta(a-14))T_{p+}(a,t) - (1-\delta(a))T_{p+}(a-1,t))/\Delta_t \\ &+ (1-\delta(a))\Delta_N(a,t)T_{p+}(a,t)/N(a,t)\end{aligned} \tag{3.21}$$

$$\begin{aligned}\dot{T}_{p-}(a,t) &= \eta d(t)D_{p-}(a,t) - \Psi T_{p-}(a,t) - \theta T_{p-}(a,t) \\ &- ((1-\delta(a-14))T_{p-}(a,t) - (1-\delta(a))T_{p-}(a-1,t))/\Delta_t \\ &+ (1-\delta(a))\Delta_N(a,t)T_{p-}(a,t)/N(a,t)\end{aligned} \tag{3.22}$$

$$\begin{aligned}\dot{T}_{np}(a,t) &= \eta d(t)D_{np}(a,t) - \Psi T_{np}(a,t) \\ &- ((1-\delta(a-14))T_{np}(a,t) - (1-\delta(a))T_{np}(a-1,t))/\Delta_t \\ &+ (1-\delta(a))\Delta_N(a,t)T_{np}(a,t)/N(a,t)\end{aligned} \tag{3.23}$$

$$\begin{aligned}\dot{F}(a,t) &= \Psi f_F^{p+} T_{p+}(a,t) + \Psi f_F^{p-}(T_{p-}(a,t) + T_{np}(a,t)) - \mu_{p+}F(a,t) \\ &- ((1-\delta(a-14))F(a,t) - (1-\delta(a))F(a-1,t))/\Delta_t \\ &+ (1-\delta(a))\Delta_N(a,t)F(a,t)/N(a,t)\end{aligned} \tag{3.24}$$

$$\begin{aligned}\dot{R}_{p+N}(a,t) &= \nu D_{p+}(a,t) - r_N R_{p+N}(a,t) - p(a)q\lambda(a,t)R_{p+N}(a,t) \\ &- ((1-\delta(a-14))R_{p+N}(a,t) - (1-\delta(a))R_{p+N}(a-1,t))/\Delta_t \\ &+ (1-\delta(a))\Delta_N(a,t)R_{p+N}(a,t)/N(a,t)\end{aligned} \tag{3.25}$$

$$\begin{aligned}\dot{R}_{p-N}(a,t) &= \nu D_{p-}(a,t) - r_N R_{p-N}(a,t) - p(a)q\lambda(a,t)R_{p-N}(a,t) \\ &- ((1-\delta(a-14))R_{p-N}(a,t) - (1-\delta(a))R_{p-N}(a-1,t))/\Delta_t \\ &+ (1-\delta(a))\Delta_N(a,t)R_{p-N}(a,t)/N(a,t)\end{aligned} \tag{3.26}$$

$$\begin{aligned}\dot{R}_{npN}(a,t) &= \nu D_{np}(a,t) - r_N R_{npN}(a,t) - p(a)q\lambda(a,t)R_{npN}(a,t) \\ &- ((1-\delta(a-14))R_{npN}(a,t) - (1-\delta(a))R_{npN}(a-1,t))/\Delta_t \\ &+ (1-\delta(a))\Delta_N(a,t)R_{npN}(a,t)/N(a,t)\end{aligned} \tag{3.27}$$

$$\begin{aligned}
\dot{R}_{p+S}(a,t) &= \Psi(1-f_D^{p+}-f_F^{p+}-f_\mu^{p+})T_{p+}(a,t) \\
&- r_S R_{p+S}(a,t) - p(a)q\lambda(a,t)R_{p+S}(a,t) \\
&- ((1-\delta(a-14))R_{p+S}(a,t) - (1-\delta(a))R_{p+S}(a-1,t))/\Delta_t \\
&+ (1-\delta(a))\Delta_N(a,t)R_{p+S}(a,t)/N(a,t)
\end{aligned} \tag{3.28}$$

$$\begin{aligned}
\dot{R}_{p-S}(a,t) &= \Psi(1-f_D^{p-}-f_F^{p-}-f_\mu^{p-})T_{p-}(a,t) \\
&- r_S R_{p-S}(a,t) - p(a)q\lambda(a,t)R_{p-S}(a,t) \\
&- ((1-\delta(a-14))R_{p-S}(a,t) - (1-\delta(a))R_{p-S}(a-1,t))/\Delta_t \\
&+ (1-\delta(a))\Delta_N(a,t)R_{p-S}(a,t)/N(a,t)
\end{aligned} \tag{3.29}$$

$$\begin{aligned}
\dot{R}_{npS}(a,t) &= \Psi(1-f_D^{p-}-f_F^{p-}-f_\mu^{p-})T_{np}(a,t) \\
&- r_S R_{npS}(a,t) - p(a)q\lambda(a,t)R_{npS}(a,t) \\
&- ((1-\delta(a-14))R_{npS}(a,t) - (1-\delta(a))R_{npS}(a-1,t))/\Delta_t \\
&+ (1-\delta(a))\Delta_N(a,t)R_{npS}(a,t)/N(a,t)
\end{aligned} \tag{3.30}$$

$$\begin{aligned}
\dot{R}_{p+D}(a,t) &= \Psi f_D^{p+}T_{p+}(a,t) - r_D R_{p+D}(a,t) - p(a)q\lambda(a,t)R_{p+D}(a,t) \\
&- ((1-\delta(a-14))R_{p+D}(a,t) - (1-\delta(a))R_{p+D}(a-1,t))/\Delta_t \\
&+ (1-\delta(a))\Delta_N(a,t)R_{p+D}(a,t)/N(a,t)
\end{aligned} \tag{3.31}$$

$$\begin{aligned}
\dot{R}_{p-D}(a,t) &= \Psi f_D^{p-}T_{p-}(a,t) - r_D R_{p-D}(a,t) - p(a)q\lambda(a,t)R_{p-D}(a,t) \\
&- ((1-\delta(a-14))R_{p-D}(a,t) - (1-\delta(a))R_{p-D}(a-1,t))/\Delta_t \\
&+ (1-\delta(a))\Delta_N(a,t)R_{p-D}(a,t)/N(a,t)
\end{aligned} \tag{3.32}$$

$$\begin{aligned}
\dot{R}_{npD}(a,t) &= \Psi f_D^{p-}T_{np}(a,t) - r_D R_{npD}(a,t) - p(a)q\lambda(a,t)R_{npD}(a,t) \\
&- ((1-\delta(a-14))R_{npD}(a,t) - (1-\delta(a))R_{npD}(a-1,t))/\Delta_t \\
&+ (1-\delta(a))\Delta_N(a,t)R_{npD}(a,t)/N(a,t)
\end{aligned} \tag{3.33}$$

where $\delta(a)$ stands for the Dirac delta function ($\delta(x=0)=1$ and $\delta(x\neq 0)=0$). There are three quantities that depend on time: the force of infection

$\lambda(a,t)$, the diagnosis rate $d(t)$ and the correction terms $\Delta_N(a,t)$, standing for any demographic variation in the population due to causes foreign to TB and aging.

It is useful 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:

$$\begin{aligned}\dot{I}(a,t) &= \omega_f L_f(a,t) + \omega_s L_s(a,t) \\ &\quad + r_N \left(R_{p+N}(a,t) + R_{p-N}(a,t) + R_{npN}(a,t)\right) \\ &\quad + r_S \left(R_{p+S}(a,t) + R_{p-S}(a,t) + R_{npS}(a,t)\right) \\ &\quad + r_D \left(R_{p+D}(a,t) + R_{p-D}(a,t) + R_{npD}(a,t)\right)\end{aligned} \tag{3.34}$$

$$\begin{aligned}\dot{M}(a,t) &= \mu_{p+}(D_{p+}(a,t) + F(a,t)) + \mu_{p-} D_{p-}(a,t) \\ &\quad + \mu_{np} D_{np}(a,t) + \Psi f_\mu^{p+} T_{p+}(a,t) \\ &\quad + \Psi f_\mu^{p-} (T_{p-}(a,t) + T_{np}(a,t))\end{aligned} \tag{3.35}$$

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:

$$i(t) = \frac{1000000 \cdot \sum_a \left(I(a,t+1) - I(a,t)\right)}{(\mathcal{N}(t+1) + \mathcal{N}(t))/2} \tag{3.36}$$

$$m(t) = \frac{1000000 \cdot \sum_a \left(M(a,t+1) - M(a,t)\right)}{(\mathcal{N}(t+1) + \mathcal{N}(t))/2} \tag{3.37}$$

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.

### 3.2.7 Initial conditions setup

Once we have detailed the forces driving the time evolution of 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.[55,108] We denote those stationary levels as $\vec{X}^*(a, d_0, \beta_0, N(a,0))$, so we have $\dot{X}_i^* = 0 \ \forall(i,t)$; provided that all the time-dependent parameters and demographic forcing terms are frozen in their initial values at $t = t_o$. Accordingly, the stationary vector $\vec{X}^*$ is used to set up the initial conditions of the system: $\vec{X}(a,0) = \vec{X}^*$.

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

$$X(a,t=0) = (1+\varsigma)X^*(a, d_0, \beta_0, \vec{N}(a,0)) \qquad \forall(X \neq S) \tag{3.38}$$

$$S(a,t=0) = S^*(a, d_0, \beta_0, \vec{N}(a,0)) \left(1 - \varsigma \sum_{X \neq S} X^*(a, d_0, \beta_0, \vec{N}(a,0))\right) \tag{3.39}$$

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

$$S(a,t=0) = S^*(a, d_0, \beta_0, \vec{N}(a,0))(1-\varsigma) \tag{3.40}$$

$$X(a,t=0) = X^*(a, d_0, \beta_0, \vec{N}(a,0)) \left(1 + \frac{\varsigma S^*(a, d_0, \beta_0, \vec{N}(a,0))}{\sum_{X \neq S} X^*(a, d_0, \beta_0, \vec{N}(a,0))}\right) \tag{3.41}$$

Taking it to their extreme values, $\varsigma = -1$ would mean that every individual is at the susceptible state (pathogen-free situation) while $\varsigma = 1$ would mean that all the population is infected with the bacterium. $\varsigma = 0$ would imply that the initial conditions of the system are those from the stationary

state. The previous definition ensures that, at any moment, the sum of individuals in all the states provides the desired population volumes regardless of how far, or in which sense $\varsigma$ shifts the initial condition from the stationary state defined by the vector $\vec{X}^*$.

### 3.2.8 Model calibration procedure

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

$$d(t) = \begin{cases} d_0 + (d_{\text{sup}} - d_0)t(t + \frac{1}{d_1})^{-1} & \text{if } d_1 > 0 \\ d_0 & \text{if } d_1 = 0 \\ d_0 - d_0 t(t - \frac{1}{d_1})^{-1} & \text{if } d_1 < 0 \end{cases} \tag{3.42}$$

$$\beta(t) = \begin{cases} \beta_0 + \beta_0 t\left(t + \frac{1}{\beta_1}\right)^{-1} & \beta_1 > 0 \\ \beta_0 & \beta_1 = 0 \\ \beta_0 - \beta_0 t\left(t - \frac{1}{\beta_1}\right)^{-1} & \beta_1 < 0 \end{cases} \tag{3.43}$$

Therefore, the diagnosis rate and the scaled infectiousness are, each of them, parameterized by two quantities ($d_0$, $d_1$ and $\beta_0$, $\beta_1$). While $d_0$ and $\beta_0$ give the value of the diagnosis rate and scaled infectiousness at the beginning of the temporal window (i.e. year 2000), $d_1$ and $\beta_1$ define their evolution, either increasing or decreasing with time depending on the sign of $d_1$ and $\beta_1$. 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.[159]

We have chosen this parameterization of the temporal evolution of the diagnosis rate and the infectiousness because, unlike previous approaches where the evolution of these parameters is assimilated to an exponential curve,[55,108] it provides a bounded growth for them, through a function that

is still continuous and differentiable but does not introduce more parameters.

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

$$H = \sum_{t=t_o}^{t_F} \left( \left( \frac{i(t) - \bar{i}(t)}{\bar{\Delta}_i(t)} \right)^2 + \left( \frac{m(t) - \bar{m}(t)}{\bar{\Delta}_m(t)} \right)^2 \right) \tag{3.44}$$

where $\bar{i}(t)$ and $\bar{m}(t)$ stand for the annual incidence and mortality rates, corresponding to the national estimations available at the WHO database for TB.[149] These measurements of TB incidence and mortality have their correspondent confidence intervals $(\bar{i}_{\text{low}}(t), \bar{i}_{\text{high}}(t))$ and $(\bar{m}_{\text{low}}(t), \bar{m}_{\text{high}}(t))$, which are not necessarily symmetrical with respect to the central values $\bar{i}(t)$ and $\bar{m}(t)$. Using these confidence intervals, and taking into consideration their asymmetry, the corresponding terms $\bar{\Delta}_i(t)$ $\bar{\Delta}_m(t)$ are constructed as follows:

$$\bar{\Delta}_i(t) = \begin{cases} \bar{i}(t) - \bar{i}^{\text{low}}(t) & \text{if } i(t) \leq \bar{i}(t) \\ \bar{i}^{\text{high}}(t) - \bar{i}(t) & \text{if } i(t) > \bar{i}(t) \end{cases} \tag{3.45}$$

$$\bar{\Delta}_m(t) = \begin{cases} \bar{m}(t) - \bar{m}^{\text{low}}(t) & \text{if } m(t) \leq \bar{m}(t) \\ \bar{m}^{\text{high}}(t) - \bar{m}(t) & \text{if } m(t) > \bar{m}(t) \end{cases} \tag{3.46}$$

In the case of China and Philippines the very small uncertainty on mortality data (directly zero for some particular years) prevents us of using the previous equation 3.44. In those cases we minimize the absolute distance given by:

$$H = \sum_{t=t_o}^{t_F} \left( \left( \frac{i(t) - \bar{i}(t)}{\langle \bar{i}(t) \rangle} \right)^2 + \left( \frac{m(t) - \bar{m}(t)}{\langle \bar{m}(t) \rangle} \right)^2 \right) \tag{3.47}$$

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

The conceptual scheme for the fitting of these parameters essentially consists in an iterative evaluation of the model across the parameter space $(\varsigma, d_0, \beta_0, d_1, \beta_1)$, which is navigated according to a certain "routing" that eventually guarantees the localization of a parameter set that yields an error

$H$ which constitutes a local minimum of the objective function $H$. 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.[160] See figure 3.3 for a graphic summary of the procedure.

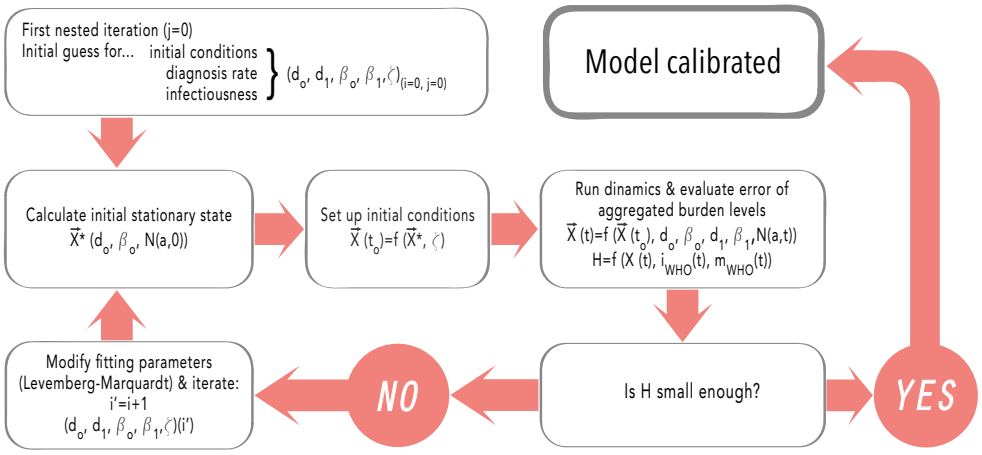


Figure 3.3: **Schematic representation of the calibration algorithm.**

### 3.2.9 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.

#### 3.2.9.1 Dynamic states

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

Table 3.6: **Description of the different dynamic states.**

#### 3.2.9.2 Literature-based epidemiological parameters

| Meaning | Parameter | Value | C.I. | Reference |
|---|---|---|---|---|
| Probability of fast progression | $p(a)$ | $(a=0)$ 0.187 | (0.1474,0.2333) | Marais et al.[53], this work |
| | | $(a=1)$ 0.0225 | (0.0200,0.0250) | |
| | | $(a>1)$ 0.15 | (0.10,0.20) | |
| Rate of fast progression ($y^{-1}$) | $\omega_f$ | 0.900 | (0.765,1.035) | Abu-Raddad et al.[55] |
| Rate of slow progression ($y^{-1}$) | $\omega_s$ | $7.500\times10^{-4}$ | $(6.375,8.625)\times10^{-4}$ | Abu-Raddad et al.[55] |
| Probability of smear-positive disease | $\rho_{p+}(a)$ | $(a<3)$ 0.100 | (0.085,0.115) | Abu-Raddad et al.[55] |
| | | $(a\geq3)$ 0.500 | (0.425,0.575) | |
| Probability of non-pulmonary disease | $\rho_{np}(a)$ | $(a<3)$ 0.250 | (0.2125,0.2875) | Abu-Raddad et al.[55] |
| | | $(a\geq3)$ 0.100 | (0.085,0.115) | |
| Mortality rate by pulmonary smear positive TB ($y^{-1}$) | $\mu_{p+}$ | 0.250 | (0.213,0.288) | Abu-Raddad et al.[55] |
| Mortality rate by pulmonary smear negative TB ($y^{-1}$) | $\mu_{p-}$ | 0.100 | (0.085,0.115) | Abu-Raddad et al.[55] |
| Mortality rate by non-pulmonary TB ($y^{-1}$) | $\mu_{np}$ | 0.100 | (0.085,0.115) | Abu-Raddad et al.[55] |
| Reduction of infection risk for previously infected individuals | q | 0.650 | (0.553,748) | Abu-Raddad et al.[55] |
| Treatment completion rate ($y^{-1}$) | $\Psi$ | 2.00 | (1.70,2.30) | Abu-Raddad et al.[55] |
| Smear progression rate ($y^{-1}$) | $\theta$ | 0.015 | (0.007,0.020) | Dye et al.[108] |
| Relapse rate for individuals who successfully completed treatment ($y^{-1}$) | $r_S$ | $9.392\times10^{-4}$ | $(6.364,12.450)\times10^{-4}$ | Korenromp et al.[62], this work |

Table 3.7: **Bibliography-based epidemiological parameters (part I).**

| Meaning | Parameter | Value | C.I. | Reference |
|---|---|---|---|---|
| Relapse rate for individuals who defaulted treatment ($y^{-1}$) | $r_D$ | $3.774\times10^{-3}$ | $(1.354,8.620)\times10^{-3}$ | Korenromp et al.[62], Picon et al.[156],this work |
| Relapse rate for naturally recovered individuals ($y^{-1}$) | $r_N$ | 0.030 | (0.020,0.040) | Dye et al.[108] |
| Natural recovery rate ($y^{-1}$) | $\nu$ | 0.100 | (0.085,0.115) | Dye et al.[108] |
| Infectiousness reduction coefficient of $D_{p-}$ with respect to $D_{p+}$ | $\phi_{p-}$ | 0.250 | (0.213,0.288) | Abu-Raddad et al.[55] |
| Infectiousness reduction coefficient of $R_{p+D}$ with respect to $D_{p+}$ | $\phi_D$ | 0.500 | (0.250,0.750) | Dye et al.[108] |
| Proportion of mothers that infect their newborn children | $m_c$ | 0.15 | (0.10,0.20) | Pillay et al.[158] |
| Diagnosis rate reduction of $D_{p-}$ and $D_{np}$ with respect to $D_{p+}$ | $\eta$ | Africa: 0.843 | (0.664,1.022) | Abu-Raddad et al.[55] , this work |
| | | Asia: 0.797 | (0.628,0.966) | |

Table 3.8: **Bibliography-based epidemiological parameters (part II).**

In tables 3.7 and 3.8 we represent the 19 epidemiological parameters used in the model, along with the eventual dependencies each of them show (to age, geographic setting, or none), the bibliographic source and the section of the appendix where the meaning of each parameter is explained.

#### 3.2.9.3 Treatment outcomes probabilities

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

- $(f_D^{p+}, f_F^{p+}, f_\mu^{p+})$: fraction of default, failure and death outcomes for smear positive pulmonary TB.[149]

- $(f_D^{p-}, f_F^{p-}, f_\mu^{p-})$: fraction of default, failure and death outcomes for smear negative pulmonary and non pulmonary TB.[149]

have been obtained from the WHO Treatment Outcomes database for each country, and their values in Ethiopia, Nigeria, India e Indonesia are presented in table 3.9:

| Parameter | Ethiopia | Nigeria | India | Indonesia | Reference | Section |
|---|---|---|---|---|---|---|
| $f_D^{p+}$ (%) | 3.84 (3.74,3.94) | 8.51 (8.36,8.66) | 5.97 (5.95,6.00) | 4.62 (4.58,4.66) | WHO Database[149] | 3.2.1.5 |
| $f_F^{p+}$ (%) | 1.04 (0.99,1.10) | 1.23 (1.16,1.29) | 2.07 (2.05,2.08) | 0.62 (0.61,0.64) | WHO Database[149] | 3.2.1.5 |
| $f_\mu^{p+}$ (%) | 3.97 (3.87,4.08) | 5.64 (5.51,5.76) | 4.48 (4.46,4.51) | 2.31 (2.28,2.34) | WHO Database[149] | 3.2.1.5 |
| $f_D^{p-}$ (%) | 3.28 (3.22,3.35) | 6.58 (6.43,6.72) | 6.11 (6.09,6.13) | 7.18 (7.12,7.24) | WHO Database[149] | 3.2.1.5 |
| $f_F^{p-}$ (%) | 0.12 (0.11,0.13) | 0.24 (0.21,0.27) | 0.42 (0.41,0.43) | 0.27 (0.26,0.28) | WHO Database[149] | 3.2.1.5 |
| $f_\mu^{p-}$ (%) | 3.53 (3.46,3.59) | 6.27 (6.13,6.41) | 3.11 (3.09,3.13) | 1.98 (1.94,2.01) | WHO Database[149] | 3.2.1.5 |

Table 3.9: **Treatment outcomes probabilities** in Ethiopia, Nigeria, India and Indonesia.

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

Once all the mentioned parameters are fixed, we obtain the time-evolving parameterization of diagnosis rates and infectiousness as the result of the calibration procedure explained in section 3.2.8, along with the initial conditions of the system in each country. These temporal evolutions are derived from equations 3.42 and 3.43, while the fitted values of the parameters $(d_o, d_1, \beta_o, \beta_1)$ are reported in table 3.10 for Ethiopia, Nigeria, India and Indonesia. Confidence Intervals are obtained through the procedure explained in section 3.2.10, as we do for any other model outcome.

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

Table 3.10: **Fitted parameters for different countries.**

In figure 3.4 we represent the evolution of $d(t)$ in these countries, which describes the average rate at which sick individuals receive their diagnosis in each country and time. The scaled infectiousness $\beta(t)$ has a less immediate epidemiological interpretation, for it is only directly proportional to the

number of infections $\tilde{R}_0$, that is caused, on average, by each infectious agent, which also depends on the distribution of individuals among the different infectious classes and age groups. This magnitude (which reduces to the basic reproductive number $R_0$ when evaluated on a fully susceptible population) is also represented in figure 3.4.

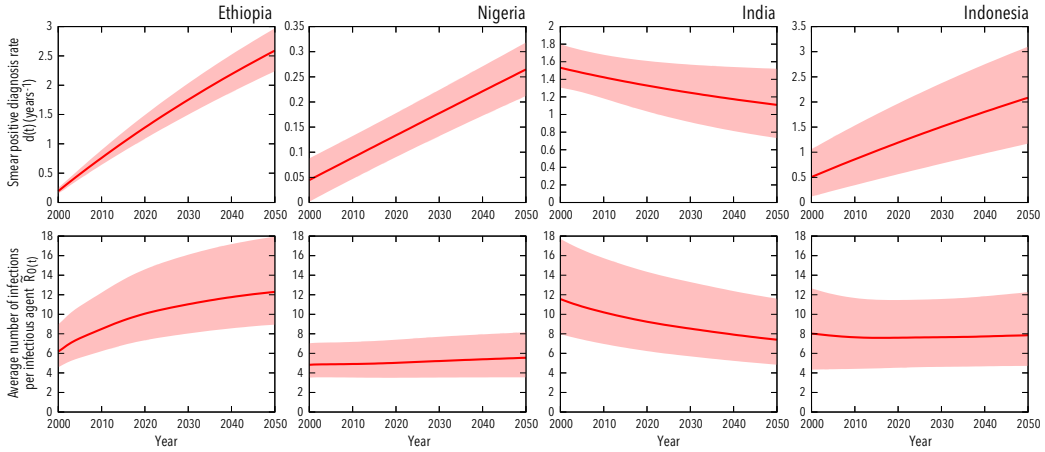


Figure 3.4: **Evolution of $\beta(t)$ and $\tilde{R}_0(t)$.** (Top) Diagnosis rates for smear-positive individuals (in years$^{-1}$). (Bottom) Average number of secondary infections per infectious agent $\tilde{R}_0$. The red curve shows the result given by the central fit, while the shadowed area represents the uncertainty obtained as described in section 3.2.10

We found very variable values of the diagnosis rates across the different countries. This is not unexpected, as the diagnosis time depends mostly on the quality of the Health System, for which we can find huge inequalities between countries. We will provide a more complete study of these inequalities and how they are captured by the model in section 3.3.5.2.

The number of secondary infections per incident case does not show the same variability. The values for $\tilde{R}_0$ lie between 5 and 12 secondary cases for the 4 countries analyzed during the entire window. These values are compatible with previous studies.[161]

#### 3.2.9.5 Fitted parameters in the reduced models

In table 3.11, we present the fitted parameters for the two reduced models in India, Indonesia, Nigeria and Ethiopia, compared to the values from the full model (already reported in 3.10).

| Country | Model | $d_0\,(y^{-1})$ | $d_1 \times 10^{-3}\,(y^{-1})$ | $\beta_0\,(y^{-1})$ | $\beta_1 \times 10^{-3}\,(y^{-1})$ | $\varsigma$ |
|---|---|---|---|---|---|---|
| Ethiopia | Complete Model | 0.19 (0.15,0.25) | 4.75 (3.94,5.64) | 7.10 (4.57,10.90) | 73.84 (20.62,103.83) | 0.16 (0.16,0.25) |
| | Constant Demography | 0.22 (0.16,0.27) | 4.40 (3.68,5.24) | 7.22 (4.64,11.02) | 70.97 (28.80,91.43) | 0.21 (0.05,0.31) |
| | Homogeneous Mixing | 0.19 (0.15,0.25) | 4.56 (3.80,5.36) | 7.18 (5.12,10.73) | 80.78 (36.04,110.66) | 0.23 (0.09,0.34) |
| Nigeria | Complete Model | 0.045 (0.003,0.088) | 0.36 (0.30,0.43) | 5.30 (3.36,8.03) | 3.71 (-3.93,11.15) | -0.29 (-0.53,-0.02) |
| | Constant Demography | 0.046 (0.004,0.090) | 0.37 (0.31,0.43) | 5.30 (3.35,8.04) | 2.71 (-3.67,9.85) | -0.27 (-0.52,-0.01) |
| | Homogeneous Mixing | 0.038 (0.003,0.080) | 0.35 (0.14,0.40) | 5.52 (3.81,8.15) | 2.91 (0.33,10.50) | -0.36 (-0.46,0.10) |
| India | Complete Model | 0.51 (0.12,1.06) | 2.98 (1.80,4.26) | 10.12 (5.10,16.99) | -1.20 (-12.01,9.19) | -0.30 (-0.44,-0.15) |
| | Constant Demography | 0.51 (0.13,1.04) | 2.66 (1.51,4.05) | 10.07 (5.12,16.78) | 6.16 (-0.39,12.97) | -0.25 (-0.37,-0.10) |
| | Homogeneous Mixing | 0.51 (0.12,1.06) | 2.94 (1.83,4.18) | 9.12 (4.55,15.20) | -1.04 (-11.38,11.52) | -0.30 (-0.45,-0.15) |
| Indonesia | Complete Model | 1.53 (1.31,1.80) | -7.60 (-20.99,-0.42) | 16.25 (10.82,25.59) | -12.73 (-17.73,-8.97) | 0.05 (0.04,0.07) |
| | Constant Demography | 1.53 (1.31,1.80) | -8.71 (-22.14,-1.40) | 16.14 (10.74,25.42) | -7.51 (-11.74,-4.83) | 0.08 (0.06,0.11) |
| | Homogeneous Mixing | 1.53 (1.31,1.80) | -8.09 (-21.64,-0.84) | 14.68 (9.63,23.38) | -12.20 (-16.72,-9.07) | 0.06 (0.05,0.08) |

Table 3.11: **Fitted parameters for different countries and models.**

### 3.2.10 Model uncertainty and sensitivity analysis

All the input data sources mentioned in the previous section, and summarized in Figure 3.1D, 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 model outcomes, produce significance estimates (i.e. p-values) as well as to evaluate the part of this uncertainty that is propagated from each of the model inputs.

#### 3.2.10.1 Uncertainty sources analysis

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

- Parameters associated with the Natural History of the disease: a total amount of 19 parameters, listed in table 3.7, each of them conservatively treated as totally independent uncertainty sources $u_i, i \in [1, 19]$.

- Burden and treatment outcomes estimations provided by WHO, listed in table 3.9. 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 the model, we have grouped these estimations produced by the WHO TB division as mutually dependent (see figure 3.1D), and considered them as one uncertainty source, labelled as $u_{20}$.

- Demographic structures $N(a,t)$: which are also considered as a single uncertainty source, labeled as $u_{21}$.

- Contact matrix $M^c_{a,a'}(t)$, whose uncertainty comes from the variability between studies, is the last single uncertainty source $u_{22}$.

By proceeding in this way, we have 22 uncertainty sources $u_i, i \in [1,22]$ that are considered independent, whose contributions to the uncertainty of a certain model outcome $x$ we want to evaluate. Here, a model outcome can be any possible magnitude that derives from the entire calibration-simulation procedure, as summarized in figure 3.1D. This includes, among others, incidence and mortality rates evaluated at any time (or averaged during the entire period), total accumulated number of incident cases or deaths, values for the parameters fitted during the calibration step, and importantly, differences between the full and the reduced models, either absolute or relative, of any of these primary outcomes. Future chapters will include other epidemiological forecasts obtained with this model, such as different measurments of the impact of a vaccine, that will follow the same uncertainty analysis.

The entire model-based calibration + simulation procedure, summarized in figure 3.1D, can be expressed, for what regards the estimation of any generic model outcome $x$, as a generic functional relationship $x = f(\vec{u})$, where $\vec{u}$ represents the 22-dimensional vector of uncertainty sources (i.e., input data).

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

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

$$d_i^{\text{low}}(x) = x(u_1, ..., u_i^{\text{low}}, ..., u_{21}) - x(\vec{u}) \tag{3.48}$$

or towards the upper limit:

$$d_i^{\text{high}}(x) = x(u_1, ..., u_i^{\text{high}}, ..., u_{21}) - x(\vec{u}) \tag{3.49}$$

Importantly, since the variation in $u_i$ that precedes the estimation of $d_i^{\text{low}}(x)$ and $d_i^{\text{high}}(x)$ occurs before model calibration, we are capturing, through this approach, the sensitivity of the entire procedure -calibration included- to the uncertainty input $u_i$, which implies that the signs of $d_i^{\text{low}}(x)$ and $d_i^{\text{high}}(x)$ cannot be trivially anticipated and can also coincide, as we see, in some cases, in figure 3.14. In that figure, we use red (blue) bars to represent the variations in the total number of TB cases/deaths that the model produces in 2000-2050 as a consequence of increasing (decreasing) the value of each uncertainty source to the upper (lower) limit of its respective confidence intervals.

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

**Grouping individual sensitivities according type of input data. Generation of confidence intervals and significance levels (Uncertainty analysis).** Once the individual sensitivity of all the n=22 sources of uncertainty are computed following the approach explained in the previous section, for the case of the 19 bibliographical parameters we separate positive versus negative sensitivities (i.e. sensitivity instances where the shift in the uncertainty source translates into an increase or a decrease in the model outcome), represented as red vs. blue bars in figure 3.14. Then, the square root of the sum of the squares of each type (positive and negative sensitivities) are computed. Denoted as $\Delta(x)^{\text{high}}_{\text{param}}$ and $\Delta(x)^{\text{low}}_{\text{param}}$, respectively, these quantities can be formally defined as follows:

$$\Delta(x)^{\text{high}}_{\text{param}} = \sqrt{\sum_{1}^{19} h(d_i^{\text{low}}(x)) \cdot d_i^{\text{low}}(x)^2 + h(d_i^{\text{high}}(x)) \cdot d_i^{\text{high}}(x)^2} \quad (3.50)$$

$$\Delta(x)^{\text{low}}_{\text{param}} = \sqrt{\sum_{1}^{19} h(-d_i^{\text{low}}(x)) \cdot d_i^{\text{low}}(x)^2 + h(-d_i^{\text{high}}(x)) \cdot d_i^{\text{high}}(x)^2} \quad (3.51)$$

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

$$\Delta(x)^{\text{high}}_{\text{WHO}} = \sqrt{h(d_{20}^{\text{low}}(x)) \cdot d_{20}^{\text{low}}(x)^2 + h(d_{20}^{\text{high}}(x)) \cdot d_{20}^{\text{high}}(x)^2} \quad (3.52)$$

$$\Delta(x)^{\text{low}}_{\text{WHO}} = \sqrt{h(-d_{20}^{\text{low}}(x)) \cdot d_{20}^{\text{low}}(x)^2 + h(-d_{20}^{\text{high}}(x)) \cdot d_{20}^{\text{high}}(x)^2} \quad (3.53)$$

For the demographic prospects ($u_{21}$):

$$\Delta(x)^{\text{high}}_{\text{demo}} = \sqrt{h(d_{21}^{\text{low}}(x)) \cdot d_{21}^{\text{low}}(x)^2 + h(d_{21}^{\text{high}}(x)) \cdot d_{21}^{\text{high}}(x)^2} \quad (3.54)$$

$$\Delta(x)^{\text{low}}_{\text{demo}} = \sqrt{h(-d_{21}^{\text{low}}(x)) \cdot d_{21}^{\text{low}}(x)^2 + h(-d_{21}^{\text{high}}(x)) \cdot d_{21}^{\text{high}}(x)^2} \quad (3.55)$$

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

$$\Delta(x)^{\text{high}}_{\text{cont.}} = \sqrt{h(d_{22}^{\text{low}}(x)) \cdot d_{22}^{\text{low}}(x)^2 + h(d_{22}^{\text{high}}(x)) \cdot d_{22}^{\text{high}}(x)^2} \quad (3.56)$$

$$\Delta(x)^{\text{low}}_{\text{cont.}} = \sqrt{h(-d_{22}^{\text{low}}(x)) \cdot d_{22}^{\text{low}}(x)^2 + h(-d_{22}^{\text{high}}(x)) \cdot d_{22}^{\text{high}}(x)^2} \quad (3.57)$$

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

$$\Delta(x)^{\text{low}} = \sqrt{(\Delta(x)^{\text{low}}_{\text{param}})^2 + (\Delta(x)^{\text{low}}_{\text{WHO}})^2 + (\Delta(x)^{\text{low}}_{\text{demo}})^2 + (\Delta(x)^{\text{low}}_{\text{cont.}})^2} \quad (3.58)$$

$$\Delta(x)^{\text{high}} = \sqrt{(\Delta(x)^{\text{high}}_{\text{param}})^2 + (\Delta(x)^{\text{high}}_{\text{WHO}})^2 + (\Delta(x)^{\text{high}}_{\text{demo}})^2 + (\Delta(x)^{\text{high}}_{\text{cont.}})^2} \quad (3.59)$$

In figure 3.5, and figure 3.9, in order to visualize the relative fraction of the total uncertainty that is due to the different 4 main uncertainty sources, we linearly weight the uncertainty error bar as follows:

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

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

In this section we summarize the structure and origin of different pieces of data used in this work. A scheme on how these data are included in the model can be found in Figure 3.1D

- Annual rates of incidence and mortality from 2000 to 2015 for the different countries studied. WHO TB burden estimates database.[149] These data are used to calibrate, for each country independently, the scaled infectiousness ($\beta_0$, $\beta_1$), diagnosis rate ($d_0$,$d_1$) and initial distance to stationarity $\varsigma$.

- Treatment outcomes for the different countries. WHO TB treatment outcomes database.[149] Used to extract the parameters: $f_D^{p+}$, $f_F^{p+}$, $f_\mu^{p+}$, $f_D^{p-}$, $f_F^{p-}$ and $f_\mu^{p-}$ (fraction of individuals experimenting the different possible treatment outcomes)

- Population of each age group and country from 2000 to 2050. UN population division database.[138] From this data we extract the demographic structures that populations are forced to follow as explained in section 3.2.5

- Age contact patterns from different experimental settings.[25,130,133,134,137?] Used to construct the contact rates between age groups (see sections 3.2.2 and 3.2.3).

- Different bibliographical sources,[55,62,108,154,156] from which we extract values for 19 epidemiological parameters (see tables 3.7 and 3.8 ).

## 3.3 Results

### 3.3.1 Baseline forecasts of TB incidence and mortality

To illustrate the ability of the method to reproduce current epidemic trends in different scenarios, the model was applied to describe the TB epidemics in India, Indonesia, Nigeria and Ethiopia (Figure 3.5). These countries, which accumulate as much as ∼40% of the total TB burden world-wide in 2015, were selected because of their different temporal evolution trends, current and prospected demographic profiles and geographic locations. Remarkably, the model does not predict, in general, a sustained decrease in TB burden for the decades to come in these cases, whose incidence rates (per million habitants and year) range between 1246 (524-2124 95% C.I.) (Ethiopia) and 3669 (2348-5247 95% C.I.) (Indonesia), in 2050.

Regarding confidence intervals, colored areas in Figure 3.5 quantify the contribution to global uncertainty that stem from the different types of input data processed by the model. This includes epidemiological parameters (purple), demographic data (orange), contact patterns (green), and, most

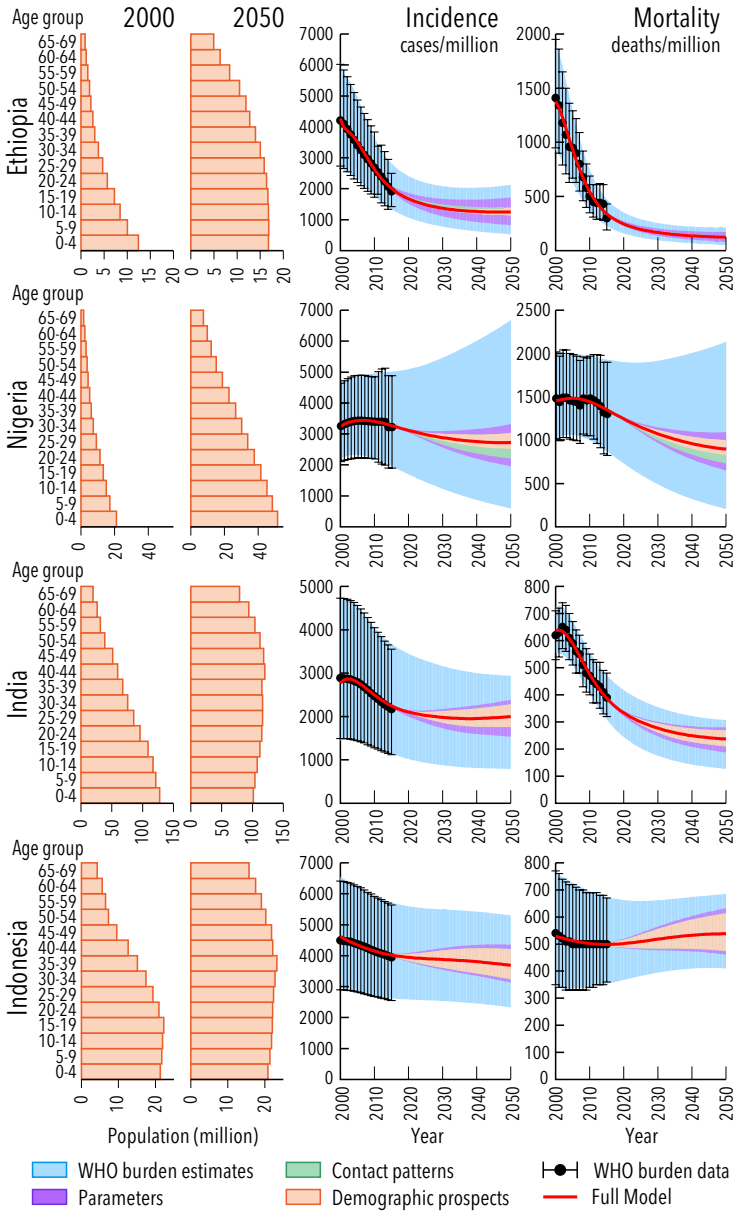


Figure 3.5: **Fitting and prediction in Ethiopia, Nigeria, India and Indonesia.** Population structure at 2000 and 2050 (projection), and annual incidence and mortality rates predicted by the model in 2000-2050. Coloured areas represent 95% confidence intervals. The contribution to overall uncertainty that stems from each of the four types of input data is disclosed. (Contributions are cumulative).

importantly, WHO-burden estimations (blue). Complementarily, in section 3.3.6, the individual contribution of each epidemiological parameter is further disclosed in an exhaustive sensitivity analysis. Of all the different individual sources of uncertainty that are susceptible to impact model's forecasts, current WHO estimates for TB burden levels is the only one that introduces more than a 15% deviation with respect to central estimates (the uncertainties in total number of TB cases prospected in 2000-2050 that are propagated from WHO-data span from 36% (Ethiopia, lower limit) to 92% (Nigeria, upper limit) with respect to central expectations).

### 3.3.2 Effects of populations aging on aggregated TB forecasts

All countries analyzed in this work are experiencing population aging to some extent, consistent with the overall trend that is forecasted for global human populations during the same period.[138] The four countries selected in Figure 3.5 lie at different points of the demographic transition by the beginning of the period under analysis (year 2000), and are expected to evolve at different paces into more or less aged populations by 2050.

In order to isolate the influence of populations aging on model outcomes, we compared the model with a simplified version where demography evolution is neglected as done in previous approaches[55,101,108] (reduced model 1, RM1). In this reduced model the demographic structures are taken from their initial configuration in 2000 and remain static until 2050. The results shows that the demographic evolution leads to a systematic and significant increase in the prospected incidence rates, which is variable in size across countries (Figure 3.6A: relative increase in incidence in 2050: full vs reduced model 1: India: 39.6% (13.9-63.6 95% C.I.), Indonesia 23.4% (7.9-36.5 95% C.I.), Ethiopia 56.0% (29.2-62.1 95% C.I.), Nigeria 34.5 % (9.1-42.9 95% C.I), see also Figure 3.9, 3.11 and table 3.13 for equivalent results in other countries). Furthermore, the relative variation between incidence forecasts obtained from the full and the reduced model by 2050 significantly correlates with the intensity of the aging shift, as given by the change in the fraction of adults (age>15 years) in 2000-2050 (Figure 3.6B, Pearson correlation r=0.66, p=0.02). This is indeed a natural consequence, since adults are burdened with higher incidence rates than children, and thus, populations' aging implies a relative increase of the demographic strata that is most affected by the disease (adults), in detriment of children, among whom TB incidence is lower (Figure 3.6C).

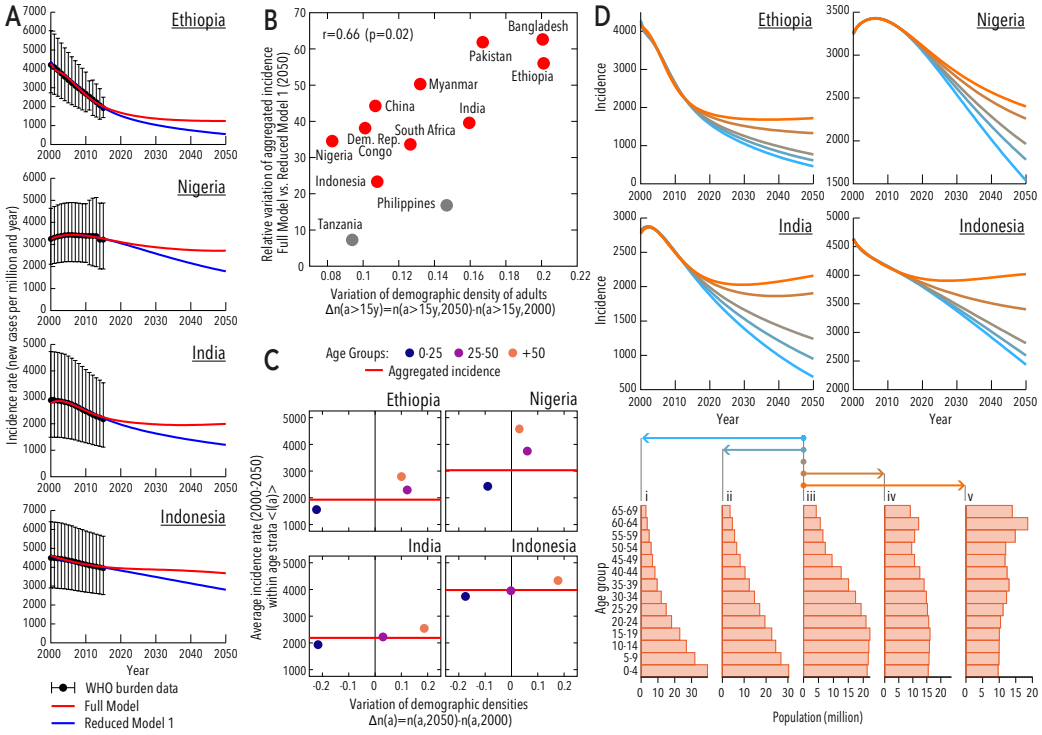


Figure 3.6: **Effects of demographic dynamics on model forecasts.** (A): Incidence rates from 2000 to 2050 obtained from the full model (red) and reduced model 1 (constant demography, blue). (B): Relative variation of aggregated incidence at 2050 for the top 12 countries with highest absolute TB burden in 2015 versus variation in the fraction of adults in the population during the period 2000-2050. In all countries but Tanzania and Philippines, in grey in the figure, the variations in incidence are significant at a nominal p=0.05. (C): Age specific average incidence rate of TB vs. variation of age-strata population density in 2000-2050. (D): Incidence projections for synthetic scenarios of demographic evolution. In each panel, the demographic evolution of each country is substituted by these synthetic scenarios: demographic transitions that go from stage iii in 2000 to different ending points in 2050. Colors in incidence series correspond to those in the arrows below, indicating the respective demographic transitions. Model calibration is repeated in each case.

Next, we built a series of synthetic demographic evolutions to simulate different scenarios (Figure 3.6D). To this end, we used three pivotal examples extracted from actual cases of populations featuring young, triangular demographic pyramids (Figure 3.6D, stage i, extracted from Ethiopia in 2000), aged, inverted-pyramids (stage v, extracted from China, 2050); as well as intermediate situations (stage iii, extracted from Indonesia, 2000, and stages ii and iv, built upon linear interpolation). Making use of these pivotal populations, we built synthetic transitions among them occurring in the period 2000-2050, which we then integrate in the TB model, in the four countries analyzed, instead of their own real demographic projections. As we can see in Figure 3.6D, population aging appears associated to increased incidence rates, while eventual transitions towards younger populations would cause incidence forecasts to decline faster.

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

### 3.3.3 Effects of aging on age-specific burden levels

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

In Figure 3.7, we show, for one example per continent –India and Ethiopia–, the infection matrices between age groups as described by each model, and their difference. The entry $(a, a')$ of these matrices represents the prospected number of infections (in 2050) from age-group $a$ (infection source) to $a'$ (infection target) per year per million people in group $a'$. For both countries, the differences between full and reduced model 1 point to a systematic underestimation of the number of infection events caused by adults as a consequence of ignoring demographic dynamics, as well as an over-estimation –only appreciable in India–, of infections caused by children during the period under analysis. Furthermore, once contagions are aggregated across infection sources within each target age-group (Figure 3.7, age-specific infection rates histograms, built as column-wise marginal sums of the infection matrices), significant differences between age-specific infection rates arise in both countries, mainly in adult age strata, where the full model predicts systematically larger incidence levels than the reduced model 1.

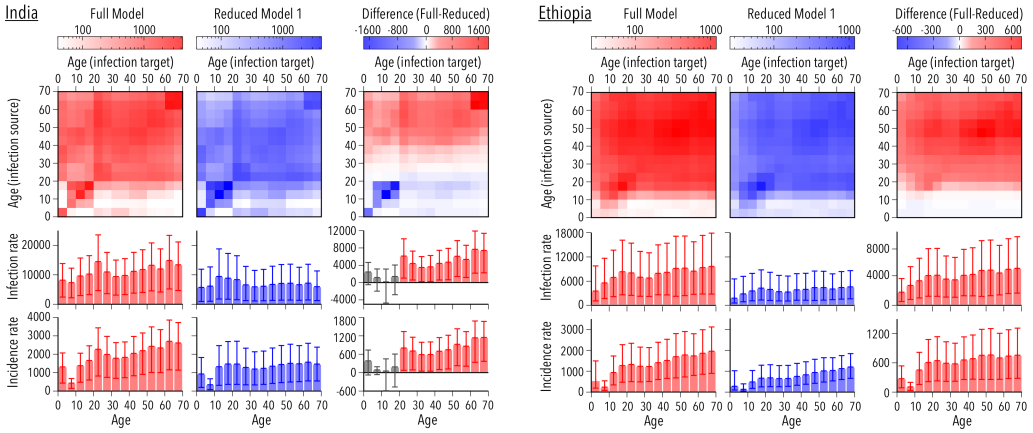


Figure 3.7: **Effect of demographic aging on age-specific TB burden.** Age-to-age infection rate matrices (number of infections from age group a to age group a' per year per million people in target age-group a'); and age-specific infection and incidence rates forecasted in 2050 for India and Ethiopia (number of contagions, or new active TB cases, respectively) per year and million individuals in a give age-group) for full model and reduced model 1. In the left column, the forecasts derive from the full model, and, in the central column, from reduced model 1 (constant demography). In the right column we represent the difference (full - reduced model 1) of these three observables: infection matrices, age-specific infection rates and age-specific incidence rates. Differences in incidence and infection rates are shown in grey when they are not statistically significant.

This ultimately translates into an increase in age-specific incidence rates of active TB cases (Figure 3.7, age-specific incidence histograms), which can be easily interpreted attending to the larger probabilities of developing the most infectious forms of pulmonary TB that adults experiment with respect to children[55]. Adults, whose proportion increases in the system as a result of considering populations' aging, not only constitute the part of the demographic pyramid most hit by the disease, but also the one that contributes the most to overall spreading. Therefore, including populations' aging on model dynamics causes not just an increase in the aggregated burden levels across all age-groups, but also within age-strata.

### 3.3.4 Effect of contact patterns heterogeneities

After discussing the impact that demographic dynamics has on model outcomes, we inspected what are the effects of including contact patterns in the TB model forecasts, either at the level of aggregated rates or within age-specific strata. To do so, we have built a second reduced model where the empirical contact matrices estimated from survey studies conducted in Africa and Asia[130,133,134,136,137] are substituted by the classical hypothesis of contacts homogeneity (reduced model 2).

In Figure 3.8, we represent the infection matrices that derive from the full and the reduced model 2 for India and Ethiopia in 2050.

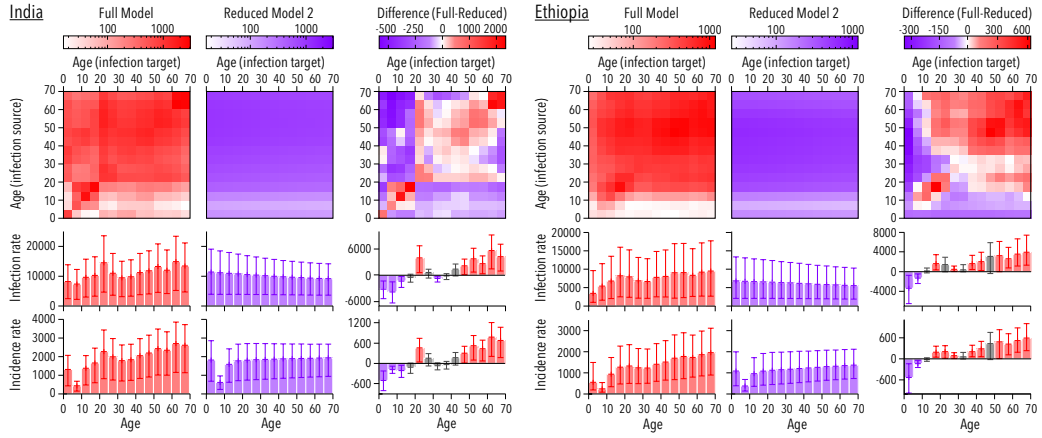


Figure 3.8: **Effect of heterogeneous contact patterns on age-specific TB burden.** Age-to-age infection rate matrices, and age-specific infection and incidence rates forecasted in 2050 for India and Ethiopia, from full (left) and reduced model 2 (center); with the difference (full - reduced model 2) in the right column in each case. Differences in incidence and infection rates are shown in grey when they are not statistically significant, otherwise they have the color associated to the model for which the rate is higher.

Clearly, empirical contact patterns reshape the distribution of contagions among age-groups, giving a larger importance to assortative infections that take place among individuals of similar ages –specially between adolescents and young adults–, while penalizing infections from children to adults, or vice-versa. As a result, in this case, the infection and incidence rates of TB among children are higher in the reduced model, while the full model predicts more infection and disease burden among adults, with slight variations

between the two countries that are due to the different contact data used in each case in the full model. In all the countries analyzed, the opposite directions of the differences between full and reduced model that are found in children versus adults tend to compensate each other. This results into similar global incidence rates produced by both models (see section 3.3.5, fig 3.12).

Summarizing this part, and despite the reduced effect observed on aggregated rates, we showed that including empirical contact structures on TB model dynamics reshapes the transmission patterns among age groups, and generates significant differences in age specific infection and incidence rates.

### 3.3.5 Complementary results and further analyses

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

In previous sections we have performed the majority of the analyses in India, Indonesia, Nigeria and Ethiopia, which were selected for their different initial age-structures, TB burden trends and demographic shifts. Now we extend the analyses to the 12 countries affected by highest absolute TB burden levels in 2015 (values of fitting rest in table 3.12 and predictions of incidence and mortality in figure 3.9).

In all these countries, except China and Philippines, the rest represents the residual sum of squares normalized by the input data uncertainty of each measure. In the last two examples, where confidence interval limits coincide with central estimates in some cases, we normalized the squared residuals by the average incidence and mortality rates, respectively (see section 3.2.8 for further details). In Figure 3.9 we see that the model is able to reproduce satisfactorily the observed trends in 2000-2015 in all cases.

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

From Figure 3.9 we see that, in general, incidence and mortality forecasts from the reduced model 1 are systematically lower than those obtained when the demography's dynamics is accounted for.

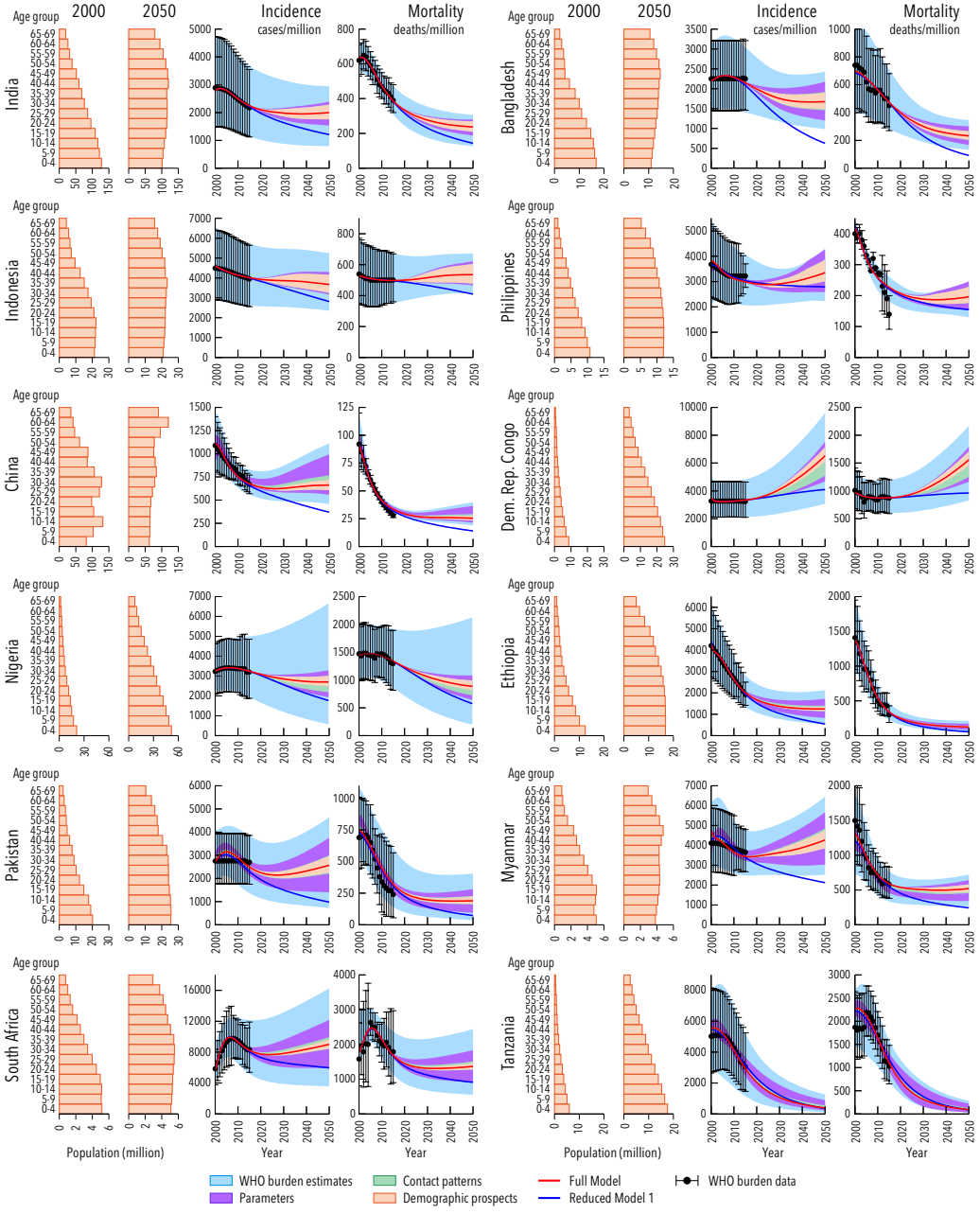


Figure 3.9: **Incidence and Mortality for top 12 countries with most TB cases** Demographic pyramids in 2000 and 2050, and incidence and mortality projections (2000-2050) for the 12 countries with the largest number of incident cases in 2015.

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

Table 3.12: **Values of fitting error $\boldsymbol{H}$.** Values of the rest $H$ and relative error for the fit (averaged in the fitting window) for the countries fitted in this work. For the ten first countries, the rest is defined as the sum of the squared deviations between model and data, normalized by the uncertainty of input data in each case (see equation 3.44). In China and Philippines (*, see equation 3.47) the normalization factors are set up as the average incidence and mortality during the training period, thus their $H$ values are not in the same scale of the rest of the countries. Further details are provided in section 3.2.8

#### 3.3.5.2 Case Notifications

The Case Notification Ratio (CNR), i.e., the fraction of all incident cases that a particular Health System detects and notifies to the WHO-surveillance systems each year, is a fundamental magnitude in the surveillance and control of TB. As a means of validating the model and its calibration procedure, we interrogated whether CNR values reported by the WHO in each country correlate to model-based Treatment coverage ratios, (defined as the fraction of incident cases per year that get diagnosed) despite the fact that these magnitudes are not considered or compared during the calibration step. In figure 3.10 we represent the results of this comparison in 2015, where we see that model-based treatment coverage fractions are strikingly correlated to CNR values reported by WHO across countries (Pearson correlation excluding Indonesia: $r = 0.96$, $p = 4.3e - 6$), which reinforces the validity of the model.

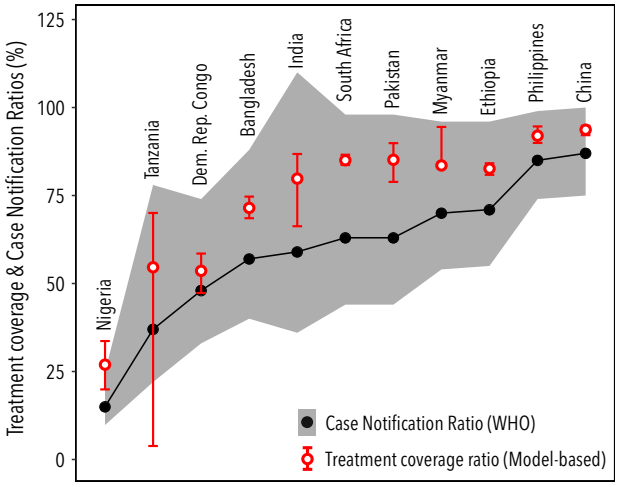


Figure 3.10: **Treatment coverage and Case Notification Rates.** Treatment coverage and associated uncertainty obtained with the model (red symbols) and Case Notification Rates, with their correspondent Confidence Intervals, extracted from WHO data (grey) in 2015.

Importantly, we find that treatment coverage is slightly higher than the CNR in every country, which can be interpreted in terms of under-reporting of diagnosed cases. The CNR and model-based Treatment Coverage ratios are closely related, but not fully equivalent, since a fraction of the total TB cases that a country detects and treats each year goes undetected to the WHO surveillance systems, despite TB notification being mandatory in most of the countries analyzed. An exception is found in Indonesia, not included in the figure, where the model predicts a Treatment coverage that is much higher than the CNR (Treatment coverage: 81.5% (CI: 78.3-83.7), CNR: 33% (CI: 23-50)). Precisely in Indonesia, previous studies have pointed out the presence of significant levels of TB under-reporting to the WHO surveillance systems[162], partly related to the fact that in this country the notification of TB cases is not mandatory.

#### 3.3.5.3 Relative Differences between Full Model and Reduced Model 1

We have compared the incidence and mortality rates predicted by the full model, (i.e., considering demographic changes), and the reduced model 1 (for which the demographic pyramid is considered constant in time), but it remains pendant to stablish whether the difference between forecasts from the full and the reduced model are statistically significant or if, instead, their uncertainty may be larger than its magnitude. To shed light on this question, in figure 3.11 we represent, for the twelve countries analyzed, the time evolution of the relative differences between the incidence rates that each model produces, alongside the correspondent uncertainty intervals, obtained as detailed in section 3.2.10.1.

As we can see in Figure 3.11, differences between the full model and the reduced model 1 become significant in most countries, often right after the end of the fitting window (2015). In Table 3.13 we register the relative difference between the incidence rates in 2050 as derived from each model, alongside the corresponding 95% confidence intervals and the significance levels. By 2050, differences between models become statistically significant (at 99% significance level) for all countries except two: Philippines and Tanzania.

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

Table 3.13: **Relative difference in the incidence rate in 2050 between full and reduced model 1** for the 12 countries with more TB cases in 2015. Significance levels: −: not significant, ∗:95%, ∗∗:99%, ∗ ∗ ∗:99.9%.

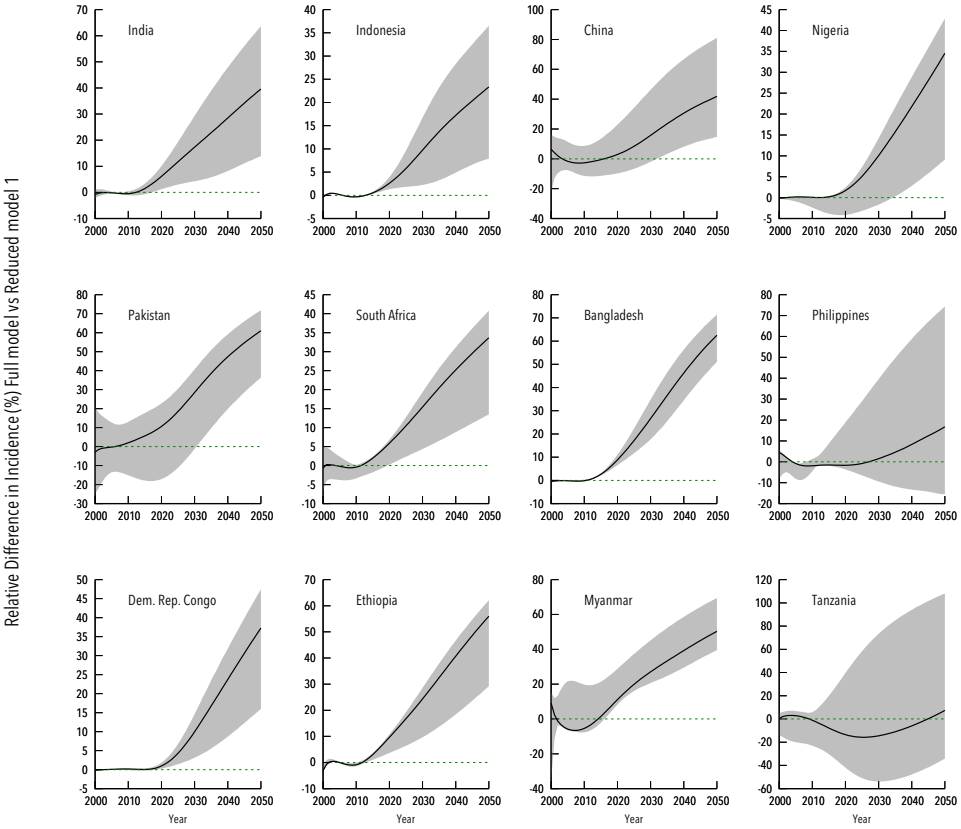


Figure 3.11: **Relative differences between the full model and the reduced model 1**, in terms of the incidence rate in the 12 countries considered.

#### 3.3.5.4 Effect of Contact Patterns at the aggregated level

Previously, we have shown that the assumption of homogeneous mixing overestimates the burden of TB in children, and underestimates it among adults. These opposite effects largely cancel each other, which makes the total effect to shrink when considering the aggregated burden across ages. In figure 3.12 we represent the forecasts of incidence and mortality for the models with heterogeneous and homogeneous mixing patterns (full vs. reduced model 2).

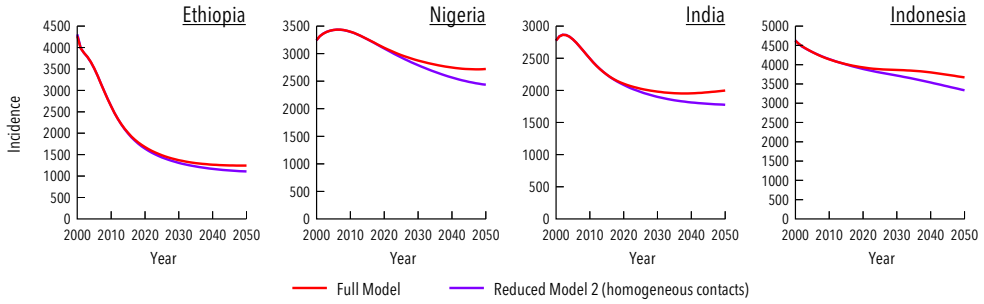


Figure 3.12: **Effect of contact patterns at the aggregated level.** Predictions of incidence and mortality rates for full (red) versus reduced model 2 (violet) in Ethiopia, Nigeria, India and Indonesia

As detailed in Table 3.14, the usage of empiric contact patterns translates into slightly larger burden rates (relative differences around 10% between the full and the reduced model 2 in 2050). These modest differences are still significant in Ethiopia, India and Indonesia, despite their relatively small values when compared to the magnitude of forecasts' uncertainty. This is not an anomalous behavior, since outcomes from the full and the reduced model are strongly dependent variables, and the uncertainty is propagated to them in a paired fashion (i.e., the same sources of uncertainty affect both models simultaneously), which is considered when comparing results from model pairs. This allows us to detect significance differences between model behaviors of lower effect sizes than the characteristic uncertainty of each independent model alone.

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

Table 3.14: **Relative difference in the incidence rate in 2050 between full and reduced model 2** in Ethiopia, Nigeria, India and Indonesia. Significance levels: $-$: not significant, $*$:95%, $**$:99%, $***$:99.9%.

#### 3.3.5.5 Effect of Contact Patterns on TB burden distribution across age

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

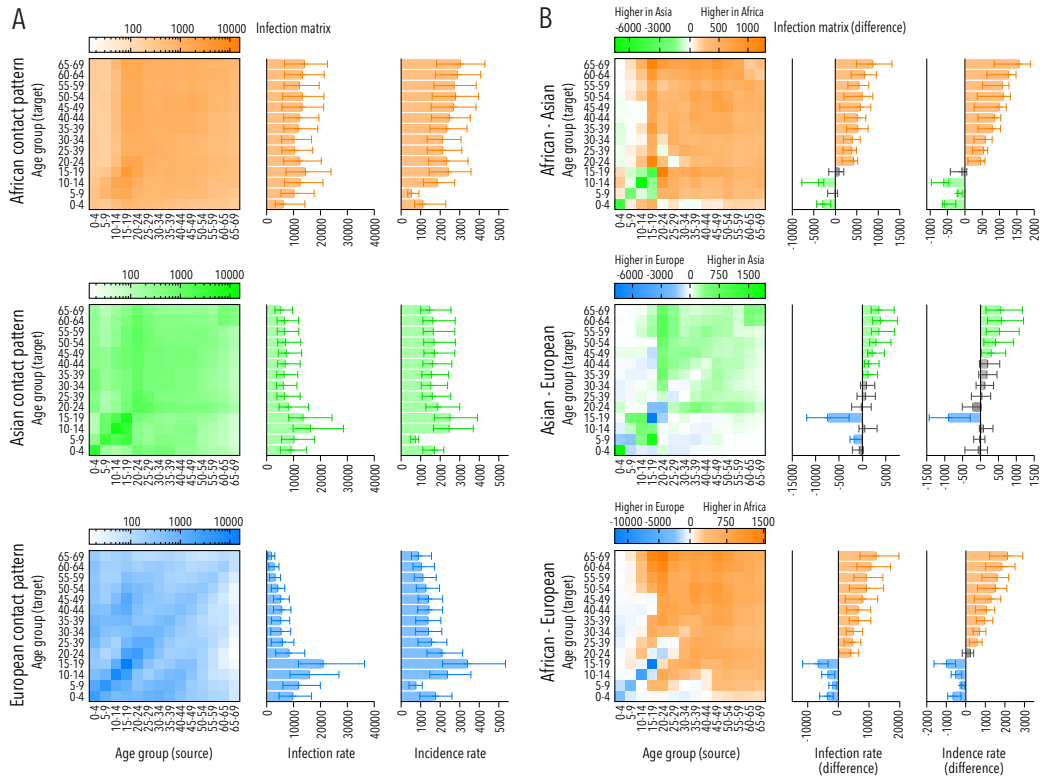


Figure 3.13: **Effect of contact patterns switching on TB burden distribution across age** (A): Matrix of infection rates, and age-specific rates of infection and incidence (averaged during the period 2000-2050) corresponding to the use of different contact patterns (African, Asian and European) in Ethiopia. (B): Pairwise differences. Non-significant differences in infection and incidence rates are represented in grey, otherwise they are coloured as the predominant contact pattern.

To answer this question in one particular example, we have chosen one of the countries analyzed, and generated forecasts based on simulations performed using contacts derived from African surveys data (i.e. the default), as well as Asian, and European contact structures.

The results of these tests, where the average incidence rate during the period (2000-2050) is reported for each case, are represented in figure 3.13. In the figure, upon pair-wise comparison between the forecasts associated to each contact matrix, we see that African contacts lead to higher burden among the eldest age-groups, while the European contact patterns tend to induce more infections and TB cases among younger individuals and the matrix used in Asian countries represents an intermediate situation.

### 3.3.6 Sensitivity analysis

In figure 3.5, as in figure 3.9, the contribution to overall uncertainty of each type of input data is disclosed: epidemiological parameters (purple), contact matrices (green), demographic prospects (orange) and WHO burden estimates (blue). In figure 3.14, the contribution of the uncertainty derived from each single epidemiological parameter is further shown.

We see that WHO burden estimates are responsible for most of the uncertainty holding relative differences with the mean value when the limits of the confidence intervals are considered almost reaching 100% in some particular settings (Nigeria, upper bound). The second source of uncertainty that produces more variability in these forecasts corresponds to the demographic projections, with a relative difference with the mean value between 5-10% when the upper or lower limits of the demographic projections are considered. This result is coherent with what we have discussed earlier: the inclusion of demographic evolution significantly changes the predicted incidence and mortality rates. As for the rest of uncertainty sources (epidemiological parameters and contact matrices), their influence in the uncertainty are significantly smaller. Only in some exceptional cases we find parameters producing a relative difference with the mean value surpassing 5%, and in most occasions these values are even smaller than 1%.

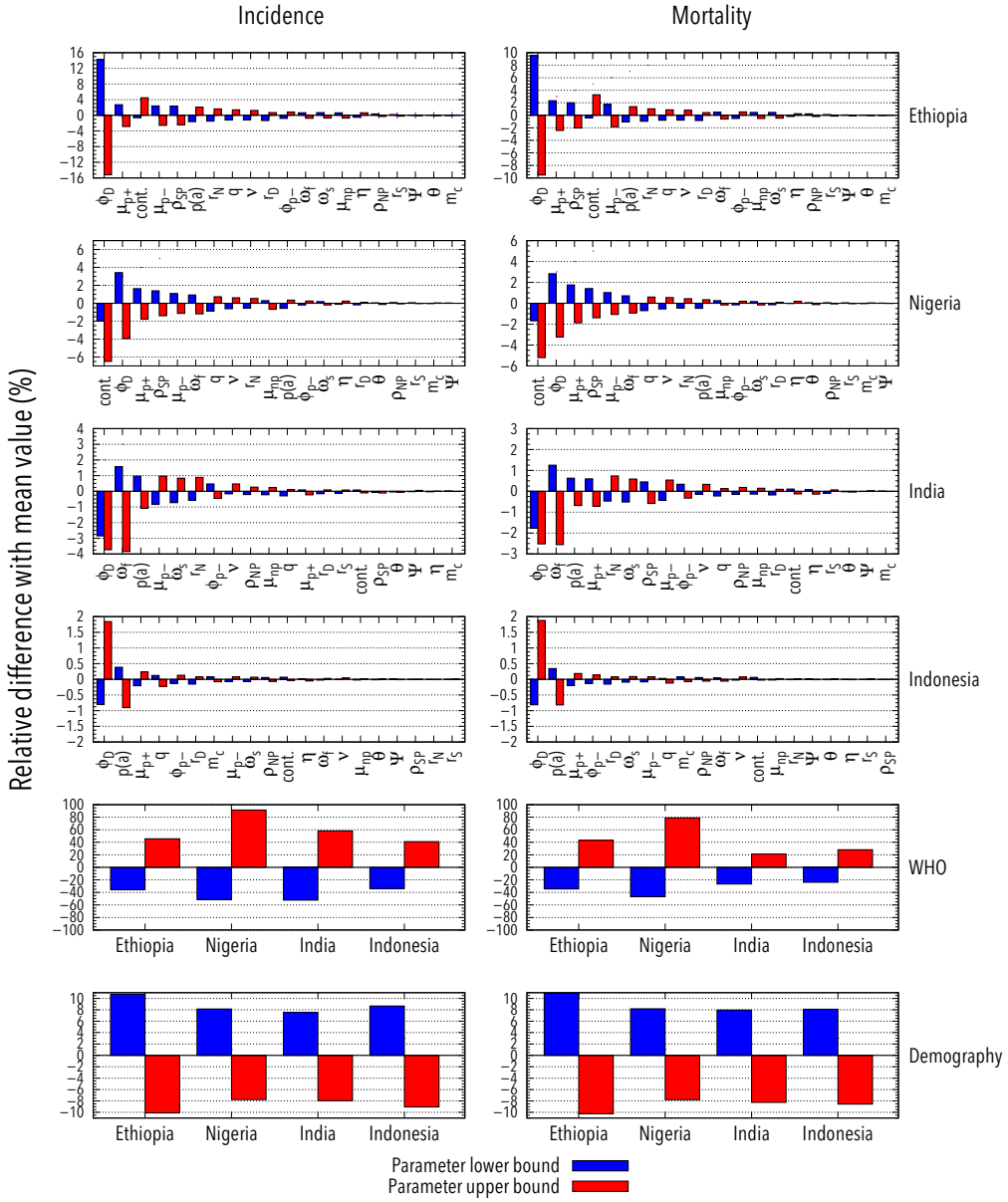


Figure 3.14: **Sensitivity Analysis** in India, Indonesia, Nigeria and Ethiopia. Red (blue) bars represent the variations in total number of TB cases/deaths that the model produces in 2000-2050 after increasing (decreasing) the value of each uncertainty source to the upper (lower) limit of its respective confidence intervals, prior to model calibration. The bottom panels contain the sensitivities associated to the WHO burden estimates and the demographic projections.

### 3.3.7 Robustness tests

#### 3.3.7.1 Effect of demographic evolution

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

**Different burden levels.** All the forecasts produced in this work are based on estimates of incidence and mortality reported by the World Health Organization. These estimates are based on a combination of epidemiological observations and empiric criteria that are known to be affected by high levels of uncertainty, as reflected by the large confidence intervals that these data present in many countries, which has been taken into account and incorporated to model forecasts. However, the broad nature of these estimates often leads to re-evaluation of methods and values reported by the WHO in their periodic TB reports, with the result that, in some cases, the burden estimates vary beyond the range of uncertainty initially assumed as a result of these methodological updates. For example, the burden estimates that we use for India turned out to be underestimated in previous publications. [71]

Taking it into account, the question of whether the effects of demographic evolution are robust under wide variations in the input burden data constitutes a valid concern. To show that those effects are indeed robust under a wide range of initial burden levels, in figure 3.15 we have repeated the simulations for the full and reduced model 1 under alternative scenarios where initial burden levels have been doubled/halved. Checking the relative difference between models in the incidence rate at 2050 (table 3.15), we see that it remains statistically significant ($p<0.05$) in every case. Thus, even if the data of incidence and mortality are not extremely reliable, and could be proven to be biased in the future, the need to improve current models and incorporate the evolution of demography holds as a general conclusion valid for a wide range of initial burden levels.

**Different Contact Patterns.** During this work, we have implemented two different contact matrices: one adapted from statistical surveys conducted in Africa, and another one from Asian studies. This supposes an arguable improvement with respect to the assumption of contacts homogeneity, and even with respect to using the same contact structure in all countries. However,

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

Table 3.15: **Relative differences for incidence rates in 2050 between full and reduced model 1, as obtained from biased TB burden estimations** (WHO estimates doubled/halved with respect to reported values). Significance levels: $-$: not significant, $*$:95%, $**$:99%, $***$:99.9%.

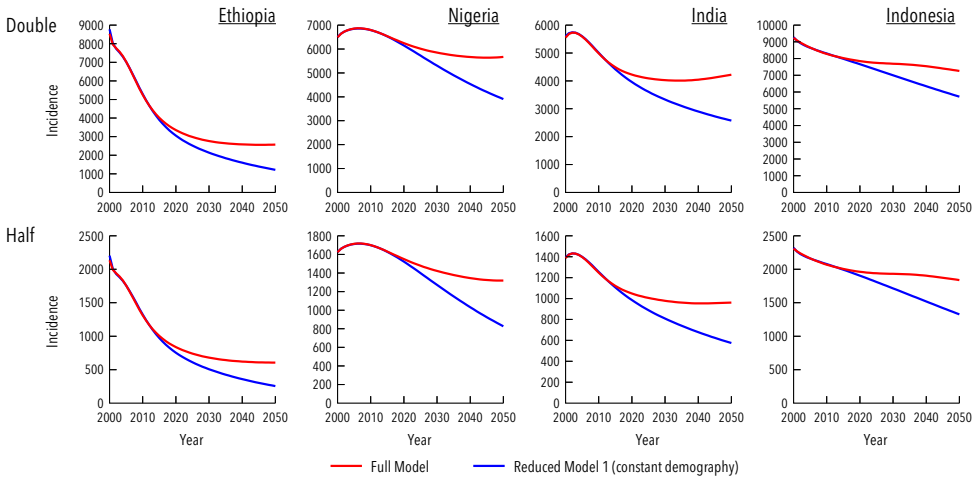


Figure 3.15: **Robustness against biased input burden estimates.** Incidence rate forecasts from full (red) and reduced model 1 (blue) under biased input burden estimates (WHO estimates doubled/halved with respect to reported values).

the available bibliography on empiric contact patterns is still reduced, and we are far away from an ideal situation where empiric data of comparable quality standards is at hand at a country-specific level. In this sense, the question of whether the effects of demography evolution might be dependent

or not of the specific contact structure that we use in each country is pertinent, and should be explicitly addressed.

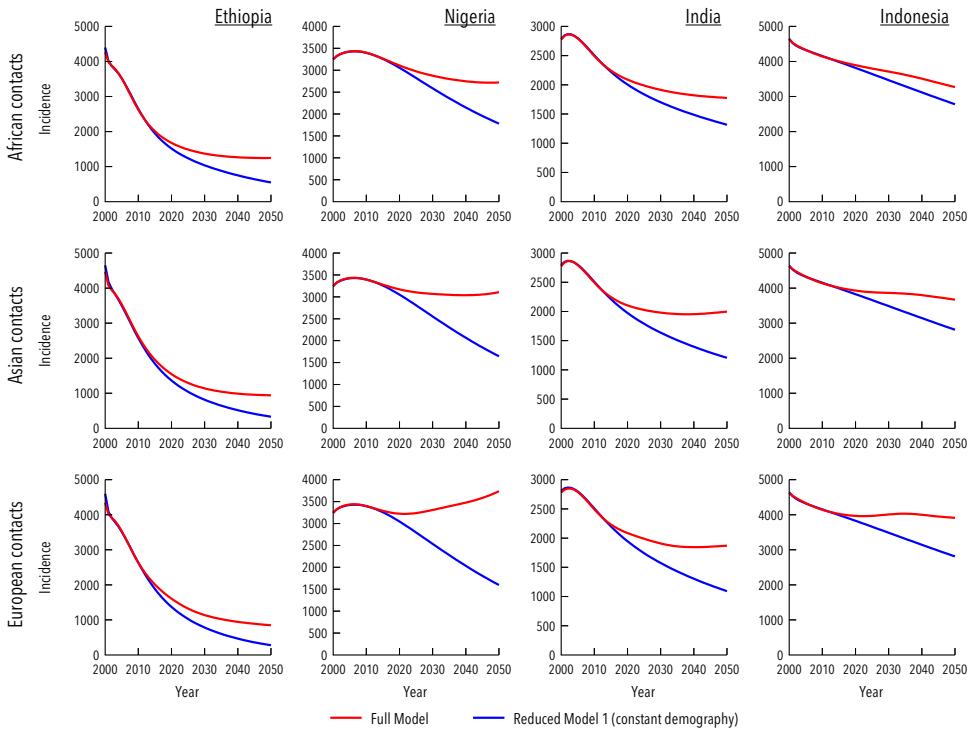


Figure 3.16: **Robustness against different contact patterns.** Incidence Rates forecasted from full (red) vs reduced model 1 (blue), for three different contact patterns (African, Asian and European).

To this end, in figure 3.16, we show epidemic forecasts derived from the full and the reduced model 1, under scenarios where the contact structure of each country has been substituted by the other matrices considered across the paper, including a contact matrix built from the European Polymod study.

Once again, we have obtained the same result: considering the evolution of demography leads to higher burden prospects, independently of the contact pattern used in the simulations. In table 3.16 we show the relative differences between models in the incidence rate in 2050 and the associated significance levels.

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

Table 3.16: **Relative differences for incidence rates in 2050 between full and reduced model 1, evaluated using different contact matrices.** Significance levels: $-$: not significant, $*$:95%, $**$:99%, $***$:99.9%.

**Different evolution of fitted parameters.** The model forecasts presented in this work are produced under the hypothesis that, after the training period (2000-2015), the time evolution of scaled infectiousness and diagnoses rates will still be governed by the sigmoid curves described by equations 3.42 and 3.43. However, if the pace of variation of these parameters slows down in future years from the expected trends, the TB burden rates will increase from the forecasts reported. To explore the behavior of the model in that situation, and re-evaluate the difference between full and reduced model 1, we have repeated the comparison in alternative scenarios where the pace of variation in the fitted parameters is slowed down from 2015 a 50% and a 100% from its expected trend.

The underestimation of TB burden that stems from ignoring demographic evolution is also robust against variations in the time-evolution of fitted parameters after 2015. In table 3.17 we show the relative difference of the incidence rates in 2050, and we check again that these differences are significant for all alternative scenarios.

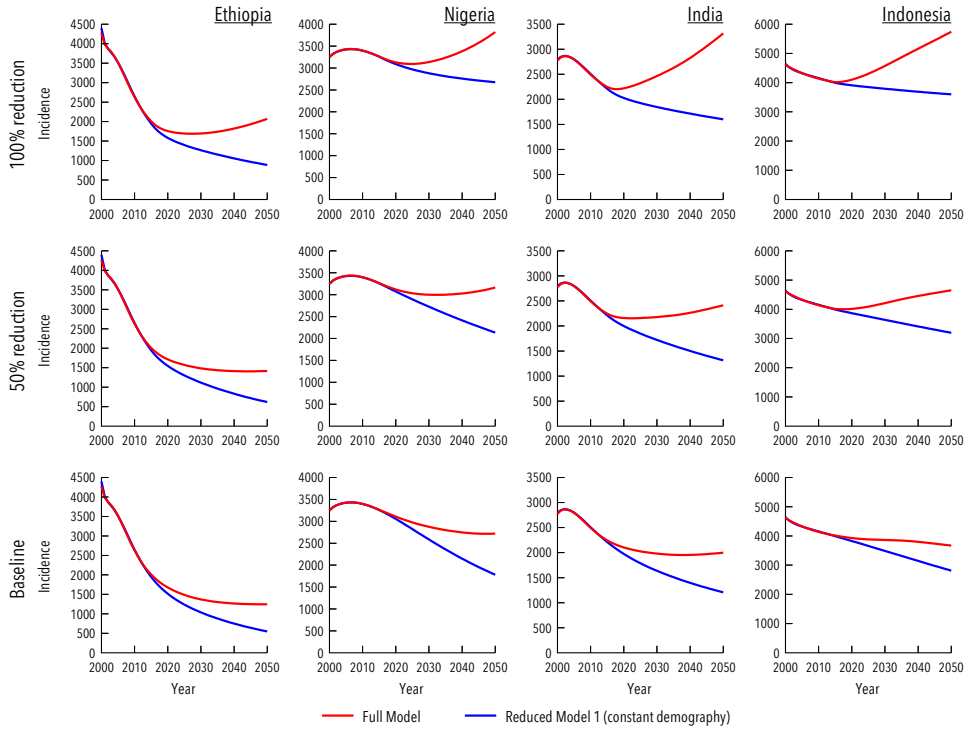


Figure 3.17: **Robustness against alternative variation rates of the fitted parameters.** Incidence Rates obtained from the full and the reduced model 1, for 2 alternative scenarios where variation rates of the fitted parameters is reduced a 50% and a 100% from their expected behaviour from 2015 on.

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

Table 3.17: **Relative differences for incidence rates in 2050 between full and reduced model 1, evaluated applying different reductions on the variation rate of the fitted parameters after 2015.** Significance levels: $-$: not significant, $*$:95%, $**$:99%, $***$:99.9%.

**Effect of Demographic Evolution without re-fitting parameters.** Comparisons between the full and -for example- the reduced model 1 across the text are performed upon independent calibration of each model, to ensure that both reproduce the initial burden trends independently. However, this procedure does not guarantee that the differences between models arise from the demographic dynamics itself, since they might be a consequence of the different parameters and initial conditions that are estimated in each case. To rule out this possibility, we present here a series of simulations where we use the values for fitted parameters and initial conditions that were estimated upon calibration of the full model, also in the reduced model 1.

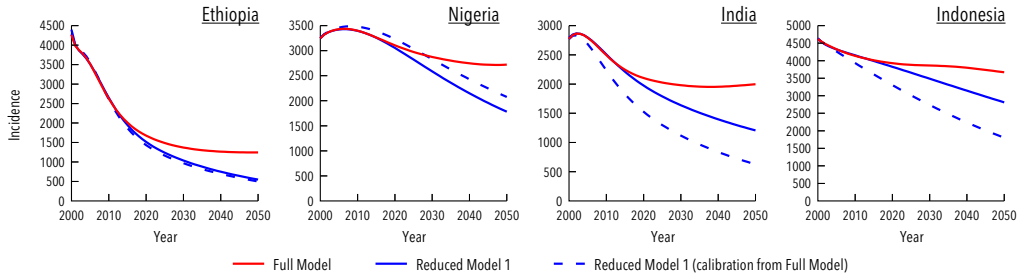


Figure 3.18: **Effect of Demographic Evolution without re-fitting parameters.** Incidence rate series produced by the full model (red), the reduced model 1 (blue, continuous), and a non-calibrated version of the reduced model 1 which makes use of the same parameters and initial conditions fitted for the full model (blue, dashed line).

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

Table 3.18: **Relative difference in the incidence rate in 2050 between full and reduced model 1 without re-fitting the model** The reduced model 1 is executed using the same initial conditions and fitted parameters inferred upon full model calibration (red minus dashed blue lines in figure 3.18). Significance levels: $-$: not significant, $*$:95%, $**$:99%, $***$:99.9%.

In Figure 3.18 and Table 3.18, we see that the reduced model 1 still reproduces significantly lower burden projections than the full model even though it is not calibrated to reproduce the data before 2015. Note that the difference with respect to the full model is in many cases amplified.

### 3.3.7.2 Effects of contact patterns

We also checked the robustness of considering contacts heterogeneity on the age distribution of TB burden in a wide spectrum of different modeling scenarios, as we did for the evolution of demography.

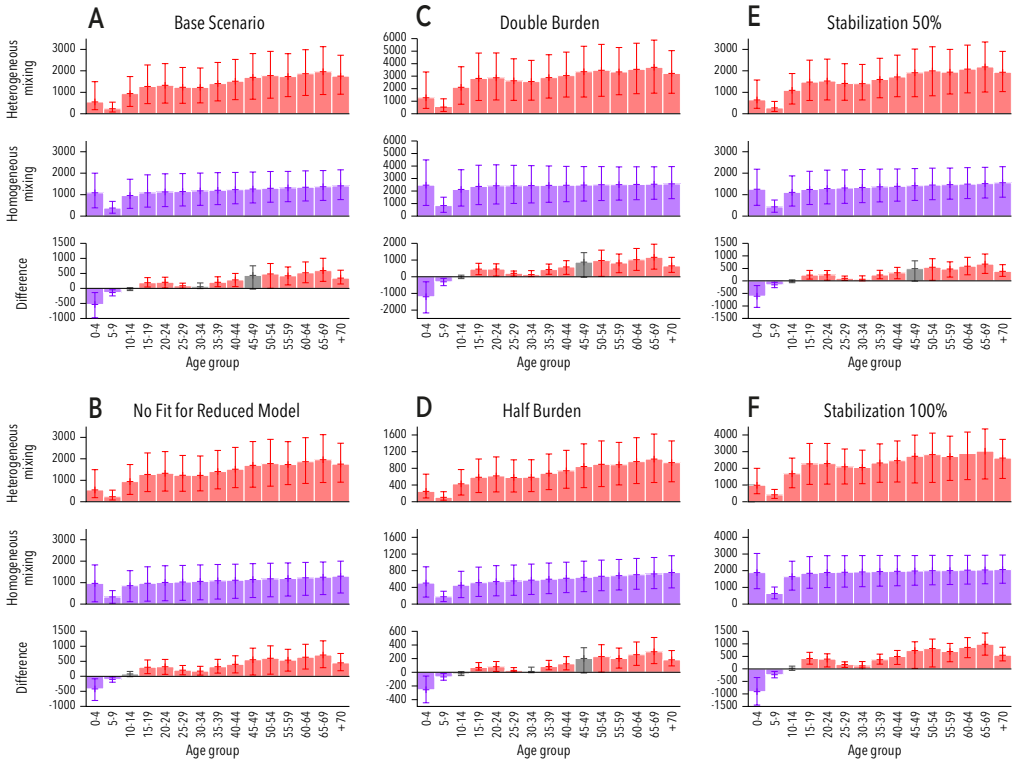


Figure 3.19: **Robustness tests for contact heterogeneity.** Age specific incidence rates (2050, Ethiopia), for the complete model, the reduced model 2 and the difference between them for different scenarios (same as in section 3.3.7.1). Non-significant differences in infection and incidence rates are represented in grey, otherwise they are colored as the predominant model.

In figure 3.19 we show the age distribution of incidence in 2050 for

Ethiopia in 6 scenarios: (A) the base scenario, (B) the case where the reduced model is not recalibrated; (C) a scenario where TB burden data is doubled, (D) a scenario where TB burden data is halved; (E) a scenario where time evolution of fitted parameters is reduced a 50% from 2015, and (F) a scenario where these variation rates are totally arrested from 2015 on. In all these different settings, the assumption of homogeneous mixing between age groups implies the emergence of significant differences between age-specific incidence rates with respect to the hypothesis of homogeneous mixing.

## 3.4 Conclussions

The model presented here was specifically designed to provide a suitable description of TB transmission dynamics in situations where demography is evolving at the same time that the epidemics unfolds. Importantly, we showed that considering current populations' aging trends in TB transmission models is followed by a systematic increase in burden forecasts. This worrisome result can be understood in terms of the known mechanisms whereby age affects the transmission dynamics of the disease. In TB, adults are affected by higher age-specific burden levels than children, and, at the same time, they are more efficient spreaders, given their increased tendency to develop infectious forms of pulmonary TB.[55] As a consequence, considering populations' aging translates in higher burden forecasts, simply by increasing the fraction of older individuals with respect to children.

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

Furthermore, the model incorporates a data-driven description of the dependency of TB transmission routes with age, which, we have showed, exerts a significant influence on the forecasted age-distribution of the disease burden, by reshaping infection routes. These results will impact the evaluation and comparison of novel epidemiological interventions, mostly if they are conceived to target specific age-strata, as is the case of new preventive vaccines. In this context, previous works have concluded that a quick immunization

of young adults through vaccination campaigns focused on adolescents is expected to produce a faster decline in TB incidence than an alternative strategy based on newborns' vaccination.[101] These results would further reinforce this hypothesis, to the extent that empirical contact patterns are followed by a relative increase of TB among adults with respect to children. It has to be noted, though, that in the decision of what is the optimal age-group to target in a hypothetical immunization campaign for a new vaccine, at least two additional aspects have to be considered, namely, whether the new vaccine is conceived to boost or substitute BCG, and whether previous exposure to environmental antigens are expected to compromise vaccine performance, as we will explore in next chapter. BCG substitutes, and/or vaccines susceptible to lose immunogenicity due to exposure to mycobacterial antigens of individuals before vaccination might not be eligible for adolescent immunization anyways.

Despite all the improvements introduced here, this approach is not exempt from the strong limitations that affect all TB transmission models operating at this level of resolution. The outcomes of the model depend on a series of epidemiological parameters and initial burden estimates that are subject to strong sources of uncertainty. Even though we have registered these uncertainties and propagated them to the final model outcomes, future improvements and reassessment of these pieces of input data are generally expected to impact the quantitative outcomes of the model, and to further delimitate the uncertainty ranges here reported. As a first example, international Health Authorities come insisting on the importance of implementing systematic surveys of TB prevalence in many countries, as a means towards more accurate TB burden evaluations. Accordingly, they revise and update their burden estimates on a regular basis, as new data become available, which obviously impacts model calibration and results. Furthermore, the demand of epidemiological studies aimed at obtaining updated estimates of key epidemiological parameters in current epidemiological settings has been pointed out as a primary need for the development of more reliable TB models.[163] Similarly, we have seen here the importance of obtaining data of contact patterns specific for each setting, by showing that different contact structures inferred from studies conducted in different parts of the world yield to significantly different distributions of TB burden across age (Figure 3.19). Importantly, the interpretation of these burden distributions of TB across age is hindered by the limited quality of the data available regarding TB distribution across age; which makes adventurous any comparison between model and data. For example, current WHO data-structure only splits TB incidence into two major age groups (0-14 vs 15+), with alleged,

heavy under-reporting biases among children.

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

Despite those limitations, in this chapter we have showed that abandoning the simplifications of constant demography and homogeneous contacts shared by previous models of TB transmission is not just technically feasible, but it has significant effects on model outcomes. Remarkably enough, TB is not the only disease where long characteristic time-scales and strong age-dependencies concur,[164,165] which, despite the specific details of the transmission dynamics of each case, implies that similar corrections to what we have proposed here for the case of TB might be pertinent to correct bias of current epidemic models of other diseases too.

The ultimate goal is the evaluation of new interventions for TB, specifically new vaccines. This model, as has been described in this chapter is not able yet to produce impact evaluation of new vaccines as we have not implement interventions. However, this model will constitute the foundation on which we will build impact evaluations for new vaccines by the end of this thesis. But before we incorporate new vaccines to this model, we should analyze carefully the different effects that vaccines can trigger in order to construct a faithfull characterization. The next chapters will focus on this precise issue.

# Chapter 4

# Quantifying the Masking and Blocking effects on BCG

## 4.1 Introduction

There is still a lot to be known about the immunology of MTB, and clinical work and development of new vaccines is hurdled by these limitations.[115] Even the current vaccine BCG, that has been around for the last century, constitutes a mistery for immunologists due to its very variable efficacy.[76]

The efficacy of BCG is consistent in protecting infants, especially from the most severe forms of meningeal and miliary TB[76] but is limited against pulmonary forms of the disease responsible for transmission fueling the growing epidemic worldwide. Accordingly, nowadays there exist tens of different research teams worldwide developing as many novel experimental vaccine candidates designed as revaccination (boosting) strategies in BCG vaccinated individuals (adolescents or adults) or as a BCG replacement strategies at birth.[82]

BCG fails to provide consistent protection to the pulmonary forms of the disease, especially in adults,[74] who are the main contributors of overall disease spreading as shown previously. Consequently, an accurate evaluation of the BCG impact under different conditions –population susceptibility, geography, environmental exposure, etc.– is essential. Such an evaluation will allow the assessing of the efficiency of BCG as a reference vaccine and, at the same time, will provide new guidelines and methodological tools to better evaluate the potential efficacy of the newly developed TB vaccines. The highly variable and apparently inconsistent results obtained in BCG's

efficacy tests and meta-analysis have been subject of intense scientific controversy,[76,166] and the use of BCG during the 20th century has been largely argued.[167,168]

The hypothesized causes underlying the observed variability of BCG efficacy in different settings include differences between the BCG strains,[169] genetic, epi-genetic or socio-economical differences between populations, study quality, parasitic co-infections, etc.[82] In addition, multi-variate meta-analysis of BCG efficacy determination studies consistently determine that latitude is a variable showing a most prominent correlation with the performance offered by BCG,[72,167,170,171] pointing to the existence of latitude-driven mechanisms influencing it, rather than other possible explanations related, for example, to the ethnicity of the tested populations.[172] Among these possible mechanisms, the hypothesis that agglutinates a greater consensus points to the existence of a complex, latitude-dependent immunological process of environmental sensitization (ES) to mycobacterial antigens which might interfere with the observed action of BCG vaccine in different ways. The hypothesis of ES being the source of BCG efficacy variability has been backed up by different epidemiological observations.[76,173–175]

ES is thought to have its origin in the exposure of individuals either to non tuberculous mycobacteriae (NTM) –whose antigenic similarity to MTB[176,177] is able to cause cross reactivity in the human immune system[178,179]– or to the reservoir of latent infection of MTB itself (and other closely related bacteria within the MTB-complex). Additional sources of sensitization have been postulated, like certain parasitic infections.[180] The diversity among the different putative sources of ES is notorious, the relationship between their prevalence and latitude is not homogeneous, and their levels of cross reactivity are variable as well. This situation portraits a complex landscape that makes specially ventured to attribute the geographical patterns of BCG efficacy variation to a single factor, as it could be a global increase in NTM prevalence levels next to the equator,[178,179,181] which has been demonstrated to be inaccurate for some species.[182] Even though, it seems clear that overall levels of ES increase both with closeness to equator and subjects' age at the time of vaccination.

Two different mechanisms have been theorized on how this exposition to environmental antigens would affect the response of the host to a vaccine like BCG.[121] The masking hypothesis postulates that ES confers a significant protection against TB in such a way that a vaccine can barely offer an additional level of protection.[121,183] As an alternative hypothesis, it has

been suggested that ES prior to vaccination may trigger an immune response capable of blocking the assimilation of the vaccine by the host, either if it's a live-attenuated vaccine or if it's a booster. This is known as the blocking hypothesis.[121,184] These two effects have the potential to explain, to a large extent, the variability observed in the trials performed, that is, both the dependence of BCG efficacy on age at the time of vaccination—as an individual gets older its exposition to mycobacteriae increases—and its geographical variations. Finally, it is worth highlighting that masking and blocking do not exclude each other: in a scenario in which both mechanisms take place, ES would contribute at the same time to reduce disease risk of non-vaccinated individuals, and to impair vaccine assimilation of immunized ones.

Several studies have tackled the problem of evaluating ES impact on BCG vaccine efficacy from different angles. Researches in animal models have shown that environmental mycobacteria strains can interfere with BCG vaccination and with susceptibility to M. tuberculosis infection.[185–187] The influence of the effects of Masking and Blocking on measurements of efficacy has been also studied from a theoretical point of view,[172,188] even though none of these works allows a quantification of masking and blocking effects on vaccine's efficacy levels measured on clinical trials performed on humans.

BCG-REVAC trials were designed to discriminate these two effects on BCG performance when applied on individuals of dissimilar ages in the Brazilian cities of Salvador and Manaus.[166,189,190] In particular, three types of trials were conducted in the study, measuring the efficacy of as many vaccination strategies in each city: newborn vaccination, school-age vaccination and school-age revaccination. The rationale behind the election of such a design is twofold. On the one hand, replicating the experiment design in two cities of the same country located at considerably different latitudes, renders reasonable the assumption that the main source of variability at the efficacies observed is due to different levels of ES, since virtually any other plausible source of variation (i.e., vaccine preparation, strain or application protocol, ethnic diversity etc.) are absent or controlled for across the study. On the other hand, the trials design allows discriminating between blocking and masking effects, since the differences across cities of the efficacies observed for each type of trial are expected to vary depending on what of the two effects is dominant.

After the analysis of BCG-REVAC trials, Barreto et al. observed that the efficacy of the vaccine, when applied to newborns and measured later in life

did not show a strong geographic variation, which suggests that spontaneous protection related to masking should play a residual role, if any. On the contrary, when BCG was applied at school age,[166] either the first time, or as a second dose, vaccine efficacy observed was in both cases lower in Manaus than in Salvador; which in principle would be compatible with the blocking hypothesis if vaccine assimilation were more efficient in Salvador as a consequence of lower levels of ES bound to its larger distance to the equator. However, even if the design of BCG-REVAC trials allowed to qualitatively asses the greater relevance of the blocking mechanism as compared to masking, no actual quantification of these two effects and their relative role has been provided up to now. In this sense, after the work by the BCG-REVAC consortium, several questions remain unanswered, as we do not know (1) what is the relative likelihood of both hypothetical mechanisms when trying to explain the observed results of the trial, (2) how much predictive power would a full model containing both effects gain with respect to single effects scenarios (masking or blocking alone) (3) whether significantly different combinations of masking and blocking strengths could be similarly compatible with the observations derived from the trials, and very relevantly, (4) what are the intensities of blocking and masking effects, and their confidence intervals, yielding a most significant agreement with the data.

In this chapter, we introduce a family of mathematical models to interpret the results from BCG-REVAC trial under the light of masking and blocking effects, in order to contribute to answering the aforementioned questions within the limitations imposed by the reduced statistical power derived from the reduced number of trials studied. By confronting the model against the results of the BCG-REVAC studies, we are able to measure extent to which these effects are sufficient to explain the efficacies measured.[166] Furthermore, we quantify the specific masking, blocking and immunity waning effects yielding best-fitted estimates for the efficacies measured. To this end, we compared the likelihoods of three different modeling scenarios: a first model in which both effects concurrently take place, a second model only considering blocking and a third one containing only masking. Translating the trial results into quantitative estimations of blocking and masking strengths constitutes a relevant step towards a deeper knowledge on how BCG efficacy depends on individuals' age and geographical areas.

Similarly, it would provide a quantitative reference for the plausible ranges of blocking and masking levels that other TB vaccines might eventually suffer as well, which could be especially relevant in the context of impact evaluation of novel vaccines. Up to our knowledge, the BCG-REVAC studies are the only

set of trials specifically designed to discriminate the effects of Masking and Blocking. Even though this chapter is restricted to this particular setting, the framework here proposed could also be extended to the interpretation of future trials and impact evaluation of future vaccines.

## 4.2 Methods

### 4.2.1 Data analyzed: the Brazilian BCG-REVAC clinical trials

BCG-REVAC consisted of a set of cluster-randomized trials involving more than 200,000 school-aged children in the Brazilian cities of Manaus and Salvador, whose principal aim was evaluating the effectiveness of BCG under different vaccination protocols. The enrolled population of the study consisted of non-infected school children between 7 and 14 years old at the moment of randomization. Within this population, individuals presenting a positive BCG scar are separated from the rest, distinguishing, this way, the enrolled individuals who were vaccinated at birth from those who were not. Each group is then split into an intervention and a control group; individuals in the intervention group were vaccinated within the context of the trial. Summing up, there are 4 cohorts in each city: non-vaccinated (1), vaccinated after birth (2), firstly vaccinated at school age during the trial (3), and revaccinated, after a first dose applied after birth, in the trial too (4). Upon such classification of enrolled individuals in cohorts, the effectiveness of BCG vaccination strategies was measured by comparing the TB incidence rate within an end-point associated to active disease in the four cohorts, according to three different types of trials: Trial I: BCG at birth vs. no intervention (cohort 2 vs.1). Trial II: BCG first dose at school age vs. no intervention (cohort 3 vs. 1). Trial III: revaccination at school age vs. first dose at birth only (cohort 4 vs. 2).

### 4.2.2 A model to describe BCG efficacy variation: masking, blocking and immunity waning

The six clinical trials conducted within the framework of BCG-REVAC study-three types of trials per two cities- output efficacies that span from 1% to 40% protection (see Table 4.1 and Fig. 4.1, red continuous lines). In order to explain this variability, we propose a model according to which the different protection levels found in each of the four cohorts in the study,

schematically shown in Fig. 4.2, result from the interplay between the intrinsic vaccine efficacy, its temporal waning patterns, masking and blocking effects. These three mechanisms of vaccine protection shifts are ultimately responsible for vaccine's performance variation, either in space or in time.

| **Trial** | **Efficacy in Salvador** | **Efficacy in Manaus** |
|---|---|---|
| Vaccination at birth vs No vaccination | 40% (22-54%) | 36% (11-53%) |
| Vaccination at school age vs No vaccination | 34% (7-53%) | 8% (-39-40%) |
| 2nd dose at school age vs Vaccination at birth | 19% (3-33%) | 1% (-27-23%) |

Table 4.1: **Vaccine efficacies (95% CI) obtained from the BCG-REVAC trials.**

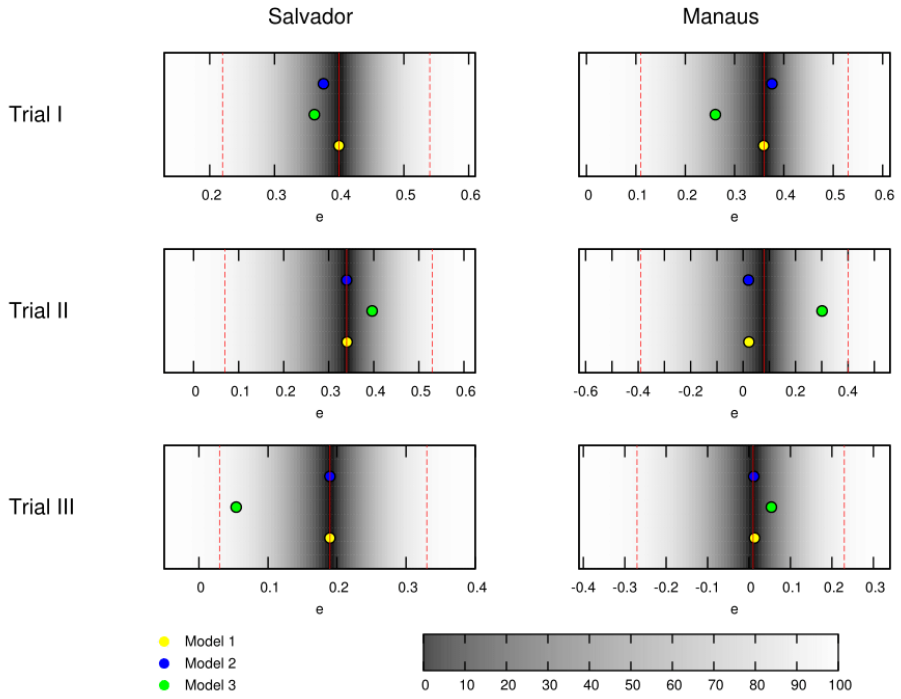


Figure 4.1: **Best fit estimates for each trial by models 1, 2 and 3** (yellow, blue and green dots, respectively) for the trials conducted in the BCG-REVAC study. The colormap represents the probability of obtaining a less extreme value of the efficacy, according to the distributions considered. The probability of zero marks the central estimate (red, continuous line) while the dashed red lines mark the 95% CI reported by Barreto et al.[166]

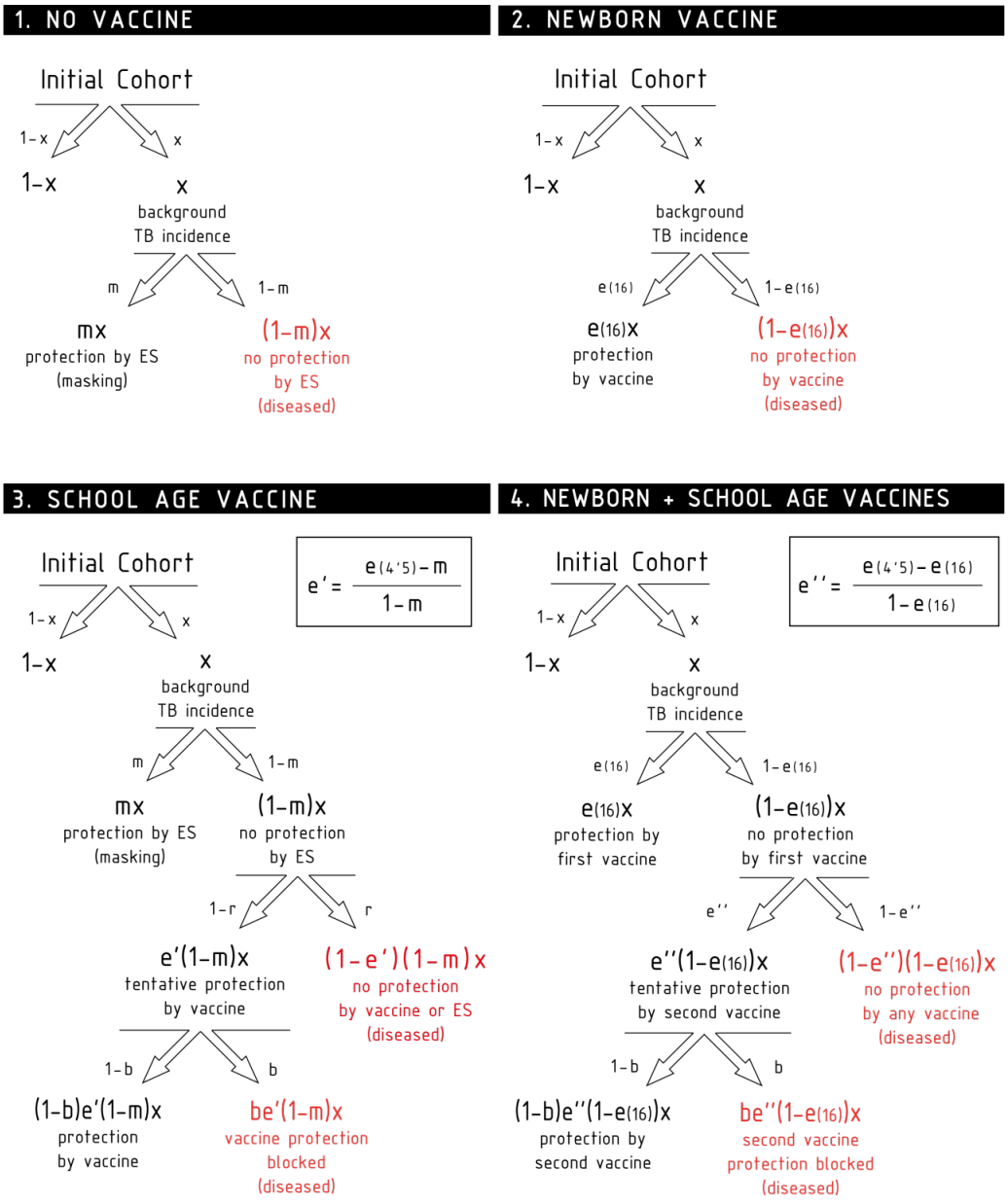


Figure 4.2: **Scheme of the different contributions to the disease risk for each cohort.**

First of all, in absence of masking or blocking, a naive vaccinated individual will receive a protection level, right after vaccination, that we call $e(0)$. As time after vaccination goes by, this protection level will wane up to $e(t) < e(0)$, generally speaking. This implies that, if we deal with a population in which the incidence rate of new TB cases per unit time is equal

to $x$; $t$ years after vaccination, this rate is modified to $(1 - e(t))x$, provided that no additional effects take place. Taking that into account, a protective vaccine will have positive efficacy values $e(t) \in (0, 1]$, being also possible for a (failed) vaccine to have a negative efficacy if it augments the disease risk among vaccinated individuals instead of reducing it. In this model, the time waning patterns of the intrinsic vaccine efficacy do not depend on the geographical area, but just on time since vaccination, which approximately is, in average, 4.5 years for school age vaccination (cohort 3) and 16 years for newborn vaccination (cohort 2), which implies the consideration of two intrinsic efficacy parameters: $e(4.5)$ and $e(16)$.

Besides vaccination, ES can also support protection against disease through the masking mechanism. The masking level, denoted by $m$, is a protection parameter formally equivalent to the intrinsic vaccine efficacy (thus verifying $m \in (0, 1]$ for a protective effect, and negative otherwise), whose effects are suffered by initially naive, non-vaccinated individuals subject to ES. Thus, in principle, the longer the time an individual has been exposed to ES—i.e., the older the individual is at the moment of observation—the higher is the masking-related protection she might show. Masking is also a geography-dependent effect, since it depends on ES, which forces us to consider two masking parameters: $m^M$ for Manaus and $m^S$ for Salvador. The dependence of these parameters on age cannot be resolved, since all the cohorts analyzed in the study have approximately the same age.

Additionally, if $e(t)$ describes the protection provided by the vaccine to a naive individual in absence of masking or blocking $t$ years after vaccination, we also need to describe how this protection is modified if the vaccine is applied to non-naive subjects. If an individual's immune system has been stimulated prior to vaccination (either by masking like in cohort 3, or by a previous vaccine, like in cohort 4 before the second dose), and consequently she is partially protected against the disease, it is unrealistic to assume that the full effect of the new dose is additive.[121]

Instead of that, the model considers that a vaccine dose applied on a previously protected individual will contribute, at most, up to resetting the initial protection levels $e(0)$, provided that no blocking of the vaccine takes place. In cohort 3, this implies that, right after the school age vaccination, if the vaccine is not blocked ($b = 0$, see below), it will have a protective effect $e'$ that will be concurrent with the masking protection $m$ so as to reduce the disease risk from $[1 - m]x$ to $[1 - e'][1 - m]x$. The estimation of $e'$ comes from assuming that such disease risk must equate what we would observe if

a vaccine of full efficacy were applied on naive individuals, and observed 4.5 years later:

$$[1-e'][1-m]x = [1-e(4.5)]x \Longrightarrow e' = \frac{e(4.5)-m}{1-m} \tag{4.1}$$

Similarly, the school-age dose at cohort 4, will add to the protection provided by the newborn dose $e(16)$, diminishing the disease risk from $[1-e(16)]x$ to $[1-e'']$ $[1-e(16)]x$. To estimate $e''$ , we assume that, if the second vaccine is not blocked, the disease risk achieved by both vaccines together $[1-e(16)][1-e'']x$ is equivalent to the disease risk reached by the same vaccine, if applied on unprotected individuals, 4.5 years after vaccination:

$$[1-e'']x = [1-e(4.5)]x \longrightarrow e'' = \frac{e(4.5)-e(16)}{1-e(16)} \tag{4.2}$$

Finally, vaccine intrinsic efficacy can be blocked by prior ES; an effect that we model through the blocking probability $b \in [0,1]$, where $b = 0$ means that no blocking appears, while $b = 1$ stands for a totally blocked vaccine, meaning that vaccinated individuals would only have the protection level that they already had before vaccination. Blocking is also a geography-dependent factor, since it is considered a consequence of ES as well, which forces us to distinguish $b^M$ and $b^S$ for Manaus and Salvador, respectively. Unlike masking, blocking does not depend on the age of the individuals at the moment of observation, but on their age at the moment of vaccination. In this case we study cohorts vaccinated at two moments in life—at birth and at the beginning of the trials—being the first of these cases (the newborn vaccination) considered blocking-free, as it is assumed that when the vaccine is applied shortly after birth, there is no place for prior ES.

Taking all these effects into account, we are left with a set of six independent parameters $\vec{P} = \{e(4.5), e(16), m^M, b^M, m^S, b^S\}$ to describe the variability observed in the trials, either temporal or geographical, under the light of blocking and masking effects, concurrently. The temporal trends of the level of protection of each cohort are schematically shown in Fig. 4.3. For Trial I the control group is cohort one that corresponds to non-vaccinated individuals. In this cohort, a level of protection above zero can only be due to masking, which is an increasing function with age. In turn, the intervention group corresponds to cohort 2 that is the newborn vaccination group: individuals are vaccinated right after birth, which provides a protection that overcomes any possible masking effect, cannot be blocked by ES and wanes with time. In the case of Trial II the vaccinated cohort is cohort 3, firstly immunized at school age. In this cohort individuals might be protected by

masking before the vaccine is applied. Then, at the moment of vaccination, if not blocked, the vaccine will overcome masking protection up to the initial value $e(0)$, which then will wane. Finally, if blocking takes place the protection provided by the vaccine will be reduced. The control cohort in this case is cohort 1 again. Finally, for Trial III the intervention group corresponds to cohort 4, joined by individuals firstly vaccinated at birth, and revaccinated at school age. At variance to the first dose, which cannot be blocked, the second dose might be blocked by ES or not, in which it will reset the initial protection levels provided by the vaccine. The control group for this trial is cohort 2, that corresponds to individuals only vaccinated at birth.

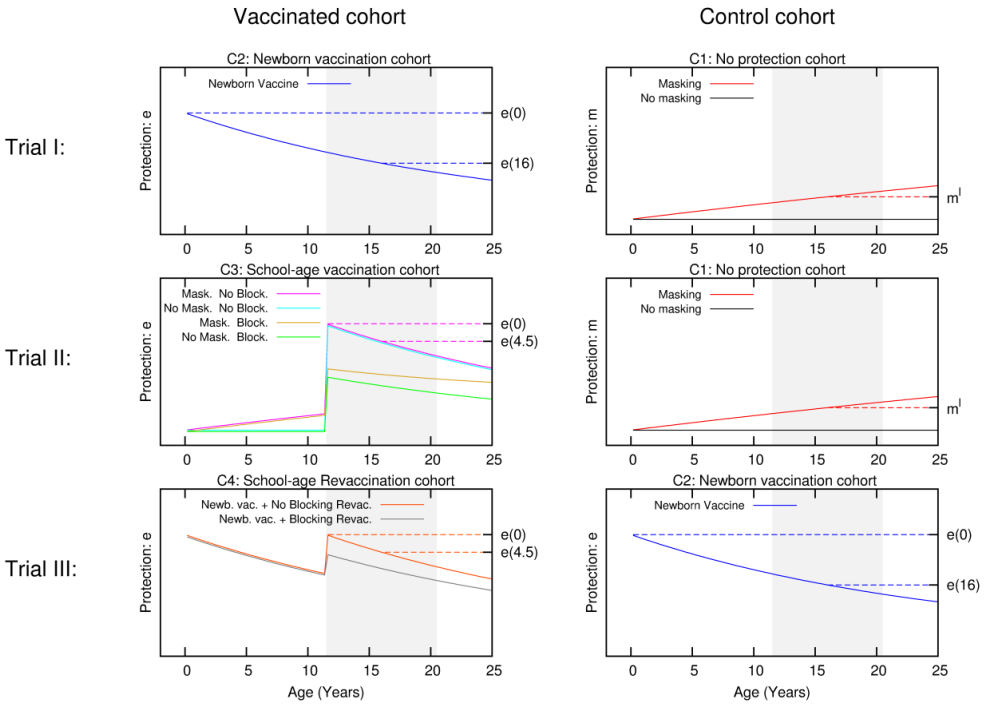


Figure 4.3: **Scheme for the temporal evolution of the level of protection for the cohorts of the three types of trials considered, according to the different vaccination strategies and ES mechanisms.** The grey shaded area represents the age window of the individuals enrolled in the study.

In the following, we will refer to this full model as model 1. In Fig. 4.2, we represent the variations on the disease rates provoked by each effect that takes place in each cohort according to model 1. Summing all the possible

contributions to the development of active disease for each cohort, we derive the general disease rates characterizing each cohort of one city as follows:

$$d_1^l = (1 - m^l)x \tag{4.3}$$

$$d_2^l = [1 - e(16)]x \tag{4.4}$$

$$d_3^l = (1 - b^l)[1 - e(4.5)]x + b^l(1 - m^l)x \tag{4.5}$$

$$d_4^l = (1 - b^l)[1 - e(4.5)]x + b^l[1 - e(16)]x \tag{4.6}$$

where the superscript indicates location, and $x$ the incidence rate observed in the population.

From equations 4.3-4.6, it is immediate to derive the expressions for the observed efficacies $\bar{e}$ of each trial according to model 1, which read as:

$$M1 : \begin{cases} \bar{e}_I^l = 1 - \frac{d_2^l}{d_1^l} = \frac{e(16) - m^l}{1 - m^l} \\ \bar{e}_{II}^l = 1 - \frac{d_3^l}{d_1^l} = \frac{e(4.5) - m^l}{1 - m^l}(1 - b^l) \\ \bar{e}_{III}^l = 1 - \frac{d_4^l}{d_2^l} = \frac{e(4.5) - e(16)}{1 - e(16)}(1 - b^l) \end{cases} \tag{4.7}$$

The system of Eq. 4.7 represents a full model for the vaccine efficacies observed during BCG-REVAC trials, which is based on the assumption that the sources of geographical variability for BCG's performance are both masking and blocking effects. From the full model, two reduced versions can be conceived: a masking-free model (model 2 in the following) in which $m^M = m^S = 0$, and a blocking free model in which $b^M = b^S = 0$ (model 3). The efficacies associated to each trial, for models 2 and 3 straightforwardly read as follows:

$$M2 : \begin{cases} \bar{e}_I^l = e(16) \\ \bar{e}_{II}^l = e(4.5)(1 - b^l) \\ \bar{e}_{III}^l = \frac{e(4.5) - e(16)}{1 - e(16)}(1 - b^l) \end{cases} \tag{4.8}$$

$$M3 : \begin{cases} \bar{e}_I^l = \frac{e(16) - m^l}{1 - m^l} \\ \bar{e}_{II}^l = \frac{e(4.5) - m^l}{1 - m^l} \\ \bar{e}_{III}^l = \frac{e(4.5) - e(16)}{1 - e(16)} \end{cases} \tag{4.9}$$

By considering these three models, this approach allows quantifying and comparing the plausibility of blocking and masking hypotheses to potentially

explain the variation in BCG efficacy trials observed in the controlled setup conceived in the BCG-REVAC trials, taking into account the non-linearities associated to each mechanism, which play a central role in the derivation of Eqs. 4.7-4.9.

### 4.2.3 Models solution: parameters estimation and confidence intervals

In order to identify the set or sets of parameters yielding a best fit for the efficacies observed in BCG-REVAC trials, we compare the model prediction associated to any parameter set $\vec{P}$ to a set of empirical probability distributions derived from BCG-REVAC data. From each of the confidence intervals reported in Barreto et al., (2014) we build a two-piece normal distribution [191] for each trial reported, centered in the reported values $[\bar{e}_i^l]_{\text{BCG-REVAC}}$ (for location $l \in \{\text{Manaus,Salvador}\}$ and trial $i \in \{I, II, III\}$), and with asymmetric variances $[\sigma_i^l]^{\pm}_{\text{BCG-REVAC}}$ equal to one half the radius of the confidence intervals reported in Barreto et al., (2014), so preserving the confidence levels of the intervals reported (see Fig. 4.1).

Once the empirical distributions have been defined, for each possible set of parameters and for each of the six trials we define the Z -score associated to the model prediction as:

$$Z_i^l(\vec{P}) = |\frac{[\bar{e}_i^l]_{\text{mod}} - [\bar{e}_i^l]_{\text{BCG-REVAC}}}{[\sigma_i^l]^{\pm}_{\text{BCG-REVAC}}}| \tag{4.10}$$

where $[\sigma_i^l]^{\pm}_{\text{BCG-REVAC}}$ will take each of its two possible values depending on the sign of $[\bar{e}_i^l]_{\text{mod}} - [\bar{e}_i^l]_{\text{BCG-REVAC}}$. From $Z_i^l(\vec{P})$, we define the corresponding p-values $p_i^l[Z_i^l(\vec{P})]$ as the probability of the empirical distributions reproducing BCG-REVAC data to have a $Z$-score $\tilde{Z}$ so that $|\tilde{Z}| > |Z_i^l(\vec{P})|$. This allows us to define the following likelihood function:

$$L(\vec{P}) = \prod_{l,i} p_i^l[Z_i^l(\vec{P})] \tag{4.11}$$

to maximize so as to identify the model's parameters $\vec{P}^*$ more likely to yield the BCG-REVAC results. The global landscape of $L(\vec{P})$ is explored using a hill-climbing algorithm designed to identify all possible local maxima in the space of parameters. Finally, a Levemberg-Marquardt algorithm [160] is used to find a more accurate value of the global maximum, if the latter is unique.

In order to estimate the confidence interval associated to model estimation, the following numerical procedure is performed. First, and starting from the maximum likelihood estimate $\vec{P}^*$ , we move on each parameters' axis until a value of $L = 0.05$ is reached in each case. We call this increment $A_j (j \in [1, 6])$ (see Fig. 4.4). These values are not symmetrical, again, and so we distinguish between $A_j^+$ and $A_j^-$. Using these asymmetric widths, we construct a two-piece normal distribution for every parameter, centered in $\vec{P}_o$ and having an asymmetric variance given by $\sigma_j^{\pm} = cA_j^{\pm}$ , where $c$ is a common modulation coefficient. Besides, the distribution is truncated at 1. Finally, we numerically estimate $c$ by generating sets of points in the parameter space whose coordinates in each axis are obtained from the split normal distributions mentioned for an initial guess of $c$. Through an iterative process we search the value $c = c^*$ for which a 95% of the points generated in the parameters space, yield efficacy estimations verifying $L(\vec{P}) > 0.05$. Once we have found the optimal value of the scaling coefficient, the reported uncertainty of the $j$-th parameter corresponds to 95% CI given the distributions we have used.

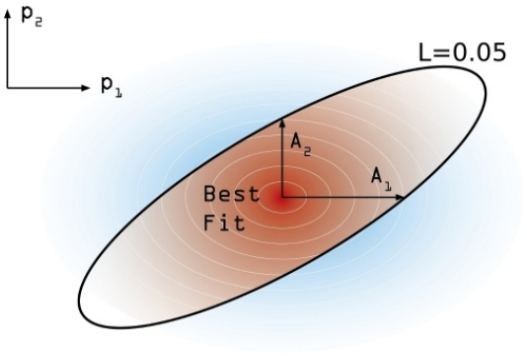


Figure 4.4: **Confidence intervals estimation scheme.** Degraded shades represent the joint probability density associated to the estimation of confidence intervals around the model best fit. The modulation coefficient $c$ is determined so as to make the brown area within the black line of $L(\vec{P}) = 0.05$ to precisely accumulate the 95% of the total joint probability distribution.

## 4.3 Results

In order to find the sets of parameters yielding best estimates of BCG efficacies according to these models, we have performed a series of numerical optimization procedures seeking for likelihood maximization. First, we are

interested in addressing whether a unique likelihood maximum exists across the parameter space of each model or whether, on the contrary, there exist multiple parameter combinations associated to comparable values of $L(\vec{P})$ close to the maximum. This is an important point to address, since the existence of different maxima in a model would indicate the inability of the model to univocally quantify the effects causing the efficacy variations observed. To tackle this question, we performed an iterative hill-climbing algorithm starting from 20,000 random points across the parameter space for each model. The algorithm works by proposing at each step a random displacement following a uniform distribution in the parameter space within a hyper-cube of size $d = 0.001$, and accepted only if it corresponds to an increasing of the likelihood function $L(\vec{P})$. The algorithm stops when no further move is accepted after $N = 10^7$ rejected displacements (i.e., the function $L(\vec{P})$ reaches a maximum) . As it can be seen in Fig. 4.5, while model 2 presents a unique likelihood maximum ($L(\vec{P}^*) = 0.53$), models 1 and 3, which contemplates masking, fails at providing a univocal vaccine's description associated to a unique solution from likelihood optimization.

Instead of that, as we can see in Figs. 4.5A, 4.5C, 4.5D and 4.5E models 1 and 3 present a parameters cliff across which, model's likelihood is near to its maximum, and largely comparable ($L(\vec{P}^*) = 0.79$ for model 1, and $L(\vec{P}^*) = 0.002$ for model 2). Furthermore, a relative likelihood test comparing models 2 and 3 (that is, comparing blocking vs. masking as exclusive mechanisms) yields a relative likelihood $\frac{L_3(\vec{P}^*)}{L_2(\vec{P}^*)} = \frac{0.002}{0.53} = 3.8 \times 10^{-3}$. This result, considering that both models share the same amount of parameters, highlights again the inability of masking to provide a picture for vaccine efficacy variation as accurate as blocking does, as we can also see in Fig. 4.1, where the best fit provided by each model is presented as well.

If the analysis of model 3 and its comparison against model 2 allows us to discard masking as an autonomous mechanism able to explain the vaccine efficacy measured in the trials, it remains to be elucidated whether its consideration in addition blocking in model 1 might still be able to significantly improve the fitting of the observed data. To answer this question, we conduct a simple likelihood ratio test in which the null and full models are, respectively models 2 and 1. From such test, we obtain that the statistic: $\chi^2 = -2\ln\left(\frac{L_2(\vec{P}^*)}{L_1(\vec{P}^*)}\right) = 0.80$, is a chi-square distributed variable with 2 degrees of freedom (difference between number of parameters of models 1 and 2) under the null model. The obtained value does not allow to discard it even with a 50% confidence ($\chi^2(p = 0.5, df = 2) = 1.39$), which indicates

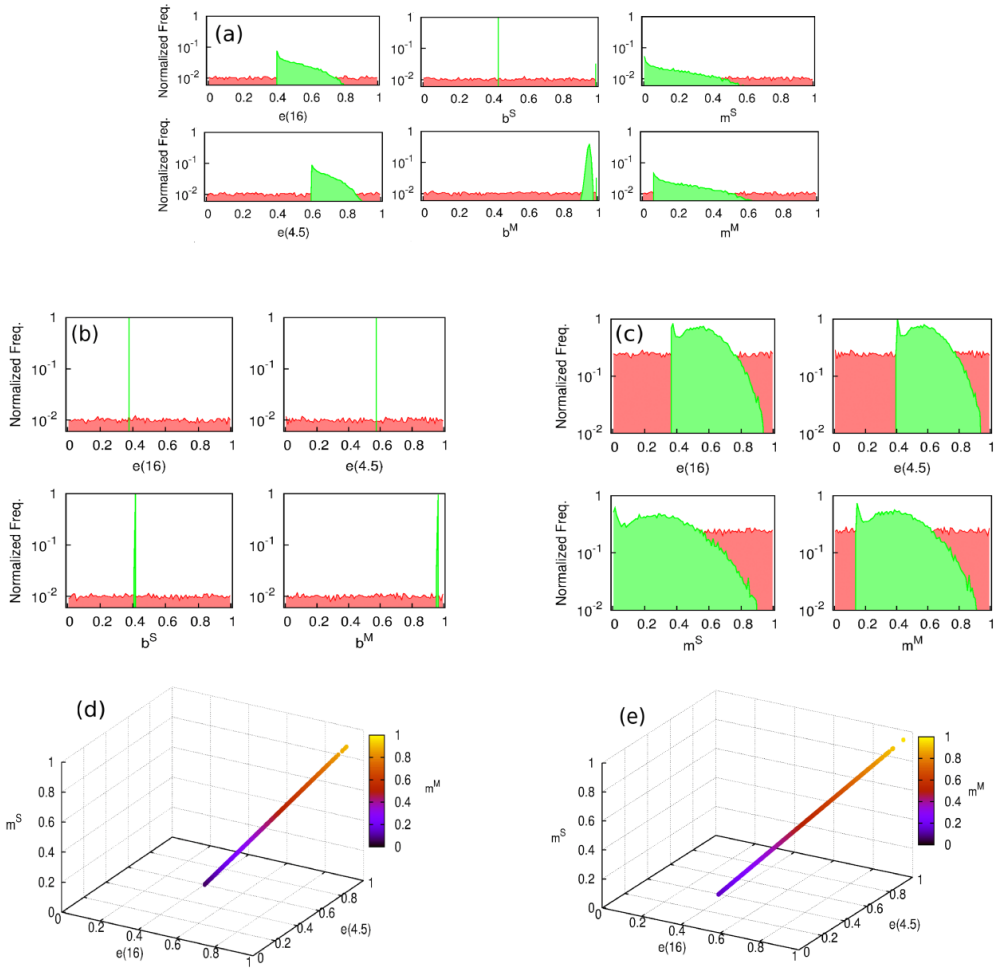


Figure 4.5: **Distribution of the parameters which yield a maximum in the likelihood function.** (a-c) Hill climbing algorithm distributions for models 1, 2 and 3, respectively. Random initial points in red. In green, we see the peaked distribution of the end points of the algorithm around the solution of the models. (d and e) parameters cliff yielding quasi-constant values of maximum likelihood $L(\vec{P}^*) = 0.79$ (d) for model 1 and $L(\vec{P}^*) = 0.002$ for model 3 (e). Model 2 is the only model capable of provide an univocal description yielding maximum likelihood.

that masking is not just unable to provide an acceptable description of the observed data by itself but also makes no significant contribution to explain the variations in vaccine efficacies observed in the trials under study, when

considered in addition to blocking. This is also reflected in the close estimates that are found for blocking parameters in models 1 and 2 (see Table 4.2 and Fig. 4.5).

| **Parameter** | **Model 2 (only blocking)** |
|---|---|
| $e(4.5)$ | $57.7\%(46.8\% - 68.6\%)$ |
| $e(16)$ | $37.6\%(29.3\% - 45.8\%)$ |
| $b^M$ | $96.4\%(51.9\% - 99.8\%)$ |
| $b^S$ | $41.1\%(14.2\% - 68.0\%)$ |

Table 4.2: **Optimal parameters of model 2.** Models 1 and 3 are unable to provide a unique parameter set yielding maximum likelihood.

Besides, if we analyze the combination of parameters that formed the cliff of maximum likelihood in model 1, we see that it consists in very similar levels of masking for the two different cities, which enters into conflict with the mentioned correlation between ES effects and closeness to equator.

In summary, these results point at blocking as the only plausible source of vaccine efficacy variation between the two mechanisms considered, validating the qualitative interpretation of the BCG-REVAC outcomes by Barreto et al.[166] The best fit of model 2 yields a likelihood $L(\vec{P}^*) = 0.53$, which corresponds to moderate blocking levels in Salvador ($b^S = 0.41$ c.i.$[0.14, 0.68]$) and to almost total blocking in Manaus ($b^M = 0.96$ c.i.$[0.52, 1.00]$). These results are consistent with the assumed correlation between ES action strength and closeness to equator.

## 4.4 Discussion

Understanding the mechanisms driving ES effects on BCG performance is a crucial task in the agenda towards the development of new tuberculosis vaccines. In this work, we have proposed a mathematical model that allows the quantitative evaluation of these two effects based on the BCG-REVAC trials performed in Brazil.[166] We have seen that the divergence in the measured efficacies of the trials is explained with high values of blocking, which concur with the qualitative discussion made in Barreto et al.[166] Furthermore, we have also observed for the first time that no alternative behavior of BCG is compatible with the observed data within the context of a model in which

BCG's variability is entirely attributed to ES sensitization.

Admittedly, the range of applications of the results here exposed must be restricted to the provision of a plausible explanation for the efficacy variation patterns observed within a controlled context such as the BCG-REVAC studies, in which the equivalent design of the trials in both cities makes reasonable to assume that ES is the only mechanism responsible for the variations observed. In that sense, the quantitative conclusions reached in this work should be interpreted as a mean to discriminate between the two mechanisms studied when it is reasonable to neglect any source of vaccine variation foreign to ES. Nonetheless, it is worth remarking that this approach cannot provide any insight on the relevance of ES itself when compared to other plausible sources of variability that could also affect vaccine's efficacy, such as diversity in production, administration, and type of BCG vaccine strain used, as well as the TB strain that circulates in a particular population, among others. In that sense, the analysis of new, hypothetical trials similarly structured, conducted in other geographical areas, could certainly yield different results, if additional sources of variation not considered in this work were playing a relevant role.

The model here proposed could however be generalized so as to address some additional questions that go beyond a simple comparison between masking and blocking, which would involve a more detailed description of the blocking effect itself.

On the one hand, it is pertinent to ask whether prior vaccination with BCG might trigger a blocking effect comparable in magnitude to that caused by ES; a question that could be tackled by an extension of the model here proposed in which two blocking parameters –one associated to each source– are considered instead of one. Remarkably enough, distinguishing between BCG vaccine and ES as possibly different causes of blocking might lead to relevant quantitative consequences in what regards impact evaluation of novel vaccines, mostly because the old vaccine is still used in the vast majority of countries, also in geographical areas in which low levels of ES would be expected.

On the other hand, an additional limitation of this study, inherited from the BCG-REVAC studies design, is due to the restriction of trials' endpoints to diseased and not diseased individuals, without measuring infection as a third relevant outcome. This limitation prevents us to address the important question of whether the vaccine is blocked in its protective role against

infection, or if, instead, blocking interferes more intensely with the vaccine's performance at reducing the progression rates from latency to active disease.[75,192]

If infection was registered as an additional endpoint in the clinical trials –something, in general, feasible[193]– this approach could then be extended so as to estimate two different blocking components associated to the impairment of vaccine's protection against infection and active disease independently. Once again, such hypothetical study could bring important insights for future vaccine development, and in particular could contribute strongly to the debate of what should be the primary goal of TB vaccines.[107] Generally speaking, more studies are needed to evaluate how general are the patterns found by BCG-REVAC trials, with the ultimate goal of assessing a positive explanation to the long lasting problem of BCG efficacy variation patterns.

## 4.5 Conclussions

The crucial implications of discriminating and quantifying masking and blocking effects for TB vaccine development are twofold. On one hand, understanding the range, and causes behind variations of BCG efficacy is essential,[194] since the efficacy of any novel vaccine will be measured against BCG. On the other hand, depending on where a new vaccine is applied and how old are the target populations, masking or –more likely– blocking effects would affect new vaccines too.

These issues affect different stages of the vaccine development pipeline, as sketched in Fig. 4.6. In the first place, during the process of vaccine evaluation in the context of clinical trials, studies of new tuberculosis vaccines should account for the possibility that prior sensitization may compromise their effects.[76] In this sense, and even if a new vaccine targeting TB in adolescents and adults rather than any other age group is expected to have the quickest impact on disease transmission and control, before we address the question of impact of novel vaccines, it is essential to know if the vaccine is more effective than BCG. The most reliable way of knowing whether a new vaccine works better than BCG is by conducting an efficacy trial in a naive population without previous ES (e.g., previous BCG vaccination, mycobacterial infection and/or TB contact) in order to avoid possible effects of masking or blocking.[74,121,166]

Furthermore, and once the efficacy estimation is complete, in order to

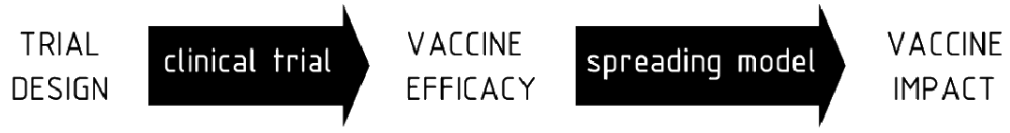


Figure 4.6: **Scheme of the basis for evaluation of anti tuberculosis vaccines in absence of universally reliable protection correlates.** First stage: design of vaccine efficacy determination clinical trials: the age of the cohorts must be elected taking into account that prior exposure to mycobacteria –either environmental, M. tuberculosis after exposure or even prior TST or also BCG– may corrupt the observed vaccine efficacies. Second stage: vaccine impact evaluations: bulk, short-term and long-term impact forecasts should be equally considered, as well as age-distributed impacts in terms of cases, infections and casualties prevented.

produce any reliable vaccine impact and cost-effectiveness forecast, modeling scenarios contemplating ES deleterious effects on TB vaccines are mandatory. The fact that, according to this analysis, blocking emerges as the driving effect behind BCG variability poses a potential pitfall to any vaccination strategy focused on individuals older than those analyzed here, including most strategies conceived so far for booster vaccines. This is especially worth noticing because blocking, unlike masking, is not supposed to degrade the vaccine-induced protection obtained further during life by individuals immunized promptly after birth. Again, even if immunizing adolescents is thought to provide better impacts than vaccination strategies focused on younger age-segments, if such a novel vaccine is affected by blocking just as BCG is, then its impact will decrease in a way that, given the high blocking levels here identified, might even revert the comparison. As suggested by Helen McShane "we should optimize deployment of BCG to administration as close to birth as possible".[194] This should be the case for new priming live vaccines candidates based on BCG replacement strategies as well.[82]

Taken altogether, the results highlight the need for measuring ES effects on novel vaccines performance, as well as of diversifying vaccination strategies. As this chapter proves, the execution of clinical trials is essential in the case of TB, and more information can be extracted from them than the classical assessment of efficacy. In the next chapter, we will study the design of efficacy clinical trials and propose a new analysis that could improve our knowledge of the new vaccines being tested, which in the long term could provide more accurate impact evaluations and less biased decisions by policymakers.

# Chapter 5

# Design principles for TB vaccines' clinical trials

## 5.1 Introduction

As it has been explained throughout this thesis, nowadays a very active research community is engaged in the development of disparate candidates for a new TB vaccine, several of them being tested in clinical trials.[82,195,196] Since the resources available for this collective endeavor are limited, the development of rational approaches for filtering vaccine candidates represents a first order priority for Public Health Organizations and funding agencies. Accordingly, a well-defined stage-and-gate system has been defined and implemented, to which the different candidates must adhere in order to advance through the subsequent phases of the vaccine development pipeline.[197] By providing a transparent framework for the evaluation and comparison of candidates at all stages, the system allows prioritizing the investment on the vaccines showing better performances regarding safety, immunogenicity and protection at each step, with the ultimate objective of maximizing the likelihood of success for the final finding of a new, safe and impactful TB vaccine.

Gathering the information needed to guarantee the eventual success of a vaccine poses a number of conceptual challenges that manifest at different stages of the clinical pipeline, making the evaluation and comparison of the different candidates under consideration an arduous task. The lack of protection correlates for TB[198,199] hinders early efficacy evaluations, affecting our ability to identify the very presence of any protective effects from a given vaccine at the early stages of its development. That major limitation factually forces researches to rely new vaccines efficacy estimates to large randomized

phase II and II-b clinical trials, which require the recruitment of thousands of individuals in high prevalence epidemiological settings during years to be completed. In this regard, the recent results from the nearly designed phase IIb trials of the novel vaccines MVA85A,[200] and H4:IC31[201] even though it fails at providing evidence of significant protection, represents a solid framework for such studies.

However, even once the task of measuring the vaccine's efficacy at reducing risk of disease ($VE_{dis}$) is achieved upon completion of a phase IIb clinical trial, the interpretations of these results are not always immediate and can be due to different mechanisms of action of the vaccine. On the one hand, a vaccine can contribute to reduce disease risk by decreasing the fraction of fast progressors, increasing the probability that newly infected individuals are able to contain immediate bacterial proliferation developing latent TB infection (LTBI). Alternatively, the vaccine can delay the onset of active TB, slowing down the dynamics of fast progression instead of preventing it. These two possible mechanisms have different dynamical interpretations in terms of epidemiological transmission models, and might appear, in principle, independently and not necessarily correlated. The question of whether an individual will progress to active disease rapidly after infection or not depends on the ability to induce an initial strong innate immune response to the pathogen along with a later T-cell mediated adaptive response both of which are needed, if not to prevent infection by completely eliminating the bacteria, at least to arrest bacterial growth and confine the pathogen within granuloma.[44] If such complex response fails, fast transition to disease will take place, as a consequence of the host's inability to both eliminate or confine bacteria around closed granuloma, yielding to rapid pathogen proliferation. Factors affecting the probability of individuals to join fast or slow paths to disease upon infection are both environmental and genetic, but little is known about how they impact, if they do it at all, the delay observed between infection and onset of symptoms in recent transmission TB cases (i.e. fast progressors).[44]

The main focus of this chapter is the analysis of this apparently innocent mechanistic degeneracy problem, in virtue of which, it is hard to distinguish, in the context of a clinical trial designed to estimate vaccine efficacy against disease, a vaccine that prevents from fast-progression from a vaccine that simply delays it. After a formal description of the question, we use Monte-Carlo methods to simulate synthetic clinical trials,[202] along with a compartmental transmission model to produce impact forecasts (an upgrade of the model described in Chapter 3)[129] of vaccines that, yet compatible with

single trial-derived lectures of $VE_{dis}$, have different effects on fast-progression dynamics. Upon such an exercise, we find that vaccines reducing the probability of fast progression are expected to offer, for the same observed levels of $VE_{dis}$, significantly larger long-term impacts than vaccines that delay it. This situation translates into an excruciating level of uncertainty in what regards model-based impact evaluation of vaccines protecting against disease, unless a method for telling apart which of the two mechanisms constitutes a more plausible model under the light of the trial data is provided. Finally, we develop such a method for analysing raw results of clinical trials, which allows to isolate and measure what are the mechanisms of action of the vaccine that prevent or delay fast-progression to disease. This translates into more accurate impact forecasts and more faithful characterizations and comparisons of different types of vaccines, in a fraction of cases that depends on trials' specifications (cohort sizes and follow-up periods duration).

## 5.2 Methods

### 5.2.1 Vaccine protection mechanisms against disease: reducing fast-progression probability vs. reducing fast transition rates to active TB

The essential goal of this work is to provide a means to interpret the results of a generic phase II clinical trial of vaccine efficacy in terms of compartmental models used downstream to evaluate vaccines' impact and cost-effectiveness. In the more elementary version of these models (figure 5.1A),[55,97,101,102,108,203,204] susceptible individuals get infected at a rate $\beta$ defined by the interplay of their intrinsic susceptibility to infection and the frequency of contacts with infectious individuals that they experiment. Upon infection, they can develop fast-progression to disease (F) with a probability $p$, or slow-progression (latent infection L) with the remaining probability $1-p$. Finally, while fast progressors develop disease (D) at a rate $r$ associated to typical transition times lower than one year, latent individuals can remain so for decades, only eventually falling sick, at a rate $r_L << r$. When we are talking about vaccinated individuals, parameters $\beta$, $p$ and $r$ get reduced to $(1-\varepsilon_\beta)\beta$, $(1-\varepsilon_p)p$ and $(1-\varepsilon_r)r$, respectively.

From this general picture, a series of assumptions are necessary to interpret the dynamics observed in a trial in terms of such minimal transmission model. A typical randomized, placebo controlled clinical trial of vaccine efficacy consists of two cohorts of initially susceptible individuals,

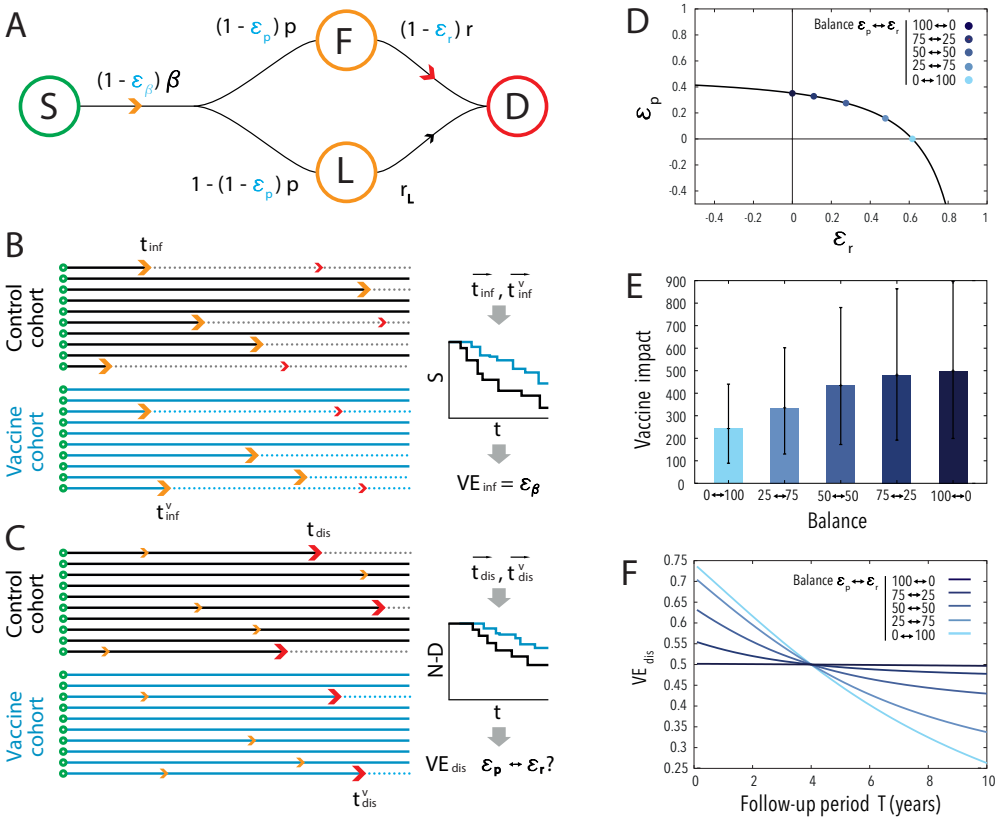


Figure 5.1: **Assessment of the degeneracy problem in efficacy evaluation.** A: Transmission model. B: Scheme of times of transition to infection, and associated survival curves. C: Scheme of times of transition to disease (after infection) and associated survival curves (of non-disease states). D: Curve of values of ($\varepsilon_p$,$\varepsilon_r$) compatible with a measurement of $VE_{dis} = 0.5$ after 4 years of follow-up (assuming $\varepsilon_\beta = 0$). We have marked 5 different points in this curve, with very different balance between the two mechanisms that cause the degeneration, to be used in next examples. E: Forecasted impacts (thousands prevented cases) obtaining after introducing the 5 highlighted vaccines in a TB spreading model[129] in Ethiopia for the period 2025-2050. F: Evolution of measured $VE_{dis}$ for the 5 highlighted vaccines as a function of the follow-up period.

(i.e. QuantiFERON-TB (QFT) negative[200]). At the beginning of the study, the vaccine is supplied to one of these cohorts, while a placebo is given to the other. Then, individuals will be periodically tested for infection (QFT)

or disease (standard TB diagnosis criteria), during a given follow-up period short enough (i.e. less than 10 years) for us to assume that all the transitions to active disease observed were due to fast progressors in the first place. Finally, analyzing transition times to infection and disease end-points, two independent vaccine efficacy parameters can be measured: efficacy against infection $VE_{inf}$ and against disease $VE_{dis}$, (figure 5.1, panels B,C). Under the light of this elementary transmission model, such efficacy observations can arise from the three putatively independent mechanisms: reduction of susceptibility to infection (via $\varepsilon_\beta > 0$, see figure 5.1A), reduction of the fraction of fast-progressors ($\varepsilon_p > 0$) and reduction of the rate of fast progression to disease ($\varepsilon_r > 0$).

All that said, the nature of the question under analysis turns evident: how to estimate three independent mechanisms of action of the vaccine from two measurements of vaccine efficacy? Regarding vaccine protection against infection, that question is easy to answer, since the only way for a vaccine to protect immunized individuals against infection is to reduce their probability of getting infected upon a contact with an infectious individual, which simply implies that $VE_{inf} = \varepsilon_\beta$. Instead, vaccine's protection against disease offers a more complex outlook, because different vaccines, showing different combinations of effects on fast progression probabilities and transition rates to disease ($\varepsilon_p$ vs $\varepsilon_r$) are compatible with a single lecture of $VE_{dis}$ (figure 5.1D). This implies that, from a trial-derived lecture of $VE_{dis}$ it is not possible to say whether the vaccine decreases individuals' risk to develop disease by reducing the probability for them to become fast progressors upon infection or, alternatively, by slowing down the rate at which they develop TB.

Once the problem is identified, a direct way to quantify its relevance is to interrogate whether vaccines acting through different combinations of $(\varepsilon_p, \varepsilon_r)$ that are still compatible with a common value of $VE_{dis}$ might produce different impacts when applied on large populations. To answer this question, we capitalized on the epidemiological model from Chapter 3, with a small modification to incorporate vaccines. Using such a model, we simulated the introduction of vaccines in Ethiopia, and obtained vaccines' impact estimations in terms of TB cases prevented from 2025 to 2050, upon a newborn-focused vaccination campaign implemented at the beginning of that period (figure 5.1E, vaccines compatible with $VE_{dis} = 50\%$ in a 4 year-trial). For this particular case, we found a difference of 256 thousand cases prevented (104-466 95% CI) between the first vaccine ($\varepsilon_p = 0.00$, $\varepsilon_r = 0.74$) and the last one ($\varepsilon_p = 0.50$, $\varepsilon_r = 0.00$), even though both vaccines provide an efficacy of $VE_{dis} = 50\%$ in a 4 year-trial. This supposes a relative difference of 51% (45-

59 95% CI)–, evidence the critical relevance of this degeneracy, which would compromise the reliability of any comparison between any disease-preventing vaccine candidates solely based on their observed efficacy levels $VE_{dis}$.

These results have a straightforward interpretation: slowing down transition rates to disease of fast progressors will always be less effective, in the long term, than removing them from the fast-transition branch. A complementary way to illustrate this is to evaluate the time evolution of the efficacy $VE_{dis}$ that would be observed, for a single vaccine, in trials of different duration depending on its mechanisms of action (figure 5.1F), which can be derived analytically (see section 5.2.2). In this sense, a vaccine that reduces the intrinsic risk of developing fast progression ($\varepsilon_p > 0$) provides a reduction in the risk of disease that is almost independent of the observation time. Instead, for a vaccine that only slows down the rate at which fast-progressors develop active disease, the observed efficacy is a rapidly decreasing function of time: the longer the trial, the less efficient the vaccine would appear to be. This situation, which is intrinsically bound to the lower impacts forecasted from $\varepsilon_r$-based vaccines, translates into a violation of the hypothesis of constant proportional risks underlying standard survival analysis (Cox regression) typically used to estimate $VE_{dis}$ for those vaccines, a question whose implications are discussed in detail in section 5.2.3.

### 5.2.2 Analytical solution of the transition model

According to the scheme of transitions that we are using to model a clinical trial (see panel A of figure 5.1), the evolution of the four states ($S$: Susceptible, $F$: Fast-progressors, $L$: Slow-progressors, latent infection and $D$: Disease) is given, in the control cohort (subindex $c$), by:

$$\frac{dS_c(t)}{dt} = -\beta S_c(t) \tag{5.1}$$

$$\frac{dF_c(t)}{dt} = \beta p S_c(t) - r F_c(t) \tag{5.2}$$

$$\frac{dL_c(t)}{dt} = \beta(1-p) S_c(t) - r_L L_c(t) \tag{5.3}$$

$$\frac{dD_c(t)}{dt} = r F_c(t) + r_L L_c(t) \tag{5.4}$$

For the vaccinated cohort (subindex $v$), parameters $\beta$, $p$ and $r$ are modified by the action of the vaccine:

$$\frac{dS_v(t)}{dt} = -(1-\varepsilon_\beta)\beta S_v(t) \tag{5.5}$$

$$\frac{dF_v(t)}{dt} = (1-\varepsilon_\beta)\beta(1-\varepsilon_p)pS_v(t) - (1-\varepsilon_r)rF_v(t) \tag{5.6}$$

$$\frac{dL_v(t)}{dt} = (1-\varepsilon_\beta)\beta(1-(1-\varepsilon_p)p)S_v(t) - r_L L_v(t) \tag{5.7}$$

$$\frac{dD_v(t)}{dt} = (1-\varepsilon_r)rF_v(t) + r_L L_v(t) \tag{5.8}$$

In this model we implicitly assume that the individuals in the cohorts correspond to a small fraction of the total population in the site, and thus, their contribution to overall transmission once sick can be neglected. Solving the system of differential equations, we obtain the proportion for every state at each cohort. Initial conditions are $S_x = 1$ and $L_x = F_x = D_x = 0$ for $x = c, v$, i.e. all individuals are susceptible at the beginning of the trial.

$$S_c(t) = \exp(-\beta t) \tag{5.9}$$

$$S_v(t) = \exp(-(1-\varepsilon_\beta)\beta t) \tag{5.10}$$

$$F_c(t) = \frac{\beta p\left(\exp(-rt) - \exp(-\beta t)\right)}{\beta - r} \tag{5.11}$$

$$F_v(t) = \frac{(1-\varepsilon_\beta)\beta p\left(\exp(-(1-\varepsilon_r)rt) - \exp(-(1-\varepsilon_\beta)\beta t)\right)}{(1-\varepsilon_\beta)\beta - (1-\varepsilon_r)r} \tag{5.12}$$

$$L_c(t) = \frac{\beta(1-p)\left(\exp(-r_L t) - \exp(-\beta t)\right)}{\beta - r_L} \tag{5.13}$$

$$L_v(t) = \frac{(1-\varepsilon_\beta)\beta\left(1-(1-\varepsilon_p)p\right)\left(\exp(-r_L t) - \exp(-(1-\varepsilon_\beta)\beta t)\right)}{(1-\varepsilon_\beta)\beta - r_L} \tag{5.14}$$

$$\begin{aligned} D_c(t) \;=\; & 1 - \exp(-\beta t) - \frac{\beta p\left(\exp(-rt) - \exp(-\beta t)\right)}{\beta - r} \\ & - \frac{\beta\left(1-p\right)\left(\exp(-r_L t) - \exp(-\beta t)\right)}{\beta - r_L} \end{aligned} \tag{5.15}$$

$$
\begin{aligned}
D_v(t) \;=\;& 1-\exp(-(1-\varepsilon_\beta)\beta t) \\
&-\frac{(1-\varepsilon_\beta)\beta(1-\varepsilon_p)p\left(\exp(-(1-\varepsilon_r)rt)-\exp(-(1-\varepsilon_\beta)\beta t)\right)}{(1-\varepsilon_\beta)\beta-(1-\varepsilon_r)r} \\
&-\frac{(1-\varepsilon_\beta)\beta\left(1-(1-\varepsilon_p)p\right)\left(\exp(-r_L t)-\exp(-(1-\varepsilon_\beta)\beta t)\right)}{(1-\varepsilon_\beta)\beta-r_L}
\end{aligned} \tag{5.16}
$$

Then, we obtain the disease-ratio at the end of the trial as follows:

$$
\rho(T)=\frac{D_v(T)}{D_c(T)}=\frac{p(B_v-C_v)\varepsilon_p+A_v-pB_v-(1-p)C_v}{A-pB-(1-p)C} \tag{5.17}
$$

where:

$$
A=1-\exp(-\beta T) \tag{5.18}
$$

$$
A_v=1-\exp(-(1-\varepsilon_\beta)\beta T) \tag{5.19}
$$

$$
B=\frac{\beta(\exp(-rT)-\exp(-\beta T))}{\beta-r} \tag{5.20}
$$

$$
B_v=\frac{(1-\varepsilon_\beta)\beta(\exp(-(1-\varepsilon_r)rT)-\exp(-(1-\varepsilon_\beta)\beta T))}{(1-\varepsilon_\beta)\beta-(1-\varepsilon_r)r} \tag{5.21}
$$

$$
C=\frac{\beta(\exp(-r_L T)-\exp(-\beta T))}{\beta-r_L} \tag{5.22}
$$

$$
C_v=\frac{(1-\varepsilon_\beta)\beta(\exp(-r_L T)-\exp(-(1-\varepsilon_\beta)\beta T))}{(1-\varepsilon_\beta)\beta-r_L} \tag{5.23}
$$

### 5.2.3 Methods for the estimation of $VE_{inf}$ and $VE_{dis}$

In order to obtain efficacy at infection, which corresponds to the parameter $\varepsilon_\beta$ in the model, we implement the Cox Regression Model,[205] since modeling of infection is trivially compatible with the premises of Cox Regression Model (i.e. Proportional Hazards). However, that is not the case for efficacy at disease. According to the modeling that we propose for post-infection processes, the applicability of Cox Regression Model is not guaranteed, since the hypothesis of proportional hazards might not be correct depending on

the features of the vaccine considered. To illustrate this situation, we have conducted Schoenfeld's Residual tests[206] for different vaccines, characterized by different values of $\varepsilon_p$ and $\varepsilon_r$. The results of such a test are represented in Figure 5.2 panel B, where we show the proportion of tests with a False Discovery Rate (FDR) below 5% (Benjamini-Hochberg correction for Multiple Testing[207]). As a conclusion, we see how vaccines that delay the fast progression rates violate more often the hypothesis of Hazards Proportionality, which is coherent with the observation shown in figure 5.1C, where we see how the observed efficacies of an $\varepsilon_r$ vaccine is strongly dependent on time.

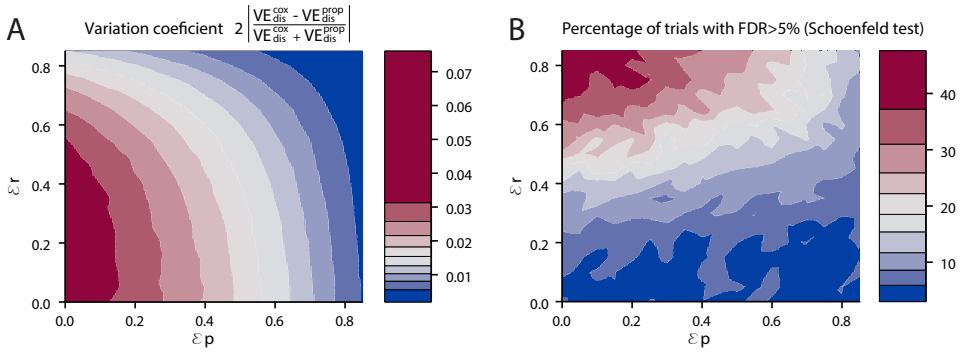


Figure 5.2: **Schoenfeld's Residual tests.** A. Coefficients of variation. Simulations have been performed with a Cohort Size of 3000 individuals and a follow-up period of 4 years. B. Proportion of tests with a FDR below 5% as a function of $\varepsilon_p$ and $\varepsilon_r$. Simulations have been performed with a Cohort Size of 3000 individuals, a follow-up period of 4 years, and 500 iterations.

Motivated by that limitation of classical survival analysis regarding $VE_{dis}$ we decided to make use of a more elementary, proportion-based estimate for measuring the efficacy of a vaccine against disease, namely $VE_{dis}(T) = 1 - \rho(T)$, where $\rho(T)$ is the disease ratio evaluated at the end of the trial, as defined in equation 5.17.

This alternative way to estimate $VE_{dis}$ (or, equivalently, $\rho$) enables a tractable analytical relationship between $\rho$, $\varepsilon_r$ and $\varepsilon_p$, which will be useful in order to determine $\varepsilon_p$ once $\varepsilon_r$ is independently estimated. Additionally, even though the proportional hazards hypothesis is not perfectly met for all vaccines, it is worth noticing that the differences observed between proportion-based and Cox-regression measures of $VE_{dis}(T)$ are small.

In Figure 5.2 we compare the results obtained with both measurements of efficacy at disease for different vaccines. We see in panel A, representing the Coefficient of Variation (CV) of such the two different measures, that differences are bound below 8% of the mean of the efficacies estimated from both methods.

### 5.2.4 Estimation of Vaccine Impact using a TB transmission model

To estimate the impacts of different vaccines (in Ethiopia during the period 2025-2050), we use the model developed in Chapter 3. The first stages of the Natural History considered in this model are equivalent to those considered for the simulation of Clinical Trials, with a first class of susceptible individuals from which individuals access two possible latency states (fast or slow) when an infection occurs.

The model, as it was described, can not be used to evaluate the impact of vaccines. Thus we need to extend the model by adding a second branch for vaccinated individuals. This second branch follows qualitatively the same Natural History, already explained in Chapter 3, but quantitatively the disease parameters that regulate all the transitions between states can be modified by the epidemiological intervention considered (in this case vaccines). Thus, for every parameter $x$ in the unvaccinated branch, we have $x' = (1 - \varepsilon_x)x$ in the vaccinated branch, where $\varepsilon_x$ represents the effect of the vaccine over the parameter $x$.

In this work, we study vaccines that act over three different parameters: the force of infection ($\varepsilon_\beta$), the probability of fast progression ($\varepsilon_p$) and the rate of fast progression ($\varepsilon_r$). We assume $\varepsilon_x = 0$ for every other parameter (i.e., the vaccine has no other effect). Individuals in the vaccinated branch contributes to the force of infection with the same weight than no vaccinated individuals (i.e., a reduction in infectiousness or contact rate as a result of vaccination is not considered).

For this study, we are considering newborn vaccination. Newborns enter the system only in the vaccinated branch starting in 2025 (which we considered as the year that, in the best case scenario, a new massive vaccination campaign could begin), while they enter the unvaccinated branch exclusively (following an unmodified Natural History) before that year. In other works, the coverage considered is 100%. It is not rare to achieve very high levels of

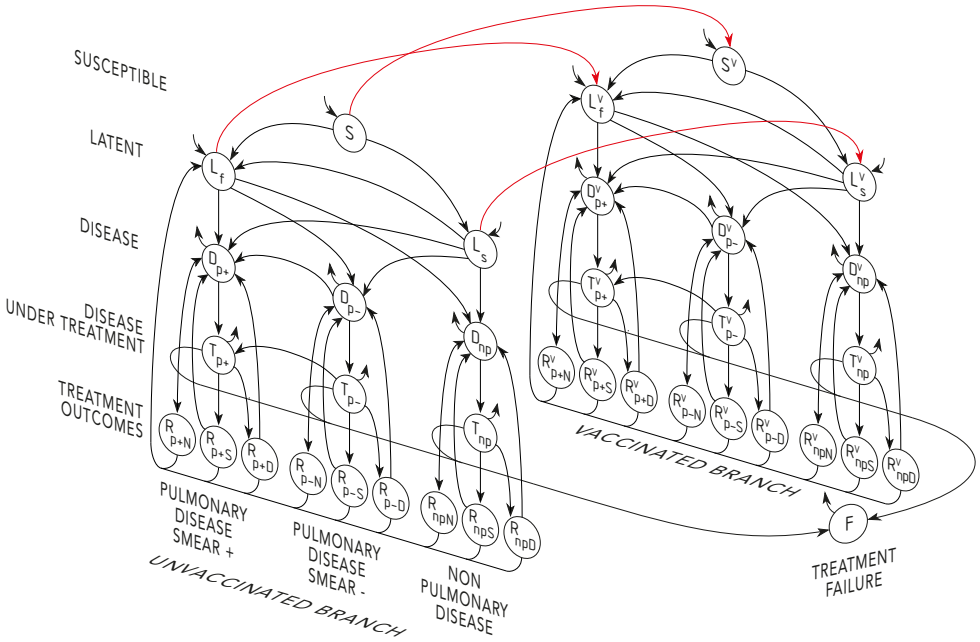


Figure 5.3: **Natural History scheme of the TB spreading model with a vaccinated branch**. To the control branch (i.e. non-vaccinated), a parallel branch where individuals have received a vaccine (i.e. red arrow transitions) is added. In the vaccinated branch, the disease transmission dynamics is the same with the exception of that, depending on the vaccine type, the infectiousness, probabilities of fast transmission and/or rates of fast-progression to disease are reduced, as dictated by the vaccine descriptors $\varepsilon_\beta$, $\varepsilon_p$ and $\varepsilon_r$, respectively.

newborn coverage, for instance BCG coverage in Ethiopia reached a 92% in 2016.[208]

The impact results that are reported in what follows correspond to the difference of incidence cases in the period 2025-2050 between the scenario in which no vaccine is introduced and the scenario where we introduce a vaccine with its corresponding specifications. Further technical specifications on model parameters, bibliographical sources, methods used to estimate uncertainty in model-based estimates, etc., can be found in Chapter 3.

### 5.2.5 Interaction between impact evaluation time window and vaccine mechanism

The impact of a vaccine will strongly depend on the period of time considered. While vaccines that are focused on the delay of fast-progression might work well in short-term, its effectiveness is compromised in the long-term as we have discussed previously. In figure 5.4, we further explore that phenomenology representing the impact, measured over different durations, for different vaccines that give the same disease efficacy in 4-year trials relaying either to different extents on $\varepsilon_r$ or $\varepsilon_p$. As we can see, in the short term (when the evaluation of impact is performed over a period that is smaller or of the same order than the period of the clinical trial itself) vaccines designed to delaying fast-progression are associated to quicker short term impacts. However, as the evaluation time window is enlarged, vaccines that prevent fast-progression more efficiently provide better impacts.

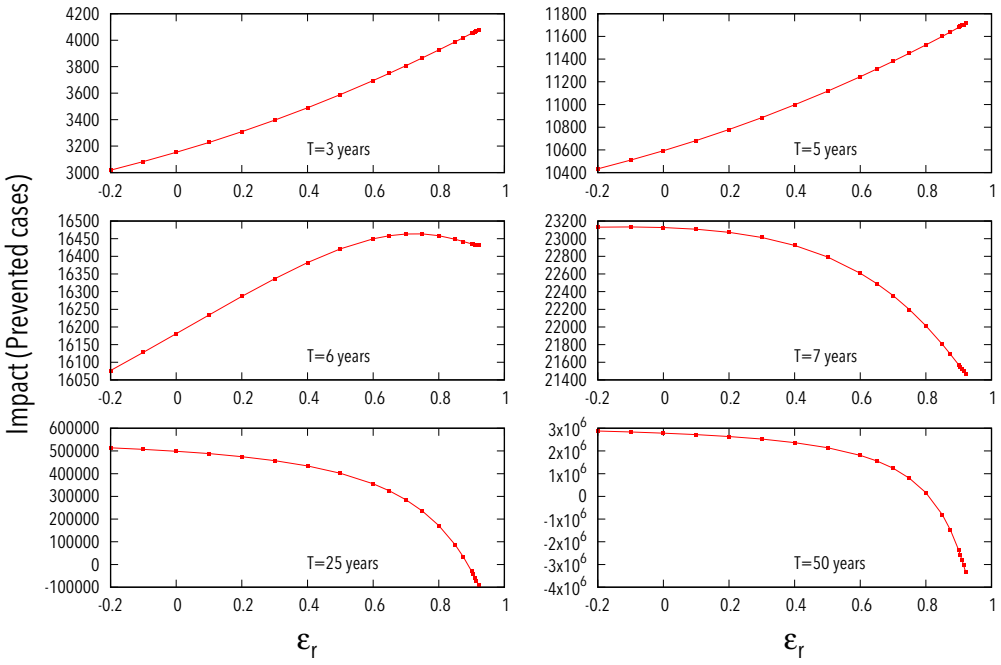


Figure 5.4: **Impact measured over different time windows, for vaccines that provides the same disease efficacy in 4 year trials.** The x-axis shows the value of $\varepsilon_r$ of the vaccines, as the parameter relevant of the curve $\varepsilon_p - \varepsilon_r$. $\varepsilon_\beta = 0$ in this case.

In principle, it is expected that the intrinsic efficacies of the vaccine are comprised between 0 and 1, where 1 would mean a perfect efficacy and 0 no effect at all. However, it is possible for a vaccine to have a negative effect. In the case of efficacies affecting rates (i.e. $\varepsilon_\beta$ and $\varepsilon_r$) there is no formal lower limit and a rate equals to $\infty$ (associated to $\varepsilon = -\infty$) would mean an instantaneous process, although a conservative enough limit of -300 is implemented to avoid numerical instabilities. On the contrary, $\varepsilon_p$ works as a modifier of a probability, which implies that $(1 - \varepsilon_p)p$ has to be comprised between 0 and 1, introducing a lower limit for $\varepsilon_p$ that is $\varepsilon_{p,\min} = 1 - \frac{1}{p}$. Furthermore, the existence of such lower bound in $\varepsilon_p$ generates in turn an upper bound for $\varepsilon_r$, since these two parameters are analytically bounded through a common value of the disease-ratio $\rho = D_v/D_c$.

Based on this result, when evaluating the state-of-the-art complete uncertainty in the evaluation of impacts, we will assign as the minimum impact that corresponding to the vaccine with the maximum possible value of $\varepsilon_r$ (and therefore minimum value of $\varepsilon_p$, given the analytical expression that links those parameters for a given $VE_{dis}$). For the maximum impact in principle there exists no bound as already mentioned. However, given the saturation of the impact that is produced for decreasing values of $\varepsilon_r$, the limit that we have imposed of $\varepsilon_r = -300$ provides a good proxy for the upper bound of impact.

### 5.2.6 Breaking the degeneracy: a method for estimating $\varepsilon_r$ and $\varepsilon_p$

Even though violation of the proportional-hazard hypothesis in trials' data constitutes a first signature suggesting delay in progression rates rather than fast progression prevention, its potential to tell apart quantitatively the possible mechanisms of action of a vaccine (see figure 5.2) is limited. Thus, we propose a different approach to provide an independent estimation of $\varepsilon_r$ and $\varepsilon_p$.

#### 5.2.6.1 Estimation of $\varepsilon_r$: truncated fit of transition rates from uncensored sub-cohorts

The method consists of adding, to the custom estimation of $VE_{inf}$ and $VE_{dis}$, a third statistical analysis aimed at directly estimating $\varepsilon_r$ from the transition times between infection and disease (figure 5.5, panel A). To do so, unlike classical survival analysis, we only make use of data associated to not-censored times: that is, individuals that have completed their transition to

disease within the follow-up period of the trial. By doing so, we can assume that all these cases correspond exclusively to fast progressors (for $r_L << r$), and obtain an analytical expression for the expected distribution of transition times observed between infection and disease $t = t_{dis} - t_{inf}$ (i.e., the time difference between the moment individuals first produce QFT positive results and when they fall sick), conditioned to the moment the infection took place. Furthermore, if we assume that the transition from active disease upon infection is a Poisson process, –as it is done customarily in compartmental models in mathematical epidemiology[55,108,209]–, the theoretical probability distribution function (PDF) of the time $t = t_{\text{dis.}} - t_{\text{inf.}}$ between infection and disease in the control cohort would correspond to an exponential curve $f(t|r) = re^{-rt}$, from which the average transition time $\langle t \rangle = 1/r$ and its associated variance $\sigma_t = \langle t^2 \rangle - \langle t \rangle^2 = 1/r^2$ can be easily obtained by integrating the moments of the PDF.

However, in the practical context of a clinical trial, the period of measure cannot be arbitrarily extended, which implies that the maximum transition time that can be observed for a subject who was initially infected at $t_{inf}$ is truncated at $t_{max} = T - t_{inf}$, where $T$ stands for the follow-up period of the trial. This situation implies that the integrals needed to obtain the expected value of the transition time need to be truncated as well, which ultimately makes $\langle t \rangle$ to depend itself on $t_{inf}$:

$$\langle t \rangle(t_{inf}) = \frac{\int_0^{T-t_{inf}} t f(t|r) dt}{\int_0^{T-t_{inf}} f(t|r) dt} = \frac{1}{r} - \frac{e^{-r(T-t_{inf})}\,(T - t_{inf})}{1 - e^{-r(T-t_{inf})}} \tag{5.24}$$

Similarly, by truncating the integrals of the second moment of the distribution we can obtain its dependence with the time at infection, $\langle t^2 \rangle(t_{inf}))$, and ultimately derive the corresponding expression for the variance of observed transition times as a function of $t_{inf}$:

$$\begin{aligned} \sigma_t^2(t_{inf}) \quad = \quad & \frac{-\exp(-r(T-t_{\text{inf}}))\left(1 + (r(T-t_{\text{inf}})+1)^2\right) + 2\left(1 + r(T-t_{\text{inf}})\exp(-r(T-t_{\text{inf}}))\right)}{r^2\left(1-\exp(-r(T-t_{\text{inf}}))\right)} \\ & -\frac{1}{r^2} - \frac{(T-t_{\text{inf}})^2 \exp(-2r(T-t_{\text{inf}}))}{\left(1-\exp(-r(T-t_{\text{inf}}))\right)^2} \end{aligned} \tag{5.25}$$

Equations 5.24 and 5.25 describe how observed transition times from infection to disease and their variance are expected to be biased towards lower values as the infections occur later during the trial; simply because the later the infection takes place, the less time available to observe a transition

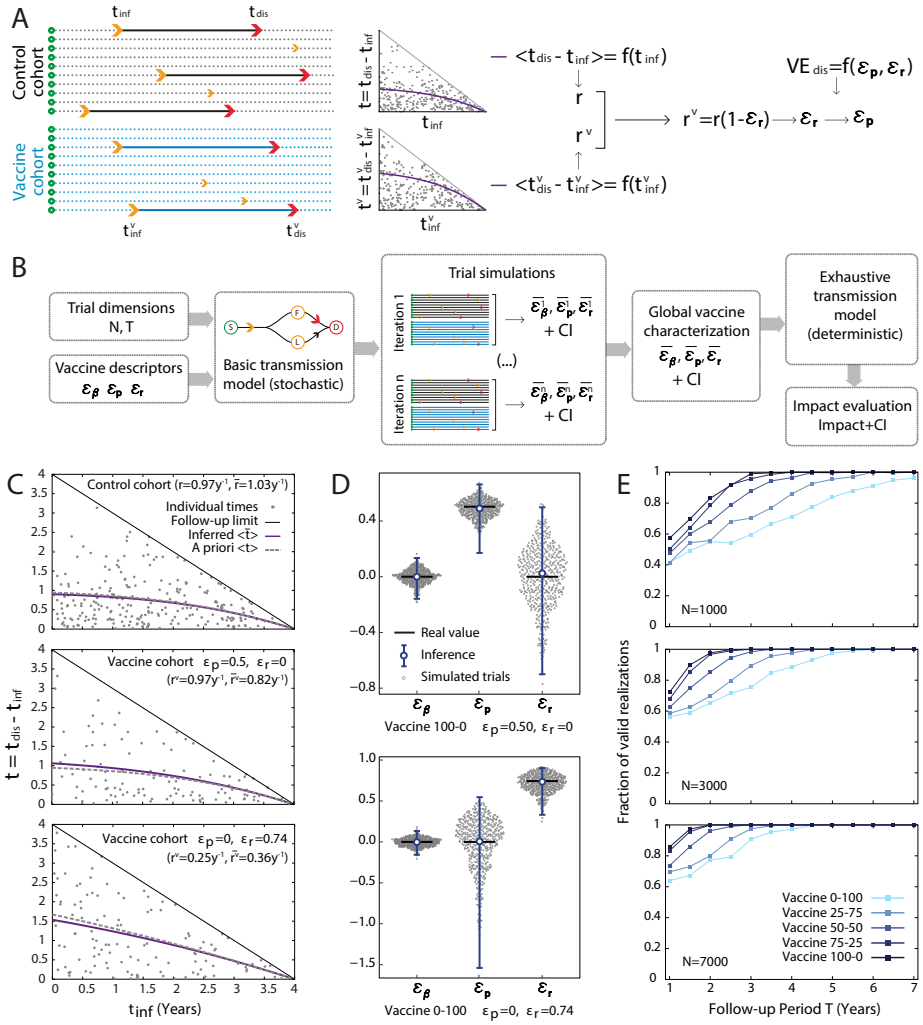


Figure 5.5: **Method for estimating $\varepsilon_r$ and $\varepsilon_p$** A. Scheme of the Method for characterizing vaccine's efficacies. B. Schematic representation of the process of this work. C. Transition times of control cohort (top) and two different vaccinated cohorts (center and bottom). Expectation values in blue for vaccinated cohorts and grey for controls. Analytical predictions in dashed lines. D. Probability density of the inferred parameters of the vaccines ($\varepsilon_\beta, \varepsilon_r, \varepsilon_p$), alongside the inferred parameters (with their respective CIs) for two different vaccines. E: Fraction of correct realizations of a trial as a function of the follow-up period, for three different cohort sizes and 5 different vaccines (same 5 as in Figure 5.1)

to disease is left. These expressions allow us to isolate the effect of that bias, and to infer, using only data from individuals developing active TB during the trial, the transition rate $r$ within the control cohort, via a Maximum Likelihood approach (R package bbmle[210]) along with its confidence intervals (95% reported). Then, the exercise is repeated in the vaccine cohort, whose transition rate $r^v$, in terms of the transmission model would be expressed as the product $r(1-\varepsilon_r)$, which yields the following expression for the vaccine effect on fast progression rate $\varepsilon_r$:

$$\varepsilon_r = 1 - \frac{r^v}{r} \tag{5.26}$$

Finally, we obtain an estimation of the CI for $\varepsilon_r$ by propagating the independent uncertainties of $r^v$ and $r$.

#### 5.2.6.2 Estimation of $\varepsilon_\beta$ and $\varepsilon_p$

The effect of the vaccine on the infection rate, codified in the model as $\varepsilon_\beta$, coincides, by construction, with the vaccine efficacy against infection $VE_{inf}$, and, as such, is inferred using Cox-regression (R package OIsurv[211]).

The third vaccine effect –reduction of fast-progression probability $\varepsilon_p$– can be trivially deduced once the other effects are known. First, we need to calculate the disease-ratio at the end of the trial (see equation 5.17). The ratio will depend only on the parameters of the vaccine ($\varepsilon_\beta$, $\varepsilon_p$ and $\varepsilon_r$) and on the natural parameters of the disease ($\beta$, $p$, $r$ and $r_L$) and on the trial follow-up period $T$. From equation 5.17 we finally obtain $\varepsilon_p$ as follows:

$$\varepsilon_p = \frac{\rho(T)\left(A - pB - (1-p)C\right) - A_v + pB_v + (1-p)C_v}{p\left(B_v - C_v\right)} \tag{5.27}$$

Finally, in order to estimate $\varepsilon_p$ uncertainty, we propagate standard error from $\varepsilon_\beta$ (Cox regression), $\varepsilon_r$ (from Maximum likelihood-based inference of fast-progression rates at control cohort and at vaccine cohort, as explained before) and $\rho$. Regarding the latter source of uncertainty, we obtain the variance associated to the disease-ratio as follows:

$$s^2 = \frac{1 - D_c(T)}{D_c(T)N} + \frac{1 - D_v(T)}{D_v(T)N} \tag{5.28}$$

which yields the following confidence interval:

$$CI = 1 - \exp\left(\ln\left(\frac{D_v(T)}{D_c(T)}\right) \mp z_{1-\frac{\alpha}{2}} s\right) \tag{5.29}$$

where $D_c(T)$ and $D_v(T)$ are the fraction of cases observed at both cohorts during the follow-up period, and $z_{1-\frac{\alpha}{2}}$ is the standard score for the chosen level of significance $\alpha$ (we will use 95%).

#### 5.2.6.3 Testing the method: Synthetic clinical trials simulations

To test the performance of this approach, we used Monte Carlo methods to simulate synthetic clinical trials of different lengths and sizes, for vaccines acting through different mechanisms and showing disparate levels of protection (Figure 5.5, panel B). To these produce synthetic simulations of clinical trials, we first calibrated the baseline parameters of the transmission model to reflect the current epidemic situation in Worcester, South Africa, where the MVA85A study took place.[200] First, the transition rate from LTBI to disease is assumed to be $r_L = 7.5 \times 10^{-4} y^{-1}$, in accordance with previous bibliographical estimations.[55] The probability of fast-progression has been fixed to $p = 0.375$ which is compatible with previous observations about the remarkably high probability of developing fast-progression during the first months of life. The value used in this work is a reference to simulate trials conducted in newborns.

Second, the baseline transmission rate was estimated to be $\beta = 0.069 y^{-1}$ to reproduce the proportion of infections in the control cohort of the MVA85A trial (12.8% after 2 years). Finally, the transition rate from fast latency to disease $r = 0.97 y^{-1}$ was calibrated to reproduce the same proportion of cases than in the control cohort of MVA85A trial (2.3% after 2 years), once all the other parameters are fixed. The fast-progression rate is also compatible with previous observations.[55] Thus, this approach is specific both to the site and age-strata of individuals joining the trial, which means that the inferred baseline parameters cannot be automatically used to simulate trials conducted in other sites, or age-cohorts, although similar re-calibrations are possible upon availability of reference data. Similarly, since BCG vaccination is mandatory in South Africa, the baseline parameters have embedded the eventual protective effects provided by the old vaccine.

Next, we arbitrarily define a vaccine by providing the triad of vaccine efficacies $(\varepsilon_\beta, \varepsilon_r, \varepsilon_p)$, describing its effects on the infection rate, the transition rate to disease, and the probability of fast progression, respectively. While a value of, for example, $\varepsilon_\beta = 0$ means no protection against infection, and $\varepsilon_\beta = 1$ means perfect protection, it is worth stressing that, in principle, vaccines where any of the three $\varepsilon$ parameters is lower than zero are possible,

as it is possible to observe a failed vaccine that indeed increases the risk of infection or disease with respect to the control cohort, as occurred in the MVA85A trial (although not significantly).[200]

Once all the dynamical parameters governing TB transmission dynamics in both cohorts are set up, we use an agent-based approach to simulate the evolution of $N$ individuals per cohort during a follow-up period $T$. We use a discrete time step of 3 months to reproduce the temporal resolution between consecutive analyses (QFT for infection and/or TB diagnosis tests for active TB) in the MVA85A trial.

Thus, we generate a set of vectors of transition times to infection and active TB that summarizes the realization of the trial. Since the model is probabilistic by construction, we iterate to obtain a set of simulated trials describing a distribution of most-likely outcomes associated to such a trial set-up as a function of the a-priori known vaccine behavior. From a given vaccine $(\varepsilon_\beta, \varepsilon_r, \varepsilon_p)$ and trial design $(N, T)$ we simulate 500 equivalent trials that we characterize using the method developed. From the inferred parameters we obtain a global estimation of the parameters and their CIs that we will use to evaluate their impact and their respective uncertainty.

In order to obtain global estimates and confidence intervals for vaccine descriptors we follow a three steps approach. First, we generate a set of 500 synthetic clinical trials for each vaccine analyzed. Second, for each of these simulated trials, we infer the values of the vaccine descriptors $\varepsilon_\beta, \varepsilon_p$ and $\varepsilon_r$ along with their confidence intervals: that of $\varepsilon_\beta$ from Cox-regression, that of $\varepsilon_r$, as explained before, and, finally, that of $\varepsilon_p$ propagated from the other two, and from the CI of the disease ratio $\rho$. Finally, we assume that the true values of these parameters come from an unweighted mixture of normal distributions each of which is associated to the log transform of one minus the outcome of each simulated trial. The final value and CI of each of the three vaccine descriptors is associated to the median and 95% CI of such distribution mixture, back in the linear scale. Through this approach we get a global estimation of the accuracy and precision of the method as a function on the predefined vaccine's characteristics and trial dimensions, which we have introduced in the TB spreading model for the forecasts of vaccine impacts (and their correspondent Confidence Interval).

Then, for each clinical trial realization, we use the method to infer the vaccine descriptors $(\varepsilon_\beta, \varepsilon_r, \varepsilon_p)$. In figure 5.5, panel C we represent three examples of how the transition times associated to a single realization of a

synthetic trial can be used to infer the transmission rates at the control cohort (upper panel) or at the vaccine cohort used, which are compared to infer $\varepsilon_r$. Here, two types of vaccines are considered: an example of a vaccine that prevents fast progression (center, where the inferred transition rates are not significantly different from the controls), and another example where the vaccine reduces transmission times to disease within the fast progression branch (lower panel). Then, once $\varepsilon_r$ is estimated from the analysis of the transition times, we infer the other two parameters as described above to retrieve estimates for the three vaccine descriptors $(\varepsilon_\beta, \varepsilon_r, \varepsilon_p)$. As it can be seen in Figure 5.5D, this closely reproduce the initial values used to simulate the trials. Results correspond to a set of 500 realizations for an example vaccine with $\varepsilon_\beta = 0$, $\varepsilon_p = 0.4$, $\varepsilon_r = 0.4$.

A final question to address concerns the range of applicability of the method. Since the key step of the proposed approach relies on the calibration of the distribution of infection-to-disease progression times, the size and duration of it must be big enough to ensure sufficient statistics. As explained before, there exists a minimum value of $\varepsilon_p$ (that is $\varepsilon_{p,\min} = 1 - \frac{1}{p}$), that generates an upper bound for $\varepsilon_r$ as these two parameters are linked through the disease ratio $\rho$. Notwithstanding this, the inference of $\varepsilon_r$ is agnostic to the value of $\rho$ or $\varepsilon_p$, and, as a consequence, for poor statistical settings –most often in the case of vaccines delaying fast-progression– some individual trial realizations lead to vaccine descriptor estimates that lie beyond these epidemiologically meaningful intervals for parameters $\varepsilon_p$ and $\varepsilon_r$. In figure 5.5E we represent the fraction of stochastic realizations that yield valid inferences of vaccine descriptors, for the same five vaccines represented in figure 5.1E. For a trial with a cohort size of 3000 individuals and follow-up period of 4 years, only a vaccine acting exclusively through $\varepsilon_r$ yields a probability of observing a trial incompatible with this method that surpasses 10%. Those ill-defined realizations are excluded from subsequent analyses.

## 5.3 Results

### 5.3.1 Method's evaluation: uncertainty reduction in impact evaluations

In the previous section, we described how we simulated stochastic realizations of clinical trials of different vaccines, and inferred, from each realization, the vaccine descriptors $(\varepsilon_\beta, \varepsilon_r, \varepsilon_p)$, in a way that is blind to the "real" values

used to generate the synthetic trials. Next, we are interested in quantifying the extent to which impact forecasts might benefit from having access to the specific values $(\varepsilon_\beta, \varepsilon_r, \varepsilon_p)$. To this end, we feed the epidemic model with these parameters, simulating the introduction of the vaccines in Ethiopia in 2025, and estimate impact forecasts in number of TB cases prevented within the period 2025-2050 (Figure 5.6A).

To evaluate the advantages of this approach in this context, we compare the impact forecasts generated using the procedure described above to the most extreme impacts that can be obtained if, after measuring $VE_{inf}$ and $VE_{dis}$ from the synthetic trials, the efficacy against disease is mapped to all sort of possible combinations of values $(\varepsilon_r, \varepsilon_p)$ that are compatible with it, a procedure that could be assimilated to the uncertainty provided by state of the art efficacy evaluation. As it can be seen in figure 5.6A, the degeneration between possible mechanisms underlying protection against disease translates into exacerbated levels of uncertainty that almost systematically prevent the rejection of the null hypothesis of null vaccine impact; an observation that is valid for a wide range of values of $VE_{dis}$.

On the contrary, analyzing simulated trials' data using this method substantially contributes to reduce the associated uncertainty of impact forecasts, in a factor ranging from 23% to 63% in the examples showed. The foreseen impacts showed in Figure 5.6A have two sources of uncertainties involved, those that come from the evaluation of vaccine parameters and those coming from the spreading model itself, except when we evaluate the real underlying vaccine for which we only have the latter. As shown in Figure 5.6A, in the case of vaccines that reduce fast-progression, the inferred parameters of the vaccine introduce a significant increase in the Confidence Interval of the forecasts. This situation is avoided in the case of vaccines that prevent fast-progression from taking place, whose confidence intervals are comparable with the uncertainty that is exclusive from the model.

Finally, we are interested in estimating the probability of registering a successful trial (i.e. a trial associated to a significant estimate in terms of vaccine descriptors and forecasted impact), as a function of both trial size and true, underlying vaccine features. To achieve this, we simulated clinical trials of different sizes and durations for the two extreme vaccines among those represented in figure 5.1E (both of which share the same reference level of protection $VE_{dis}(T = 4) = 50\%$), which act either by reducing proportion of fast progressors $(\varepsilon_r = 0, \varepsilon_p = 0.50)$ or by slowing them down in their path to disease $(\varepsilon_r = 0.74, \varepsilon_p = 0)$. In figure 5.6B, we show, as a function of trials

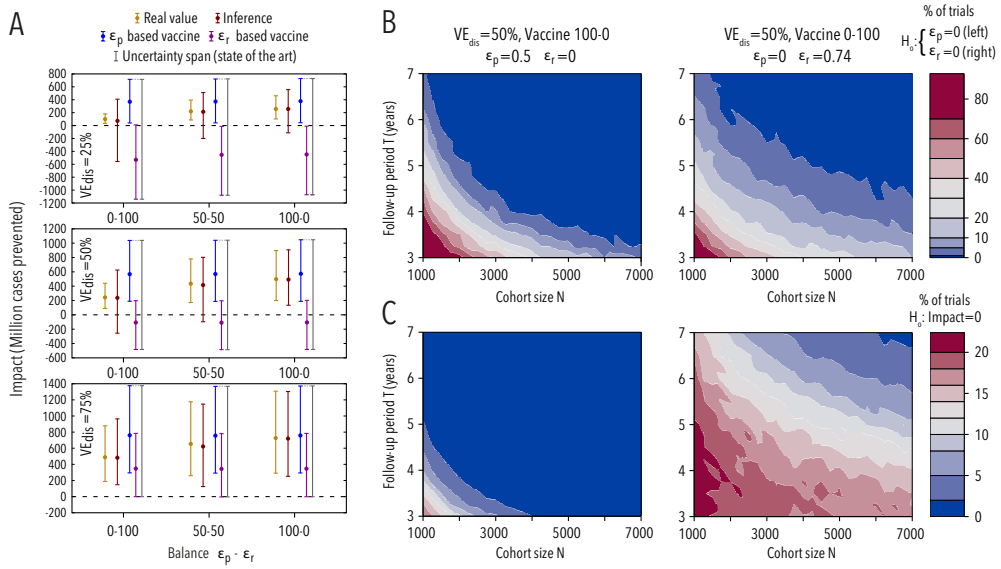


Figure 5.6: **Impact evaluation of synthetic vaccines characterized with the developed method.** A. Real impact (i.e. derive from the real vaccine parameters), inferred impact (derived from the vaccine parameters obtained using the method proposed here) and state-of-the-art uncertainty for 9 different vaccines (3 different levels of efficacy at disease $\times$ 3 different balances between $\varepsilon_p$ and $\varepsilon_r$). In all cases $\varepsilon_\beta = 0$, $N = 3000$ and $T = 4y$. While the real impact carries only the uncertainty related to the spreading model; the inferred impact from this approach also involves the uncertainty given by the characterization of the vaccine. State-of-the-art uncertainty is obtained after simulating the impact with the two points of the curve $(\varepsilon_p, \varepsilon_r)$ that will provide the most extreme impacts. B. Proportion of false negatives in the characterization of the vaccine (i.e. fraction of realizations that leads to a characterization of $\varepsilon_p$ (left) or $\varepsilon_r$ (right) that is not significantly larger than zero at a 95% CI). We study two different vaccines: $\varepsilon_p$-based (left) and $\varepsilon_r$-based (right), focusing on the characterization of the positive effect of the vaccine (i.e. $\varepsilon_p$ and $\varepsilon_r$ respectively). C. Probability of obtaining a positive impact for the two vaccines under study –$\varepsilon_p$-based (left) and $\varepsilon_r$-based (right)–. We assume that the impact obtained and its associated uncertainty correspond to a Normal Distribution with 95% CI.

dimensions, the probability of obtaining a valid trial simulation yielding an inferred value for the corresponding $\varepsilon$ statistically significant (95% CI not crossing zero). In figure 5.6C, we complement these results by characterizing

how likely it is to obtain a clinical trial result associated to a forecast with significant impact.

### 5.3.2 Evaluation of vaccines with protection against infection ($\varepsilon_\beta > 0$)

Throughout this chapter we have explained how we break the degeneracy of two mechanisms capable to offer protection against disease, and quantified the importance of this method in terms of posterior evaluation of vaccination impact. However, as we have discussed, protection against infection can be totally solved with the current methodology. Therefore, for a vaccine that prevents infection the framework presented in this work becomes unnecessary. Although it is to be expected that future vaccines will most likely affect pathways from infection to disease, we have repeated some of the previous analyses with vaccines that act partially against infection (a situation similar to BCG, who offers a small protection against infection[75,192]). More specifically, we include vaccines that are equivalent to those previously explored, in the sense that are compatible with the same measurements of efficacy against disease (i.e. $VE_{dis}(T = 4 \text{ years})$) but incorporating an efficacy against infection given by $\varepsilon_\beta = \frac{VE_{dis}(T=4 \text{ years})}{2}$, meaning that, at a first-order approximation, we can state that half of the protective power of the vaccines showed in this section is given at the infection process.

As shown in Figure 5.7A, including a protective effect on infection evidently lowers the curve of values compatible with the same measure of efficacy at disease. The efficacy at infection of these vaccines poses a common ground of posterior impact that significantly reduces the differences between the five vaccines analyzed on terms of foreseen impact (Figure 5.7B), being the relative difference between the first vaccine ($\varepsilon_\beta = 0.25, \varepsilon_p = 0.35, \varepsilon_r = 0$) and the last one ($\varepsilon_\beta = 0.25, \varepsilon_p = 0, \varepsilon_r = 0.62$) a 29% (26-33) 95% CI, while in the case of $\varepsilon_\beta = 0$ we have shown relative differences of 51% (45-59 95% CI). Also, the differences in the time evolution of $VE_{dis}$ (Figure 5.7C) suffer a reduction with respect to the case with $\varepsilon_\beta = 0$, although remain relevant.

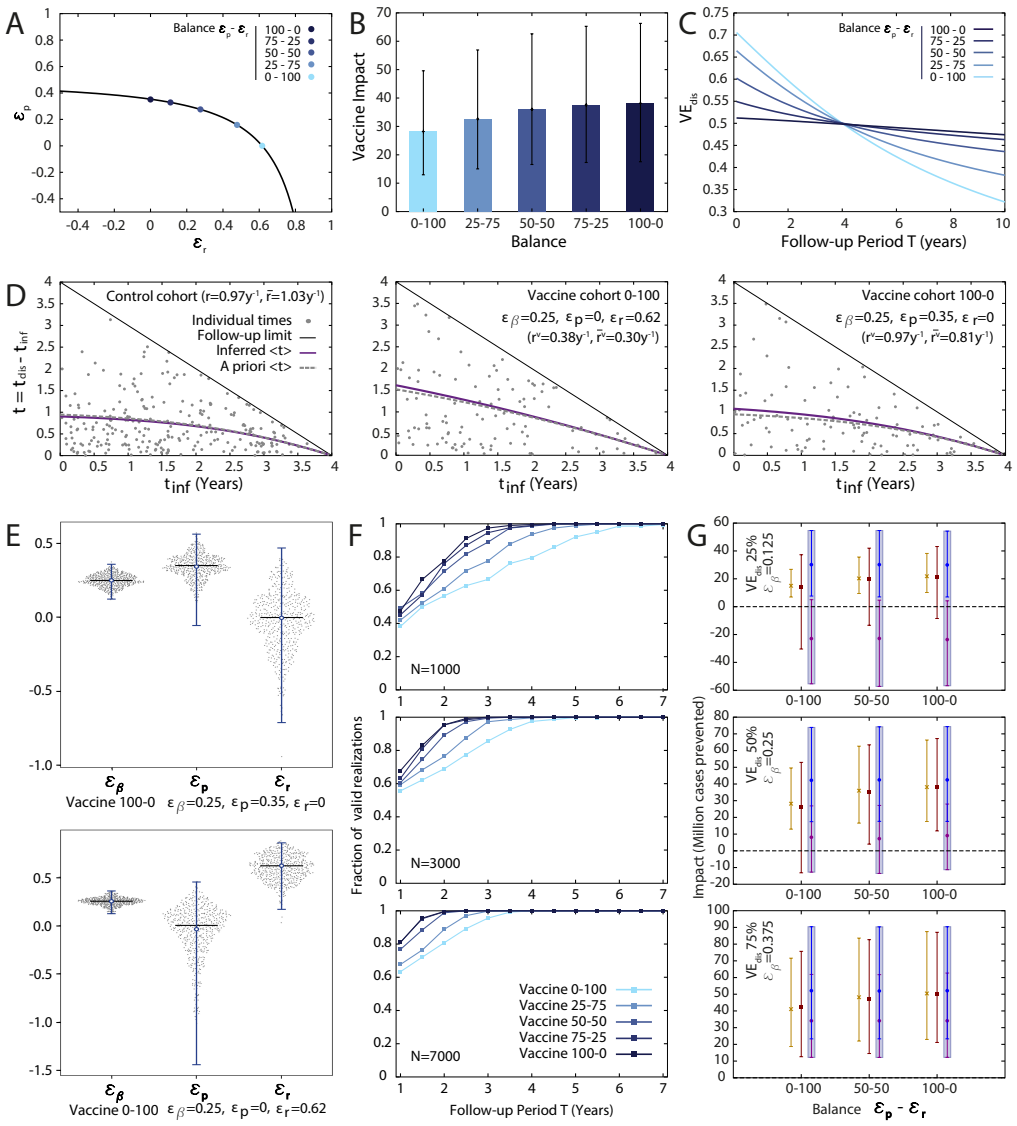


Figure 5.7: **Results for a vaccine with $\varepsilon_\beta > 0$.** A: Curve of values of ($\varepsilon_p$,$\varepsilon_r$) compatible with a measurement of $VE_{dis} = 0.5$ after 4 years of follow-up (assuming $\varepsilon_\beta = 0.25$). B: Predicted impacts obtaining after introducing the 5 highlighted vaccines in a TB spreading model in Ethiopia evaluated in the period 2025-2050. C: Evolution of measurement of $VE_{dis}$ for the 5 highlighted vaccines as a function of the follow-up period. (Continued on next page ...)

Figure 5.7: (... continued) D: Transition times for the control cohort, and the vaccine cohort for the two extreme vaccines considered (balance 0-100 and 100-0). E. Probability density of the inferred parameters of the vaccines ($\varepsilon_\beta, \varepsilon_r, \varepsilon_p$), alongside the inferred parameters (with their respective CI) for two different vaccines. F: Fraction of correct realizations of a trial as a function of the follow-up period, for three different cohort sizes and 5 different vaccines (the 5 vaccines remarked in panel A). G. "Real" impact, inferred impact (using the proposed method) and state-of-the-art uncertainty for 9 different vaccines.

We see in figure 5.7D the individual transition times for a trial realization with two different vaccines. A protective effect at infection will reduce the pool of possible posterior transitions to disease, which might constitute a problem for the inference of vaccine's parameters. However, as in this section we are comparing equivalent vaccines to those studied before (in the sense that they showed the same $VE_{dis}$ after 4 years), these vaccines have a smaller effect in the pathways that protect against disease, so the number of transitions remains, approximately, the same; and this method is able to extract correctly the rates of fast-progression.

As for the distribution of inferred parameters (Figure 5.7E), $\varepsilon_\beta$ is significantly assessed as positive in this examples, as $\varepsilon_r$ in the correspondent vaccine. The fraction of valid realizations (Figure 5.7F) gets improved with respect to their counterparts with $\varepsilon_\beta = 0$ Finally, when we study the improvement associated with this method with respect to the state-of-the-art framework, is only slightly reduced with respect to the examples presented before, ranging from 17% to 59% in the examples showed in Figure 5.7G.

Therefore, even if a significant fraction of a vaccine protective power comes from prevention of infection, the application of this methodology would still be pertinent.

## 5.4 Discussion

Despite TB being one of the diseases for which a vaccine was first developed, there is still much to learn about the mechanisms that TB vaccines, both BCG and novel candidates, unfold to disrupt the pathogen's life cycle. Recent studies have challenged the classical view according to which BCG works solely by priming T-cell mediated adaptive immune responses, proposing more complex models of vaccine function. According to these new views, BCG vaccination is also able to boost host's innate immune system, so

providing an additional layer of non-specific, sustainable protection against infection, (usually referred to as "trained immunity"), which can be observed both in-vivo and in-vitro .[212,213] How these observations might contribute to explain certain epidemiological aspects of BCG's behavior is still unclear; like its ability to reduce *M.tuberculosis* infection risk;[192] its variable protection levels against TB disease,[75] or their correlations with age, latitude and/or previous exposure to mycobacterial antigens,[78,166] or its recently discovered ability to disrupt the transmission chain of TB at other points.[116] In any case, there is increasing scientific evidence suggesting that vaccines confer protection against infection and disease by mediating host-pathogen interactions in ways that, specially in the TB case, are more complex than classically accepted. An additional piece of this intricate puzzle, which we begin to analyze here, is that of understanding the possible repercussions of these improved vaccine descriptions on disease transmission dynamics.

Throughout this chapter, we have described how vaccine protection against disease can manifest in different ways at the epidemiological level, e.g., by slowing down fast progression to disease ($\varepsilon_r$-based vaccines) and/or preventing it completely ($\varepsilon_p$-based vaccines), among others. As we have shown, these putative mechanisms of action cannot be disentangled by applying simple survival analysis. This limitation of standard methodologies when applied to TB is unexpectedly relevant, for we have shown that vaccines relying its protective action on these two mechanisms, even if they might appear as equally effective in the context of a clinical trial of vaccine efficacy, behave in remarkably different ways. As a result, this affects both the feasibility of their characterization as successful vaccines and the ultimate impact on TB burden reduction that they can provide. In this sense, $\varepsilon_r$-based vaccines appear associated to lower prospected impacts for equivalent lectures of vaccine efficacy, and they are harder to analyze, as they would require bigger clinical trials of longer duration in order to be characterized at comparable levels of statistical significance. Therefore, distinguishing between these two different mechanisms of action improves early evaluation of TB vaccine candidates by contributing to a deeper description of vaccines' expected behavior when applied on large populations and, more importantly, by avoiding previously unnoticed biases in their comparison, a crucial aspect in a context where resources are limited and committed only to the most promising candidates at each stage. As a result, considering the questions explored in this study will contribute both to achieve a better interpretation of its outcomes as well as to improve the probability of success of a clinical trial of vaccine efficacy.

The method here proposed is instrumental in a number of situations, but

has intrinsic limitations regarding both the dimensions of the trials that can be analyzed as well as the types of vaccines that can be characterized more accurately by using it. On the one hand, the framework proposed in this work is of application in the most favorable event of a vaccine being able to prevent TB infection ($\varepsilon_\beta$-based vaccine) too. However, as explored in section 5.3.2, when protection against disease is primarily achieved through a reduction of the probability of infection, the considerations made in this work are quantitatively less relevant. In that case, the appearance of additional vaccine effects, either via $\varepsilon_p$ or via $\varepsilon_r$ yields more comparable vaccine behaviors, because the common, main vaccine effect on reducing the infection rate defines the impact expected for the vaccine.

On the other hand, we have simulated trials of cohort sizes comprised between 1000 and 7000 individuals, during follow-up periods between 3 and 7 years. In the limit of short, small trials, low statistics constitutes a fundamental limitation that reduces the chances of obtaining a successful vaccine characterization, even for highly efficient vaccines. In turn, if the trial size and duration has to be increased, while increasing cohort size is always positive, implementing longer follow-ups may require different analytic approaches to those here presented. The reason is that the fundamental hypothesis behind this method is that all individuals progressing to disease during the trial can be associated to immediate progression upon infection: if the trial period is arbitrarily extended, such hypothesis does not hold anymore, and this method for estimating $\varepsilon_r$ and $\varepsilon_p$ would be biased.

In addition, the method of vaccine characterization can in principle be extended to more complex scenarios where more subtle phenomenologies can be considered. This includes to study the effect of vaccine heterogeneity, either within or across age (i.e., immunity waning), as well as to include in the simulated trials the dependence of some of the dynamical parameters on subjects' age. Admittedly, even if such refinements might be relevant from a quantitative stand, other limitations regarding the amount of information that can be extracted from a short clinical trial will most likely persist, regardless of these considerations. This comprises the intrinsic inability to estimate the effects of a vaccine on the slow dynamics of LTBI from a short-course trial, as well as the existence of multiple unavoidable sources of uncertainty that affects the precision of model-based impact and cost-effectiveness evaluations, independently on how precise the efficacy of the vaccine itself is estimated from a trial.

Despite these limitations, the methodology presented in this work is in-

strumental for designing clinical trials that are more likely to succeed in the characterization of novel TB vaccines – and for providing a deeper characterization of them–, able to reduce uncertainty in impact forecasts evaluations.

In the next final Chapter we will finally assess the impact of different vaccines, focused on different age-groups.

# Chapter 6

# Impact evaluation of novel TB vaccines

## 6.1 Introduction

As has been explained throughout this thesis, it is an unsolved problem how to determine the performance of a vaccine at its first development steps, due to the lack of immune correlates of protection against TB disease.[214] The field has already suffered the failure of the promising vaccine candidate MVA85A, that despite achieving great successes at the clinical level and in animal models,[215–221] failed to provide protection against infection or disease at an efficacy clinical trial.[200,201]

In this context, transmision modeling is crucial, for models will be the ultimate tool to quantify the impact expected from a given vaccine and to compare it with possible alternative candidates. It is crucial to note that an efficacy estimate emanating from a clinical trial is not the only readout of a vaccine defining its impact, as we have shown in Chapter 5. In other words, not always the vaccine with the largest trial-estimated efficacy will translate into a higher impact, or viceversa.

In TB, several additional factors affect vaccines' impact in a crucial manner, beyond the intrinsic efficacy of the vaccine. The first one is the age strata on which the immunization takes pace. Among the different candidates we can identify two different groups: prime vaccines, designed to be replacements of the current BCG vaccine;[117] and boosters, enhancers for the immunogenic power of BCG.[79] As we have explained previously, at least in principle, these two groups of vaccines are intended to be administered at dif-

ferent ages: while prime vaccines would be inoculated in newborns as BCG is, boosters would be used at adolescence, when BCG loses its protective effects. This is of utmost importance, since age-distributions of pulmonary TB incidence and prevalence of infectious individuals are strongly age-dependent, with adolescents and young adults having much larger burden levels than infants, as shown in Chapter 3. This translates into the fact that, a priori, a vaccine applied on adolescents will be expected to provide a larger impact than an equivalent vaccine with the same efficacy, applied on newborns, a result already confirmed by models.[101] To that, also contributes the fact that now, BCG offers few to none protection against TB in those age groups (it wanes before that) and, since the impact of a new vaccine is to be measured in comparison with BCG, the old vaccine is much easier to overcome on adolescents/young adults than on children.

However, that general observation is affected by several critical aspects that should be explicitly considering for compared prime-vs-booster vaccines. First, achieving similar levels of vaccine coverage is much harder in adolescents than in newborns.[222] Second, previous sensitization to environmental mycobacteria (which will be greater in adolescents than in newborns) can affect the performance of a vaccine. As was shown in Chapter 4, for BCG this environmental sensitization works by blocking the effect of the vaccine.[78,166] Third, new vaccines might suffer from waning of immunity,[223] which is hard to be determined in an early-phase trial. The waning of immunity is a key parameter to determine the impact and cost-effectiveness of a vaccine.

The proper consideration of all these aspects (which is essential to guide unbiased decision making in the future), needs comprehensive spreading models able to provide, very specially, a description of the ways through which TB spreading is coupled to age-structure. In Chapter 3, we developed a first approximation of such model,[129] including evolving demographic structures (following predictions) and empirical age-dependent contact patterns (properly coupled to the evolving demographic structures[224] as explained in Chapter 2). In Chapter 5 we expanded it to include the evaluation of new vaccines.

Here we use this model to explore how these aspects might impact the evaluation of novel vaccines, either boosters applied on adolescents or prime aimed at substituting BCG. That supposes the first framework for vaccine evaluation that considers to full detail the effects of populations' age structure and its evolution.

## 6.2 Methods

### 6.2.1 Modification of the transmission model

We use the transmission model developed in Chapter 3 and expanded in Chapter 5 to the problem of evaluations of epidemiological interventions. This expansion consisted in adding a second vaccinated branch with the exact same Natural History but with a modification in the epidemiological parameters produced by the vaccine. Thus for every class $X$ that we have for unvaccinated individuals (see table 3.6 in Chapter 3), we have a class $X^v$ corresponding to vaccinated individuals that are in the same stage of the disease and each parameter $x$ in the vaccinated branch transforms to $(1-\varepsilon_x)x$ (where $\varepsilon_x$ represents the effect of the vaccine to that particular epidemiological parameter $x$).

For this work we will only consider vaccines that prevent disease by modifying the probability of fast-progression (parameter $p$). Thus we will have $\varepsilon_p > 0$ and $\varepsilon_x = 0 \forall x \neq p$. Although having a vaccine that also protects against infection could be very useful, and it is known that BCG offers some protection agains infection (even though it is not its main effect),[75] in this chapter I will only focus on the study of vaccines that provide protection against disease, which is considered the main mechanism of action in BCG.[225] In Chapter 5 we also discussed the possibility of having two different pathways to prevent disease (the probability of fast-progression and the rate from infection to disease), and how they could be measured independently. For the sake of simplicity we have focused on only one interrupting mechanism from infection to disease, in this case the probability of fast-progression $p$.

In Chapter 5 we only included newborn vaccination, and a coverage of 100%. In that case, after 2025 (which we consider as the date the vaccination campaign begins) all newborns are added to the vaccinated branch. Now we expand this scenario to include an arbitrary coverage level $c \in [0,1]$ and adolescent vaccination. For newborn vaccination the modification is trivial as we only have to divide the flux of newborns between both branches (vaccinated and unvaccinated) according to the coverage $c$. However, due to mother-child transmission of the disease, we also introduce newborns to the latency states. Thus, when newborn vaccination is considered, individuals are introduced in the system at 6 different states, $S$, $L_f$, $L_s$, $S^v$, $L_f^v$ and $L_s^v$, according to the following fluxes:

- Birth of $S(0,t)$ individuals (susceptible newborns, unvaccinated): $(1 - c_{\text{newb.}}(t))(1 - m_c m_d(t))\Delta_N(a=0,t)$

- Birth of $L_f(0,t)$ individuals (infected after birth who develops fast progression, unvaccinated): $(1-c_{\text{newb.}}(t))m_c m_d(t)p(0)\Delta_N(a=0,t)$
- Birth of $L_s(0,t)$ individuals (infected after birth who develops slow progression, unvaccinated): $(1-c_{\text{newb.}}(t))m_c m_d(t)(1-p(0))\Delta_N(a=0,t)$
- Birth of $S^v(0,t)$ individuals (susceptible newborns, vaccinated): $c_{\text{newb.}}(t)(1-m_c m_d(t))\Delta_N(a=0,t)$
- Birth of $L_f^v(0,t)$ individuals (infected after birth who develops fast progression, vaccinated): $c_{\text{newb.}}(t)m_c m_d(t)(1-\varepsilon_p)p(0)\Delta_N(a=0,t)$
- Birth of $L_s^v(0,t)$ individuals (infected after birth who develops slow progression, vaccinated): $c_{\text{newb.}}(t)m_c m_d(t)(1-(1-\varepsilon_p)p(0))\Delta_N(a=0,t)$

where:

$$c_{\text{newb.}}(t) = \begin{cases} 0 & \text{if } t < 2025 \\ c_{\text{newb.}} & \text{if } t \geq 2025 \end{cases} \tag{6.1}$$

Equation 6.1 simply states that the vaccination campaing (with a coverage that we will specify later) begins at 2025, therefore before that year the coverage is zero and no individual enters in the vaccinated states.

Adolescent vaccination is coupled to the aging process. When unvaccinated individuals pass from $a = 2$ (10-15 years old) to $a = 3$ (15-20 years old), a fraction of them (given by the coverage) are moved to the vaccinated branch. We also consider that only susceptible individuals receive vaccination (i.e., we are modeling pre-exposure vaccines). For other type of interventions such as post-exposure vaccines, treatment of LTBI, etc., we should consider vaccination in other classes as well. Thus, adolescent vaccination can be reduced to the following fluxes:

- Transition from $S(a=2,t)$ to $S(a=3,t)$ (Non vaccinated adolescents): $(1-c_{\text{adol.}}(t))S(a=2,t)/\Delta_t$ individuals/unit time.
- Transition from $S(a=2,t)$ to $S^v(a=3,t)$ (Vaccinated adolescents): $c_{\text{adol.}}(t)S(a=2,t)/\Delta_t$ individuals/unit time.

where, as for newborn vaccination:

$$c_{\text{adol.}}(t) = \begin{cases} 0 & \text{if } t < 2025 \\ c_{\text{adol.}} & \text{if } t \geq 2025 \end{cases} \tag{6.2}$$

The rest of aging process remain as described in Chapter 3, both for the vaccinated and unvaccinated branches.

### 6.2.2 Coverage Data

Although we have generalized the modeling framework to any possible value of the coverage, it is still interesting to perform simulations with a perfect coverage of 100%, as this scenario represents the maximum impact we can extract from a vaccination campaing in an ideal situation.

But, of course, we are also interested on how different levels of coverage might compromise the performance of a vaccine and, specially, the comparison between adolescents and newborns campaings, as it is expected that adolescent vaccination will have lower coverage values. When establishing reasonable coverage values, we assume that a newborn vaccination campaign would have a similar coverage than BCG. Thus, we will use the data on BCG coverage, as extracted from WHO database.[149] For adolescents we follow the assumption made by Knight et al.[101], and we take as coverage values the school attendance rates. In table 6.1 we show the values for the countries discussed for this work.

| **Country** | **Newborn coverage (%) $c_{\text{newb.}}$** | **Adolescent coverage (%) $c_{\text{adol.}}$** |
|---|---|---|
| India | 89 | 72 |
| Nigeria | 64 | 49 |

Table 6.1: **Reference coverage values.**

As seen in table 6.1, for both countries we consider a higher coverage in newborn than in adolescent. In order to discuss the importance of achieving high coverage rates in adolescents, we will also implement lower values than these reference values, which are highly speculative anyway.

### 6.2.3 Update of the risk of progressing to disease after infection

In a recent paper by Andrews et al.[226] it was established that in the case of reinfection, the risk of progressing to disease is reduced by a 79%. We

incorporate this new information by changing the parameter $q$ (see Chapter 3) to $q = 0.21(0.14 - 0.30)$C.I.. The model has been recalibrated with this new parameter, and all results shown in this chapter incorporate this update.

Besides, this value serves as an upper bound of the expected efficacy for a vaccine, as no vaccine will provide better protection than the actual infection. Thus, we will consider an initial efficacy of 80% for the vaccines we modelize (see section 6.2.5).

### 6.2.4 New reduced model

In Chapter 3 we defined two reduced versions of the model for TB spreading, in order to compare the results obtained with the complete model with those obtained in more simplistic frameworks:

- Reduced model 1 (RM1): Considers the demographic structure as constant in time, but maintains empiric age-dependent contact patterns.
- Reduced model 2 (RM2): Considers homogenous mixing, but maintains the evolution of demography.

For simplicity in this chapter we will only compare results with one reduced model that combines both classical assumptions:

- Reduced model 3 (RM3): Considers the demographic structure as constant and homogeneous mixing between age-groups.

### 6.2.5 Vaccine descriptions

A critical point in the modelization of new TB vaccines is the fact that we already have a vaccine, BCG, whose use is widely spread. Therefore the population we consider, and we have been modeling in Chapter 3 also, has the effect of BCG embedded in their epidemiological parameters.

Thus, in order to implement a new vaccine we should consider the masking effect produced by BCG. This is similar to the masking effect described in Chapter 4, but in this case previous protection does not come from environmental sensitization but from BCG. Therefore we do not rule out this effect, as we know that BCG offers protection to some extent.

For example, if we have a new vaccine that offers exactly the same protection as BCG, in a clinical trial we will measure an efficacy value of zero,

which is exactly what we should incorporate in the model, as the unvaccinated branch is not really unvaccinated but vaccinated with BCG. Or in a more extreme situation, if we have a vaccine with an efficacy that is lower than that of BCG, its *effective* efficacy will be negative (and it will have a negative effect on the population if we replace BCG with that new vaccine), even if its real protective power is positive.

In order to correct for BCG-masking, we first should establish what is the immunity offered by BCG. In Chapter 4, alongside the parameters of masking and blocking, we deduced two values for the intrinsic efficacy of BCG, one at the age of 4.5 years $\varepsilon_{\text{BCG}}(4.5) = 0.577$ and another at the age of 16 $\varepsilon_{\text{BCG}}(16) = 0.376$.[78] We modelize the waning of immunity in a vaccine with an exponential:

$$\varepsilon(t) = \varepsilon_0 \exp(-\gamma t) \tag{6.3}$$

Thus, a vaccine will be described by two parameters, the initial efficacy $\varepsilon_0$ and the decay rate (the immunity waning) $\gamma$. For the case of BCG we can easily obtain these parameters using the two points that we know, which give us $\varepsilon_{0,\text{BCG}} = 0.698$ and $\gamma_{\text{BCG}} = 0.087$ (i.e. a 8.7% annual decay).

In this work we will consider three different vaccines, whose parameters are shown in table 6.2. First we will consider an ideal vaccine with no waning, then a vaccine with a weak annual waning of 1%, and finally a vaccine with a stronger waning effect (5%) but still less than BCG (we will refer to this latter vaccine as reference vaccine). All vaccines will present an initial efficacy of 80%.

| **Vaccine name** | $\boldsymbol{\varepsilon_{0,\text{new}}}$ | $\boldsymbol{\gamma_{\text{new}}}$ |
|---|---|---|
| Ideal | 0.800 | 0.00 |
| Weak waning | 0.800 | 0.01 |
| Reference | 0.800 | 0.05 |

Table 6.2: **New vaccines' parameters.** The real efficacy of new vaccines will be given by $\varepsilon_{\text{new}}(t) = \varepsilon_{0,\text{new}} \exp(-\gamma_{\text{new}} t)$

The effective efficacy $\bar{\varepsilon}(t)$ will be given by:

$$\bar{\varepsilon}(t) = \frac{\varepsilon_{\text{new}}(t) - \varepsilon_{\text{BCG}}(t)}{1 - \varepsilon_{\text{BCG}}(t)} \tag{6.4}$$

In Figure 6.1 we show the real efficacies for BCG and the hypothetical new vaccines considered, as well as the effective efficacy of the new vaccine when compared to BCG.

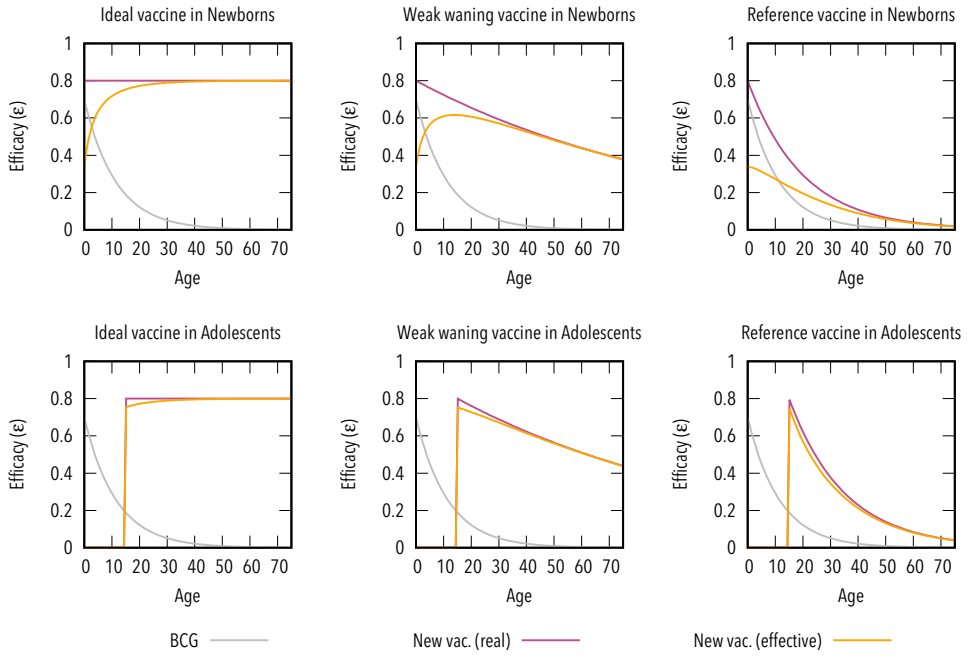


Figure 6.1: **Immunity profiles offered by hypothetical new vaccines.** Values of the efficacy of BCG and the hypothetical new vaccine (real and effective). We show the vaccines considered in table 6.2, applied on newborns and adolescents. BCG is the same in all cases.

In equation 6.3 the time variable $t$ stands for the time since the vaccine is inoculated. Thus, for adolescent vaccination this time starts at 15 years old.

The age description in the model is discrete, thus we need to provide discrete values for the efficacy for each age-group. We assign the efficacy corresponding to the medium age of each age-group (2.5 years old for $a = 0$, 7.5 years old for $a = 1$, and so on).

Eventually we will also consider the possibility of having a blocking effect (due to environmental sensitization or BCG itself) that compromises the efficacy on adolescent vaccination. In that case, the efficacy experienced by

the immunized individuals is reduced by a factor $1-b$ where $b$ is the blocking level considered.

### 6.2.6 Measurement of impact for epidemiological interventions

As the ultimate goal is to evaluate and compare epidemiological interventions, in this case new vaccines, we need to establish a quantitative measurement of the impact of the aforementioned interventions. An immediate possibility could be to measure the reduction of new TB cases (that we already used in Chapter 5). Thus, we define the impact in incidence, $\mathcal{I}_I(t)$ as the difference in TB cases (or deaths) between the baseline and the scenario with the new vaccine that we compute at the end of a certain follow-up period $t$:

$$\mathcal{I}_I(t) = I_{\text{Baseline}}(t) - I_{\text{NewVac}}(t) \tag{6.5}$$

where $I(t)$ is the acumulated number of TB cases measured from the start of the vaccination campaign in 2025 until year $t$. The subindex Baseline refers to the simulation in which no new vaccine is added, while the subindex NewVac stands for the scenario with the new vaccine.

In some cases, instead of the absolute impact (total number of cases or deaths prevented) it might be more useful to compare relative impacts $\mathcal{F}$, fraction of cases or deaths prevented as given by:

$$\mathcal{F}_I(t) = \frac{\mathcal{I}_I(t)}{I_{\text{Baseline}}(t)} \tag{6.6}$$

As we are specially interested in the comparison between adolescent and newborn vaccination, we will also use the relative difference between impacts $\mathcal{RD}$ :

$$\mathcal{RD}_I(t) = \frac{\mathcal{I}_I^{\text{adol.}}(t) - \mathcal{I}_I^{\text{newb.}}(t)}{\mathcal{I}_I^{\text{adol.}}(t)} \tag{6.7}$$

In the last years, the measurement unit disability-adjusted life years (DALYs)[227–230] is growing in popularity.[231,232] DALYs can be understood as the total number of years that a disease is taking from the people. It is the sum of two different contributions, the years of life lost due to death (YLL) and the years lost due to disability (YLD), as it is considered that while the disease is attacking, life is not enjoyed at its full:

$$\text{DALY} = \text{YLL} + \text{YLD} \tag{6.8}$$

The Years of Life Lost (YLL) are defined as:

$$\text{YLL} = M \times L \tag{6.9}$$

where $M$ is the number of deaths and $L$ is the standard life expentancy at the age of death (in years). Thus, in simulations we compute the evolution of YLL as the sum of all TB-related death fluxes (during active disease or after treatment, vaccinated or unvaccinated), weighted by the life expentancy of the age-group where those deaths are ocurring, $L(a,t)$:

$$\dot{\text{YLL}}(t) = \sum_a L(a,t)\left(\dot{M}(a,t) + \dot{M}^v(a,t)\right) \tag{6.10}$$

The evolution of $\dot{M}(a,t)$ can be found in Chapter 3. The values of $L(a,t)$ change with time (life expentancy increases as demographic structure ages). Fortunately, data on $L(a,t)$ are given by the UN population database,[138] which is the same source we use for data on demographic projections.

The Years Lost due to Disability (YLD), are defined as:

$$\text{YLD} = P \times DW \tag{6.11}$$

where $P$ is the prevalence of the disease, and $DW$ is the disability weight. In this case we consider as the prevalence of the disease the sum of all classes with active disease (states $D$ and $T$), so we can compute the evolution of YLD simply as:

$$\begin{aligned}\dot{\text{YLL}}(t) \quad = \sum_a DW( \quad & D_{p+}(a,t) + D_{p-}(a,t) + D_{np}(a,t) \\ & + T_{p+}(a,t) + T_{p-}(a,t) + T_{np}(a,t) \\ & + D^v_{p+}(a,t) + D^v_{p-}(a,t) + D^v_{np}(a,t) \\ & + T^v_{p+}(a,t) + T^v_{p-}(a,t) + T^v_{np}(a,t))\end{aligned} \tag{6.12}$$

The value for the Disability Weight (DW) is taken from the study by Salomon et al.[233] as 0.331. It is assumed to be the same for all types of TB as no further data is available.

Thus, the same definitions of impact (absolute and relative) that we have applied to the incidence, can be applied to DALYs:

$$\mathcal{I}_{\text{DALYs}}(t) = \text{DALYs}_{\text{Baseline}}(t) - \text{DALYs}_{\text{NewVac}}(t) \tag{6.13}$$

$$\mathcal{F}_{\text{DALYs}}(t) = \frac{\mathcal{I}_{\text{DALYs}}(t)}{\text{DALYs}_{\text{Baseline}}(t)} \tag{6.14}$$

$$\mathcal{RD}_{\text{DALYs}}(t) = \frac{\mathcal{I}^{\text{adol.}}_{\text{DALYs}}(t) - \mathcal{I}^{\text{newb.}}_{\text{DALYs}}(t)}{\mathcal{I}^{\text{adol.}}_{\text{DALYs}}(t)} \tag{6.15}$$

## 6.3 Results

### 6.3.1 Basal comparison adolescent vs newborn vaccination

We simulate the introduction of the vaccines described in table 6.2 and Figure 6.1 in India and Nigeria in 2025. We will begin the comparison with the ideal situation of having a coverage of 100% and no blocking effect. Results are shown in Figure 6.2, as well as Table 6.3.

| **Endpoint** | **Waning** | **Incidence** | | | **DALYs** | | |
|---|---|---|---|---|---|---|---|
| | | Newb. | Adol. | $\mathcal{RD}$(%) | Newb. | Adol. | $\mathcal{RD}$(%) |
| 2050 | 0% | 9.9 (2.8-19.7) | 14.2 (5.3-20.2) | 30.3 (-5.9-49.6) | 65 (24-113) | 90 (44-111) | 27.4 (-9.9-47.2) |
| | 1% | 8.5 (2.3-17.0) | 13.1 (4.9-18.5) | 35.1 (0.5-53.5) | 56 (20-98) | 83 (41-103) | 32.6 (-2.9-51.4) |
| | 5% | 4.2 (1.1-8.6) | 9.5 (3.5-13.5) | 56.1 (30.0-69.6) | 28 (10-51) | 62 (30-78) | 54.8 (28.3-68.4) |
| 2100 | 0% | 91.2 (27.1-138.2) | 104.6 (38.0-136.0) | 12.8 (-10.3-27.3) | 474 (167-683) | 517 (222-651) | 8.3 (-16.2-23.3) |
| | 1% | 72.5 (20.8-111.9) | 90.4 (32.4-118.9) | 19.8 (-2.8-34.2) | 386 (132-567) | 458 (195-584) | 15.6 (-8.3-30.5) |
| | 5% | 26.5 (6.8-44.2) | 53.6 (18.6-73.7) | 50.5 (32.7-61.6) | 153 (47-240) | 294 (122-387) | 48.0 (29.4-59.4) |

Table 6.3: **Impact evaluation in India at 2050 and 2100 with different immunity waning.** We present the impact (in millions) in incidence and DALYs (eqs. 6.5 and 6.13) for newborn and adolescent vaccination beginning in 2025, as well as the relative difference between both (eqs. 6.7 and 6.15)

The vaccine with no waning, an initial efficacy 80%, full coverage and no blocking (continuous lines in figure 6.2) corresponds to the maximum impact we can obtain from a single vaccination campaign. Even though the impact obtained is quite significant, eradication is not achieved during this century.

Vaccination in adolescents offers a greater and most immediate impact than vaccination in newborns, a result that was already seen by previous models.[101] However, the distance between both strategies grows shorter as time goes by, specially when the waning is low and the immunity offered

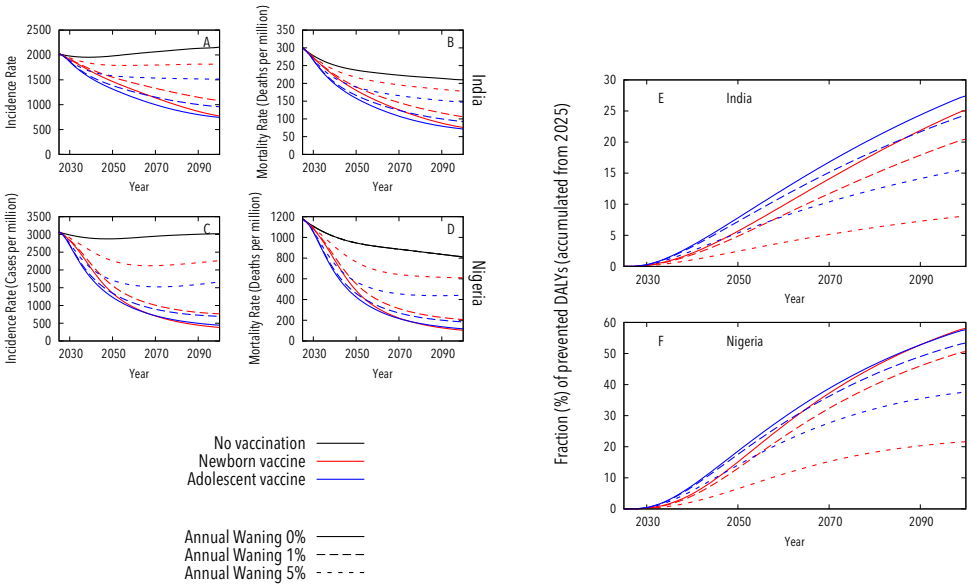


Figure 6.2: **Basal comparison adolescent vaccination vs newborn vaccination.** A-D: Incidence and Mortality Rates vs Time for India (A-B) and Nigeria (C-D) obtained after no vaccination and after applying the vaccines described in table 6.2 and Figure 6.1 at newborns or adolescents. E-F: Fraction (in %) of DALYs prevented (eq. 6.14) vs Time after the inoculation of the different vaccine campaigns in India (E) and Nigeria (F). Vaccines considered in this figure have a coverage of 100% and no blocking.

by the vaccines reaches the eldest age strata. In the case of Nigeria we can even see an inversion for the ideal case of no waning: if the time window of impact evaluation is long enough, newborn vaccination offers better performance than adolescent vaccination.

The reasons underlying these differences between newborn and adolescent vaccination become clearer when we analyze the age distribution of the impact provided by those vaccines instead of remaining at the age-aggregated level.

### 6.3.2 Age distribution of vaccine impact

In Figure 6.3 we show the accumulated impact from the start of the vaccination campaign (either newborn or adolescent) to different endpoints, for

the different age groups of the model.

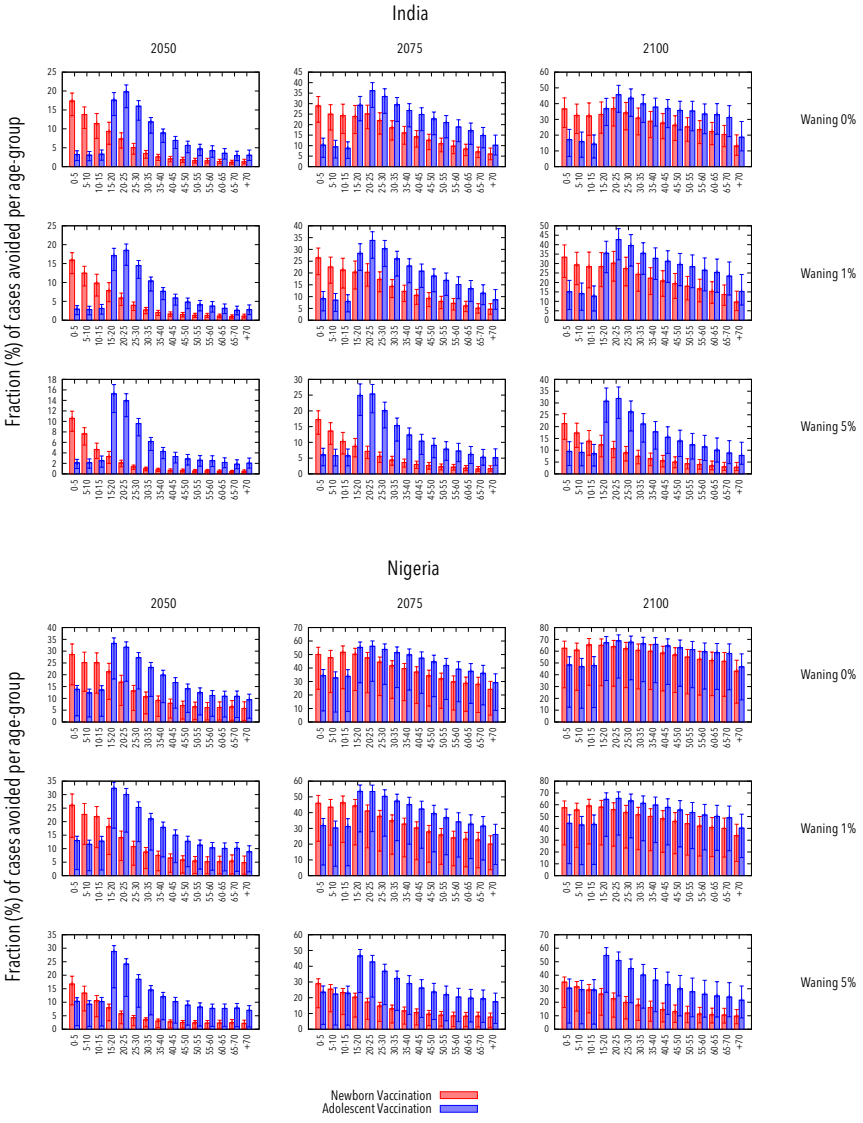


Figure 6.3: **Age distribution of vaccine impact.** Accumulated impact (measured as fraction of cases prevented, eq 6.6) at different endpoints (2050, 2075 and 2100) for the different endpoints. Results are shown for India and Nigeria, the three levels of waning considered (0%, 1% and 5%) and both vaccination strategies (adolescents or newborns)

Without immunity waning a newborn vaccine will eventually cover all the population as time goes by, significantly improving its performance on the adult age strata as we expand the observation window. However, an adolescent focused vaccine will always leave a reservoir of susceptible children. Thus, for this case, if the observation window is long enough, newborn vaccination can offer a greater impact than adolescent vaccine. But when the immunity offered by the vaccine wanes with time, if we vaccinate only newborns we might lose the effect of the vaccine before it reaches the age strata that are most affected by the disease, thus in this case to vaccinate adolescent would be a better strategy.

The situation in both countries analyzed, India and Nigeria, presents significative differences. While in India an adolescent vaccination only provides minimal protection in the younger age strata (less than 5% in 2050), in Nigeria the same vaccination campaign is capable to protect children almost to the same levels as a newborn focused vaccine.

### 6.3.3 Effect of demographic evolution and contact patterns

In order to study the effect of demographic evolution and heterogeneous contact patterns, the novel ingredients added by this thesis, in the evaluation of vaccines' impact we repeat the same analysis with a reduced model (RM3) in which both features are neglected (i.e., demography is constant and mixing is homogeneous). Figure 6.4 is the analogous to Figure 6.2 and Figure 6.5 is the analogous to Figure 6.3.

When impact of new vaccines is evaluated with the reduced model, results change significantly. In this case newborn vaccination have in most cases a much better performance than the adolescent counterparts, that is only able to outperform newborn vaccination for the highest level of waning considered (5%).

In this model, that assumes homogeneous mixing, adolescents stop being the most socially active individuals, thus immunizing them does not have the same effect on stopping transmission. Besides, we are not considering the aging of the population in this model. Therefore, the youngest age strata will not lose density of individuals, and will remain as the most populated age-groups. Thus, in this reduced model, newborn vaccination shows a higher effective coverage than in the complete model that presents a reduction with

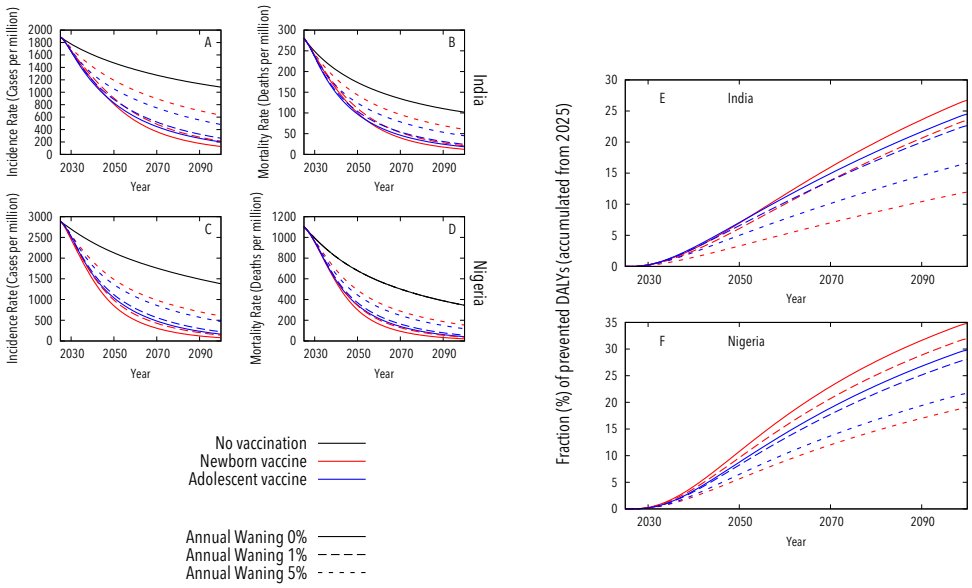


Figure 6.4: **Basal comparison adolescent vaccination vs newborn vaccination with RM3.** A-D: Incidence and Mortality Rates vs Time for India (A-B) and Nigeria (C-D) obtained after no vaccination and after applying the vaccines described in table 6.2 and Figure 6.1 at newborns or adolescents. E-F: Fraction (in %) of DALYs prevented (eq. 6.14) vs Time after the inoculation of the different vaccine campaigns in India (E) and Nigeria (F). Vaccines considered in this figure have a coverage of 100% and no blocking.

time in the birth rate.

Even when these differences between models can be easily traced back to the changes in the underlying assumptions, the results are quite puzzling when we explore past bibliography. Knight et al.[101] show in their model a clearly greater impact in adolescent vaccination than in newborn vaccination, even for life-long immunization, despite not having in their model heterogeneity in contact patterns or population aging. One possible explanation to this phenomenon is the description of the vaccination process itself. In Knight et al.[101] they consider a vaccine that prevent disease by cutting the fluxes from latency to active disease, that is applied in infected and uninfected individuals. In this case we consider only pre-infection vaccines, as we argue that a vaccine probably will not offer more protection than latent infection. Thus, the coverage in adolescents will be much greater in the work

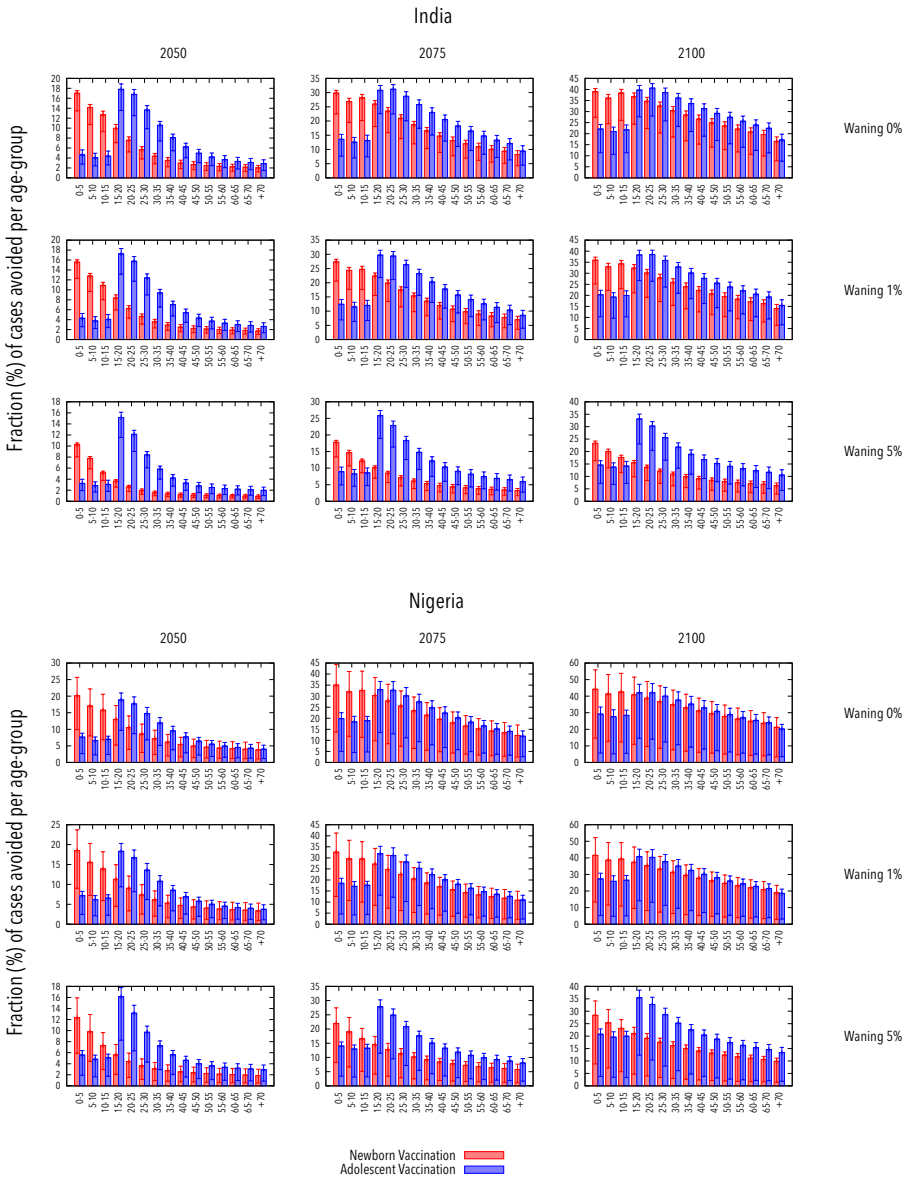


Figure 6.5: **Age distribution of vaccine impact obtained with RM3.** Accumulated impact (measured as fraction of cases prevented, eq 6.6) at different endpoints (2050, 2075 and 2100) for the different endpoints. Results are shown for India and Nigeria, the three levels of waning considered (0%, 1% and 5%) and both vaccination strategies (adolescents or newborns)

by Knight et al.[101] than in the situation considered here. In high burden settings prevalence of infection can reach 50% at 15 years of age[234] so this distinction can heavily influence results.

### 6.3.4 Influence of coverage and blocking

Even when for expected values of waning (from now on we assume 5% as reference value) the vaccination campaign focused on adolescents obtains greater impact, there are two aspects that will worsen the performance of adolescent vaccination and have not been considered yet.

To begin with, we have considered until now an ideal coverage of 100%. However, as we have discussed earlier, coverages from real vaccination campaigns are rarely close to that number. We will implement now the coverage values from Table 6.1 as reference. We will also implement lower values of the coverage in the adolescent campaign, to explore worst-case scenarios. Thus, we repeat the analysis from Figure 6.2, with the new values of coverage considered, in Figure 6.6.

With a non-negligible waning and a more realistic coverage, prospects of the impact of new vaccines are more pessimistic (notice that the y-axis on panels A-D does not start in zero anymore). In this case, we are not only far from eradication, but also vaccines can not prevent the rebound in TB burden that is caused by the aging of populations (as we discussed in Chapter 3). For the reference values of coverage, adolescent vaccination works better than newborn, but this tendency can be reversed if an adolescent campaign is not able to reach those levels of coverage.

Another effect we must take into account is the possible existence of blocking (as defined in Chapter 4, it is the effect of having an underperformance of the vaccine caused by previous sensitization to environmental mycobacteria), that won't appear in newborns but can compromise an adolescent vaccination campaign. Starting from the reference coverage values and a waning of 5%, we repeat once again the analysis from Figure 6.2 with different values of blocking in Figure 6.7.

In Chapter 4 we established that the blocking levels suffered by BCG at school age are 41.1% in Salvador and 96.4%. in Manaus. In this case, we will explore a 20% of blocking alongside the 41.1% that we have in Salvador. The different levels of blocking suffered by the adolescent vaccine are analogous to having different penalizations on coverage, thus results are very similar

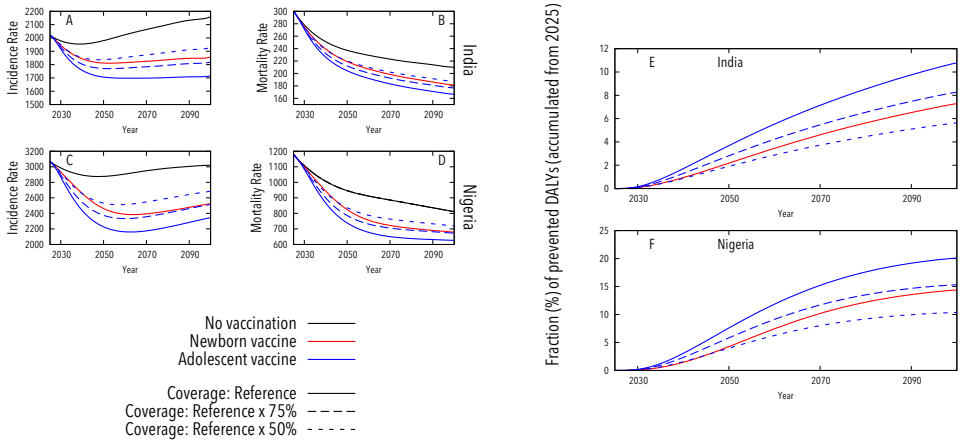


Figure 6.6: **Comparison adolescent vaccination vs newborn vaccination for different coverage levels.** A-D: Incidence and Mortality Rates (Cases per million) vs Time for India (A-B) and Nigeria (C-D) obtained after no vaccination and after applying the vaccines described in table 6.2 and Figure 6.1 at newborns or adolescents. E-F: Fraction (in %) of DALYs prevented (eq. 6.14) vs Time after the inoculation of the different vaccine campaigns in India (E) and Nigeria (F). Vaccines considered in this figure have no blocking.

to what we obtained in previous Figure 6.6. For low values of blocking the adolescent campaign is still able to outperform newborn vaccination, but as blocking approaches the levels of Salvador,[78] this result does not hold anymore.

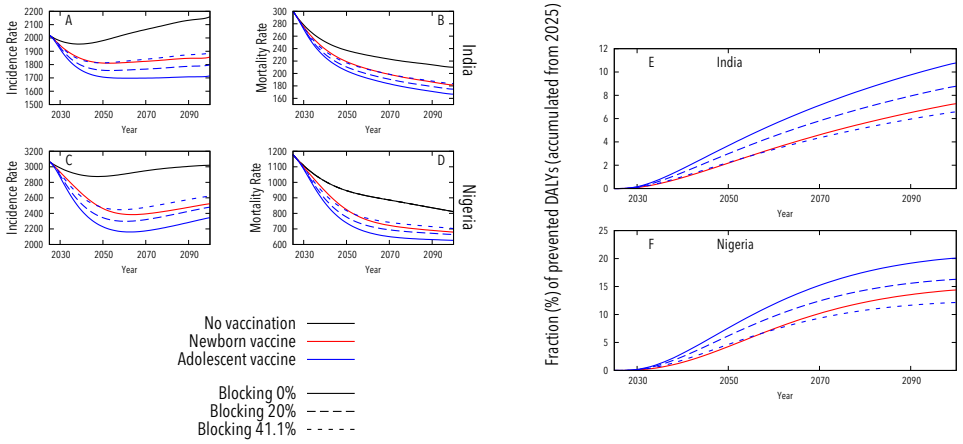


Figure 6.7: **Comparison adolescent vaccination vs newborn vaccination for different blocking levels.** A-D: Incidence and Mortality Rates (Cases per million) vs Time for India (A-B) and Nigeria (C-D) obtained after no vaccination and after applying the vaccines described in table 6.2 and Figure 6.1 at newborns or adolescents. E-F: Fraction (in %) of DALYs prevented (eq. 6.14) vs Time after the inoculation of the different vaccine campaigns in India (E) and Nigeria (F).

## 6.4 Conlussions

The ability of a new vaccine to provide a significant impact at reducing the burden of TB and getting closer to eradication is as low as our ability to foresee that impact.

As we have seen throughout this chapter, in order to obtain a significant impact on the population we need vaccines with low immunity wanings, so protection can reach all age-groups. The same effect could eventually be achieved by having mass vaccination campaigns at all ages, but probably the cost of such enterprise would be prohibitive.

The fundamental problem lies in our capacity to determine the waning of new vaccines. Clinical trials span for short periods of time (a couple of years,[200] maybe 5 in an exceptionally resource consuming trial), but immunity waning is by definition a long-term effect. Thus, it is almost impossible to extract information about the immunity waning before the deployment of the vaccine. The situation could not be more worrisome; the information

taken from these clinical trials is what determines which vaccine candidates move to the next stage and which are rejected but it leaves aside the most important parameter of the vaccine, it's immunity waning.

Furthermore, the forecasts about the performance of new vaccines have turned out to be very pessimistic. In Chapter 3 we discussed how the expected aging of the population is often associated with a stabilization, or even a rebound, in TB burden levels. Now we have seen that, unless we achieve incredibly lasting immunity, we can not avoid this effect. Most likely, heterogeneities in contact patterns and the aging of demography is after the unexpected underperformance of these vaccines. On the one hand, assortativity prevents the reduction of transmission, as contacts remain mostly within age-groups; and on the other hand, as population gets older transmission ceases to play an important role, as most cases in elderly people come from old infections that a pre-exposure vaccine can not avoid in the short-term. Thus, as we stated in Chapter 3, we should not rule out these features when modeling TB spreading, specially for the evaluation of new vaccines.

As for the long debate about which vaccination strategy (newborn or adolescent) offers a stronger performance we don't have a clear answer. Although adolescent vaccination offers a faster impact as it targets the age strata most affected by the disease, it will always leave a reservoir of unprotected children (or with the variable protection offered by BCG to be more precise). However, a newborn vaccine with a sustained protective power, holds the promise of protecting eventually the entire population, offering a better long-term performance than its adolescent counterpart. Besides the feasibility of an adolescent campaign is a substantial practical difficulty (in terms of coverage and blocking).

In this work we have limited the study to the case of vaccines that reduce the probability of fast-progression. The space of parameter on which a new vaccine could act, that was hinted in Chapter 5, remains vastly unexplored.

Although an effective vaccine could undoubtly play a role in the control of TB, if the goal is to eradicate the disease during this century other approaches should be considered as well (post-exposure vaccine, treatment of LTBI, better diagnostic tools, better treatments, ...)

# Conclussions

The goal of this thesis is to develop a model of TB spreading for the evaluation of new vaccines. Specifically we intend with this thesis to achieve a more advanced description of those processes that depend on the age (demographic evolution, contacts among age groups, ...) than what can be found in current models. The importance of these advances come for two different paths. On one hand, the way in which TB develops strongly depends on the age of the individuals that suffers it. For example, we know that, after getting infected, newborns and adults tend to develop the disease faster than school-aged children, who tend to develop asymptomatic latency. We also know that adults tend to develop the most infectious forms of TB (pulmonary TB) while children develop other forms of TB (meningeal, milliar, ...) considered as non contagious. On the other hand, vaccines also act on a different way depending on the age of the vaccinated individual. Among all possible effects that might compromise the immunization provided by the vaccine, the previous exposure to environmental mycobacteria stands out. By definition, exposure will be greater the older the individual is. In addition, currently there is a debate about which age group should be immunized: immunizing adolescents offers more rapid protection (since they are more infectious than children) but leaves a reservoir of unvaccinated children.

We have seen throughout these thesis that, for all these reasons, it is necessary to update the current models with a more realistic description of those mechanisms that depend on age. Specifically, in this thesis we propose to include demographic predictions in the simulations, so that the demographic pyramids evolve in time in a realistic way instead of being considered constant. In a similar way, we include heterogeneous contact patterns depending on age, extracted from experimental studies, instead of assuming that individuals mix among them in a homogeneous way without accounting for age.

Nonetheless, implementing these two ingredients is not as trivial as it might seem at first. As we discuss in chapter 2, contact patterns must be

adapted to the demographic pyramid of the population we simulate. In this chapter we study how (and in which cases) we can adapt contact patterns, that were measured empirically in a determined location, to other populations. The fundamental problem, that we have studied in this chapter, is the lack of reciprocity that is produced when contact patterns, measured in a certain setting, are used in a different population with a different age structure. Thus, we have studied in this chapter different methods to adapt contact patterns and how they influence in the propagation of a SEIR-type disease.

We have also addressed the differences between contact patterns measured in different countries. We have concluded after our analysis that contact matrices belonging to African countries show larger assortativity and a smaller influence of young individuals than European matrices. Our results warn against the potential biases that could occur when contact matrices are extrapolated blindly from one country to another. These findings will be fundamental in our model of TB spreading, in which the age structure changes at every time step; but also might be useful for the modeling of other infectious diseases.

In chapter 3 we develop our model of TB spreading. We include heterogeneous contact patterns and demographic evolution over a state-of-the-art natural history with 17 states for the disease (including fast and slow latency, different types of active disease, different treatment outcomes, etc.). After abandoning the oversimplifying hypotheses of constant demographic pyramids and homogeneous mixing, we prove how the predictions of the model are modified. Specifically we have seen how including the predicted aging of the population produces more pessimistic outcomes (more TB burden) than when this ingredient is forgotten. Adults suffered a higher TB burden than children and they are also more efficient spreaders, and thus, the aging of populations produces higher predicted incidence and mortality. This result has important consequences regarding the management of Public Health: the efforts made to lower the incidence of TB in the last decades should be redoubled in the future. Also, implementing the heterogeneity in contact patterns changes the age distribution of the burden of TB, which has implications for age-focused interventions.

After developing the model of TB spreading, during the next chapters we study how to correctly describe the mechanism of the vaccine, as well as the possible effects that might compromise its power. We start studying the effect of previous exposure to mycobacteria suffered by BCG. In order to do

that, through chapter 4 we have analyzed the BCG-REVAC clinical trials to quantify the effects of masking and blocking suffered by this vaccine. We show how blocking is the effect that most plausibly is behind the variability that we find in BCG's efficacy, and we measured its value for two cities in Brazil: Manaus and Salvador. The possible occurrence of blocking in new vaccines must be considered in the debate about which age group should receive immunization, something that usually goes unnoticed.

In addition to these effects that can make a vaccine more or less effective, there are different ways or mechanisms by which a vaccine can interrupt the life cycle of the pathogen. In chapter 5 we study how protection against disease might be achieved via two different (not exclusive) pathways: the slowdown of rapid progression and/or the prevention of rapid progression (producing LTBI). These two mechanisms are indistinguishable in a classical clinical trial but produce very different impacts when evaluating through a model of TB spreading. We have proposed a new analysis that potentially could disentangle these two mechanisms, producing much more accurate impacts and increasing the probability of having a successful trial. This new design of clinical trials does not require to take more data than current trials; and this framework is potentially extensible to more complex scenarios.

After acquiring all this knowledge about the modeling of TB spreading and the details of the vaccines, we study the debate about which is the optimal age group to receive vaccination. Prior to this work it was accepted without qualms, that vaccinating adolescents offers a faster and more pronounced impact than vaccinating neonates. However, previous studies not only did not consider the new ingredients that we have added to the TB spreading model (demographic evolution and heterogeneous contact patterns) but also did not consider many of the details around the description vaccines that we have studied in this thesis, and did not explore the entire parameter space of vaccines. When we expand the possibilities offered by the vaccines, the statement that adolescent vaccination is better than neonatal vaccination loses generality. Vaccinating adolescents offers a much more immediate impact than vaccinating newborns as correctly pointed out by previous models. Nonetheless vaccinating adolescents leaves a reservoir of unprotected children. That is why a newborn vaccine, if it is able of sustain its immunity in time, on the long-term can offer a greater impact than adolescent vaccination. Besides there are two extra factors that can affect the comparison between adolescent and newborn vaccination. On the one hand, the coverage on adolescent will be, presumably, lower than newborn coverage, and on the other hand, the blocking that we have studied in chapter 4 will compromise

adolescent vaccination but will not affect newborn vaccination (because newborns have not been exposed to environmental mycobacteria).

Besides the advances proposed by this thesis, there are still many limitations. Our model of TB spreading rely on a series of epidemiological parameters and TB burden estimations that are subject to strong sources of uncertainties. Thus, it would be recommendable in a future to improve the quality of the data introduced as input of the model (and in this sense the WHO suggests the implementation of prevalence studies), and it would be specially positive the acquisition of age-stratified data.

We also find enormous limitations in the parameterization of new vaccines. The long-term effects of a vaccine, by its own nature, can not be measured by clinical trials, leaving a wide range of uncertainty in our predictions. Therefore, we should be aware of these limitations when we evaluate the impact of new vaccines.

# Conclusiones

El objetivo de esta tesis sa desarrollar un modelo de propagación de la Tuberculosis (TB) para la evaluación de nuevas vacunas. Concretamente pretendemos con esta tesis una descripción más avanzada de aquellos procesos que dependen de la edad (evolución demográfica, contactos entre grupos de edad, ...) que la que podemos encontrar en los modelos actuales. La importancia de estos avances vienen por dos vertientes diferentes. Por un lado, la forma en que se desarrolla la TB depende enormemente de la edad del individuo que la padece. Por ejemplo, sabemos que, tras infectarse, los neonatos y los adultos tienden a desarrollar la enfermedad más rápidamente que los niños en edad escolar, quienes suelen desarrollar latencia asintomática. También sabemos que los adultos tienden a desarrollar las formas de TB más infecciosas (Tuberculosis pulmonar) mientras que los niños desarrollan formas de TB (meníngea, miliar, ...) consideradas no contagiosas. Por otro lado, las vacunas también actúan de forma distinta en función de la edad del individuo vacunado. Entre todos los posibles efectos que pueden comprometer la inmunización proporcionada por una vacuna, destaca la exposición previa a micobacterias ambientales. Por definición, está exposición acumulativa será mayor cuanto más edad tenga el individuo. Además, actualmente se plantea el debate de a qué grupo de edad debería inmunizarse: inmunizar adolescentes ofrece una protección más rápida (ya que son más infecciosos que los niños) pero deja un reservorio de niños sin vacunar.

Hemos visto a lo largo de esta tesis que, por todos estos motivos, es necesario actualizar los modelos actuales con una descripción más realista de aquellos mecanismos que dependen de la edad. En concreto, en esta tesis proponemos incluir predicciones demográficas en las simulaciones, de forma que las pirámides demográficas evolucionen en el tiempo de una manera realista en lugar de ser consideradas constantes. De forma similar, incluimos patrones de contactos heterogéneos dependientes de la edad, extraídos de estudios experimentales, en lugar de asumir que los individuos se mezclan entre sí de forma homogénea sin tener en cuenta la edad.

Sin embargo, implementar estos dos ingredientes no es tan trivial como podía parecer en un principio. Como discutimos en el capítulo 2, los patrones de contacto deben adaptarse a la pirámide demográfica de la población que simulamos. En este capítulo estudiamos cómo (y en qué casos) los patrones de contactos, medidos empíricamente en una localización determinada, pueden ser adaptados a otras poblaciones. El problema fundamental que hemos estudiado aquí es la falta de reciprocidad que se produce cuando se utilizan unos patrones de contactos medidos en una población determinada, en una población distinta con una estructura de edades diferente. Así, hemos estudiado en este capítulo distintos métodos para adaptar patrones de contacto, y como influyen en la propagación de una enfermedad tipo SEIR.

También hemos estudiado las diferencias que existen entre los patrones de contacto medidos en diferentes países del mundo. Hemos concluído tras nuestro análisis, que las matrices de contactos de países africanos presentan una mayor asortatividad y una menor influencia de los individuos más jovenes que la que encontramos en las matrices europeas. Nuestros resultados advierten de los sesgos potenciales que pueden producirse cuando se extrapolan matrices de contacto indiscriminadamente de un país a otro. Estos hallazgos serán fundamentales en nuestro modelo de propagación de TB, en el que la estructura por edades cambia en cada paso temporal; pero también pueden ser de utilidad para la modelización de otras enfermedades infecciosas.

En el capítulo 3 desarrollamos nuestro modelo de propagación de TB. Como venimos diciendo, incluimos patrones de contacto heterogéneos y evolución demográfica sobre una historia natural con 19 estados para la enfermedad (incluyendo latencia rápida y lenta, distintos tipos de enfermedad activa, distintas posibilidades de recuperación, etc.). Tras abandonar las hipótesis sobre-simplificadoras de pirámides demográficas constantes y mezcla homogénea, comprobamos como se modifican las predicciones realizadas por el modelo. Específicamente, hemos visto como incluir el envejecimiento previsto de las poblaciones produce predicciones más pesimista (carga de Tuberculosis mayor) que cuando este ingrediente es olvidado. Los adultos sufren una mayor carga de Tuberculosis y son propagadores más eficientes, y por ello el envejecimiento de las poblaciones produce predicciones de incidencia y mortalidad más elevadas. Este resultado tiene importantes consecuencias en lo que concierne a la gestión de la Salud Pública: los esfuerzos realizados para bajar la incidencia de la Tuberculosis en las últimas decadas deberán ser redoblados en un futuro. También implementar la heterogeneidad en los patrones de contactos cambia la distribución por edades de la carga de

Tuberculosis, lo cuál tiene implicaciones para intervenciones centradas en la edad.

Tras desarrollar el modelo de propagación de TB, durante los siguientes capítulos estudiamos cómo describir correctamente el funcionamiento de la vacuna, así como los posibles efectos que pueden comprometer el potencial de la misma. Empezamos estudiando el efecto de la exposición previa a micobacterias que sufre la vacuna BCG. Para ello, en el capítulo 4, hemos analizado los ensayos clínicos BCG-REVAC para cuantificar los efectos de enmascaramiento y bloqueo sufridos por esta vacuna. Mostramos como el bloqueo es el efecto que, más plausiblemente, está detrás de la variabilidad que encontramos en la eficacia de BCG, y medimos su valor para dos ciudades de Brasil: Manaus y Salvador. La posible aparición de bloqueo en nuevas vacunas debe ser considerada en el debate sobre qué grupo de edad debería recibir inmunización, algo que suele pasar desapercibido.

Además de estos efectos que pueden hacer que una vacuna sea más o menos efectiva, existen distintas vías o mecanismos por los cuales una vacuna puede interrumpir el ciclo vital del patógeno. En el capítulo 5 hemos estudiado como la protección contra enfermedad puede conseguirse por dos vías diferentes (no excluyentes): la ralentización de la progresión rápida y/o la prevención de la progresión rápida (produciendo LTBI). Estos dos mecanismos son indistinguibles en un ensayo clínico clásico pero producen impactos muy distintos cuando se evaluan con un modelo de propagación de Tuberculosis. Hemos propuesto un nuevo análisis que podría potencialmente desenredar estos dos mecanismos, produciendo impactos mucho más precisos en la evaluación de vacunas y aumentando las probabilidades de tener un ensayo exitoso. Este nuevo diseño de los ensayos clínicos no requiere tomar más datos que los ensayos actuales, y este marco es potencialmente extensible a situaciones más complejas.

Después de adquirir todos estos conocimientos acerca de la modelización de propagación de Tuberculosis y los detalles de las vacunas, estudiamos el debate sobre cual es el grupo de edad óptimo para recibir la vacunación. Antes de este trabajo se aceptaba sin reparos que vacunar adolescentes ofrece un impacto más rápido y acusado que vacunar neonatos. Sin embargo, los estudios previos no solo no consideraban los ingredientes que hemos añadido al modelo de propagación de TB (evolución demográfica y patrones de contacto heterogéneos) sino que tampoco consideraban muchos de los detalles alrededor de las vacunas que hemos estudiado en esta tesis, y no exploraban todo el espacio de parámetros propio de las vacunas. Cuando expandimos las

posibilidades ofrecidas por las vacunas, la declaración de que la vacunación adolescente es mejor que la vacunación neonatal pierde generalidad. Vacunar adolescentes ofrece un impacto más inmediato que vacunar neonatos como bien apuntaban modelos previos. Sin embargo, la vacunación a adolescentes deja un reservorio de niños sin protección. Es por ello que una vacuna neonatal, si es capaz de sostener su inmunidad en el tiempo, a largo plazo puede llegar a ofrecer un mayor impaco que la vacunación a adolescentes. Además, hay otros dos factores que pueden afectar la comparación entre vacunación adolescente y neonatal. Por un lado la cobertura a adolescentes será presumiblemente inferior que la cobertura neonatal, y por otro, el bloqueo que hemos estudiado en el capítulo 4 afectará a la vacunación adolescente pero no a la vacunación neonatal (ya que los recién nacidos no han sido expuestos a micobacterias ambientales).

A pesar de los avances propuestos por esta tesis, todavía hay muchas limitaciones. Nuestro modelo de propagación de Tuberculosis dependen de una serie de parámetros epidemiológicos y estimaciones de la carga de Tuberculosis que están sujetos a grandes fuertes de incertidumbres. Sería recomendable en un futuro mejorar la calidad de los datos que introducimos como input del modelo (y en este sentido la Organización Mundial de la Salud sugiere la implementación de estudios de prevalencia), y sería especialmente positivo la obtención de datos estratificados por edades.

También encontramos grandes limitaciones en la parametrización de las nuevas vacunas. Los efectos a largo plazo de una vacuna, por su propia naturaleza, no puede ser medida en ensayos clínicos, dejando un amplio rango de incertidumbre en nuestras predicciones. Por tanto, deberíamos ser conscientes de las limitaciones cuando evaluamos el impacto de nuevas vacunas.

# Bibliography


[1] David J Robinson. A pest in the land: New world epidemics in a global perspective. *The Americas*, 61(2):274–275, 2004.

[2] Ole J Benedictow. *The Black Death, 1346-1353: the complete history*. Boydell & Brewer, 2004.

[3] Philip Ziegler. *The black death*. Faber & Faber, 2013.

[4] Norman F Cantor. *In the wake of the plague: the Black Death and the world it made*. Simon and Schuster, 2001.

[5] Daniel Bernoulli. Réflexions sur les avantages de l'inoculation. *Mém. Paris*, pages 439–482, 1758.

[6] Daniel Bernoulli. Essai d'une nouvelle analyse de la mortalité causée par la petite vérole, et des avantages de l'inoculation pour la prévenir. *Histoire de l'Acad., Roy. Sci.(Paris) avec Mem*, pages 1–45, 1760.

[7] Klaus Dietz and JAP Heesterbeek. Daniel bernoulli's epidemiological model revisited. *Mathematical biosciences*, 180(1-2):1–21, 2002.

[8] Fred Brauer. Mathematical epidemiology: Past, present, and future. *Infectious Disease Modelling*, 2(2):113–127, 2017.

[9] Steven Johnson. *The ghost map: The story of London's most terrifying epidemic–and how it changed science, cities, and the modern world*. Penguin, 2006.

[10] John Snow. On the mode of communication of cholera. *Edinburgh medical journal*, 1(7):668, 1856.

[11] Sandra Hempel. *The medical detective: John Snow, cholera and the mystery of the Broad Street pump*. Granta Books, 2014.

[12] William Budd. *Typhoid fever: its nature, mode of spreading, and prevention*. Longmans, Green, 1873.

[13] Robert Moorhead. William Budd and typhoid fever. *Journal of the Royal Society of Medicine*, 95(11):561–564, 2002.

[14] William Heaton Hamer. *The Milroy lectures on epidemic disease in England: the evidence of variability and of persistency of type*. Bedford Press, 1906.

[15] Ronald Ross. The prevention of malaria, with addendum on the theory of happenings. *Murray, London*, 1911.

[16] George Macdonald et al. The epidemiology and control of malaria. *The Epidemiology and Control of Malaria.*, 1957.

[17] William O Kermack and Anderson G McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 115(772):700–721, 1927. ISSN 0950-1207. doi: 10.1098/rspa.1927.0118. URL `http://rspa.royalsocietypublishing.org/content/115/772/700`.

[18] William O Kermack and Anderson G McKendrick. Contributions to the mathematical theory of epidemics. ii.—the problem of endemicity. *Proc. R. Soc. Lond. A*, 138(834):55–83, 1932.

[19] William O Kermack and Anderson G McKendrick. Contributions to the mathematical theory of epidemics. iii.—further studies of the problem of endemicity. *Proc. R. Soc. Lond. A*, 141(843):94–122, 1933.

[20] Michael Y Li, John R Graef, Liancheng Wang, and János Karsai. Global dynamics of a SEIR model with varying total population size. *Mathematical biosciences*, 160(2):191–213, 1999.

[21] Carlos Castillo-Chavez, Wenzhang Huang, and Jia Li. Competitive exclusion and coexistence of multiple strains in an sis std model. *SIAM Journal on Applied Mathematics*, 59(5):1790–1811, 1999.

[22] Masashi Kamo and Akira Sasaki. The effect of cross-immunity and seasonal forcing in a multi-strain epidemic model. *Physica D: Nonlinear Phenomena*, 165(3-4):228–241, 2002.

[23] Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemic spreading in scale-free networks. *Physical review letters*, 86(14):3200, 2001.

[24] Yamir Moreno, Romualdo Pastor-Satorras, and Alessandro Vespignani. Epidemic outbreaks in complex heterogeneous networks. *The European Physical Journal B-Condensed Matter and Complex Systems*, 26(4): 521–529, 2002.

[25] Joël Mossong, Niel Hens, Mark Jit, Philippe Beutels, Kari Auranen, Rafael Mikolajczyk, Marco Massari, Stefania Salmaso, Gianpaolo Scalia Tomba, Jacco Wallinga, et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS medicine*, 5(3):e74, 2008.

[26] Stephen Eubank, Hasan Guclu, VS Anil Kumar, Madhav V Marathe, Aravind Srinivasan, Zoltan Toroczkai, and Nan Wang. Modelling disease outbreaks in realistic urban social networks. *Nature*, 429(6988): 180, 2004.

[27] Lauren Ancel Meyers, Babak Pourbohloul, Mark EJ Newman, Danuta M Skowronski, and Robert C Brunham. Network theory and SARS: predicting outbreak diversity. *Journal of theoretical biology*, 232 (1):71–81, 2005.

[28] Hazhir Rahmandad and John Sterman. Heterogeneity and network structure in the dynamics of diffusion: Comparing agent-based and differential equation models. *Management Science*, 54(5):998–1014, 2008.

[29] Liliana Perez and Suzana Dragicevic. An agent-based approach for modeling dynamics of contagious disease spread. *International journal of health geographics*, 8(1):50, 2009.

[30] Dirk Brockmann and Dirk Helbing. The hidden geometry of complex, network-driven contagion phenomena. *Science*, 342(6164):1337–1342, 2013.

[31] Duygu Balcan, Hao Hu, Bruno Goncalves, Paolo Bajardi, Chiara Poletto, Jose J Ramasco, Daniela Paolotti, Nicola Perra, Michele Tizzoni, Wouter Van den Broeck, et al. Seasonal transmission potential and activity peaks of the new influenza A (H1N1): a Monte Carlo likelihood analysis based on human mobility. *BMC medicine*, 7(1):45, 2009.

[32] Paolo Bajardi, Chiara Poletto, Jose J Ramasco, Michele Tizzoni, Vittoria Colizza, and Alessandro Vespignani. Human mobility networks, travel restrictions, and the global spread of 2009 H1N1 pandemic. *PloS one*, 6(1):e16591, 2011.

[33] Sandro Meloni, Nicola Perra, Alex Arenas, Sergio Gómez, Yamir Moreno, and Alessandro Vespignani. Modeling human mobility responses to the large-scale spreading of infectious diseases. *Scientific reports*, 1:62, 2011.

[34] Vitaly Belik, Theo Geisel, and Dirk Brockmann. Natural human mobility patterns and spatial spread of infectious diseases. *Physical Review X*, 1(1):011001, 2011.

[35] Michele Tizzoni, Paolo Bajardi, Adeline Decuyper, Guillaume Kon Kam King, Christian M Schneider, Vincent Blondel, Zbigniew Smoreda, Marta C González, and Vittoria Colizza. On the use of human mobility proxies for modeling epidemics. *PLoS computational biology*, 10(7):e1003716, 2014.

[36] Alberto Aleta, Andreia NS Hisi, Sandro Meloni, Chiara Poletto, Vittoria Colizza, and Yamir Moreno. Human mobility networks and persistence of rapidly mutating pathogens. *Royal Society open science*, 4 (3):160914, 2017.

[37] Wouter Van den Broeck, Corrado Gioannini, Bruno Gonçalves, Marco Quaggiotto, Vittoria Colizza, and Alessandro Vespignani. The GLEaMviz computational tool, a publicly available software to explore realistic epidemic spreading scenarios at the global scale. *BMC infectious diseases*, 11(1):37, 2011.

[38] Jeffrey Shaman and Alicia Karspeck. Forecasting seasonal outbreaks of influenza. *Proceedings of the National Academy of Sciences*, 109(50): 20425–20430, 2012.

[39] Elaine O Nsoesie, John S Brownstein, Naren Ramakrishnan, and Madhav V Marathe. A systematic review of studies on forecasting the dynamics of influenza outbreaks. *Influenza and other respiratory viruses*, 8(3):309–316, 2014.

[40] Marcelo FC Gomes, Ana Pastore y Piontti, Luca Rossi, Dennis Chao, Ira Longini, M Elizabeth Halloran, and Alessandro Vespignani. Assessing the international spreading risk associated with the 2014 West African Ebola outbreak. *PLoS currents*, 6, 2014.

[41] Isaac I Bogoch, Oliver J Brady, Moritz UG Kraemer, Matthew German, Marisa I Creatore, Manisha A Kulkarni, John S Brownstein, Sumiko R Mekaru, Simon I Hay, Emily Groot, et al. Anticipating the international spread of zika virus from brazil. *The Lancet*, 387(10016):335–336, 2016.

[42] Robert Koch. The etiology of tuberculosis. *Reviews of infectious diseases*, 4(6):1270–1274, 1982.

[43] Thomas Dormandy. *The white death: a history of tuberculosis*. Hambledon Press, London, 1999.

[44] Anne O'Garra, Paul S Redford, Finlay W McNab, Chloe I Bloom, Robert J Wilkinson, and Matthew PR Berry. The immune response in tuberculosis. *Annual review of immunology*, 31:475–527, 2013.

[45] GH Mukesh, CM Theresa, OS Jo-Anne, PK Jane, MS Manohar, and RM Peter. Tuberculosis from head to toe. *Radiographics*, 20:449–470, 2000.

[46] ED Carrol, JE Clark, and AJ Cant. Non-pulmonary tuberculosis. *Paediatric Respiratory Reviews*, 2(2):113–119, 2001.

[47] Clifton E Barry 3rd, Helena I Boshoff, Véronique Dartois, Thomas Dick, Sabine Ehrt, JoAnne Flynn, Dirk Schnappinger, Robert J Wilkinson, and Douglas Young. The spectrum of latent tuberculosis: rethinking the biology and intervention strategies. *Nature Reviews Microbiology*, 7(12):845, 2009.

[48] Surendra K Sharma, Sandeep Mohanan, and Abhishek Sharma. Relevance of latent TB infection in areas of high TB prevalence. *Chest*, 142(3):761–773, 2012.

[49] C Dye, S Scheele, P Dolin, V Pathania, and MC Raviglione. Consensus statement. global burden of tuberculosis: estimated incidence, prevalence, and mortality by in an urban community. *Int J Tuberc Lung Dis*, 9(9):970–6, 2005.

[50] George W Comstock, Verna T Livesay, and Shirley F Woolpert. The prognosis of a positive tuberculin reaction in childhood and adolescence. *American journal of epidemiology*, 99(2):131–138, 1974.

[51] Emilia Vynnycky and Paul EM Fine. Lifetime risks, incubation period, and serial interval of tuberculosis. *American journal of epidemiology*, 152(3):247–263, 2000.

[52] Elizabeth L Corbett, Catherine J Watt, Neff Walker, Dermot Maher, Brian G Williams, Mario C Raviglione, and Christopher Dye. The growing burden of tuberculosis: global trends and interactions with the HIV epidemic. *Archives of internal medicine*, 163(9):1009–1021, 2003.

[53] BJ Marais, RP Gie, HS Schaaf, AC Hesseling, CC Obihara, JJ Starke, DA Enarson, PR Donald, and N Beyers. The natural history of childhood intra-thoracic tuberculosis: a critical review of literature from the pre-chemotherapy era [State of the Art]. *Int J Tuber Lung Dis*, 8(4): 392–402, 2004.

[54] Thomas R Frieden, Timothy R Sterling, Sonal S Munsiff, Catherine J Watt, and Christopher Dye. Tuberculosis. *The lancet*, 362(9387):887–899, 2003.

[55] Laith J Abu-Raddad, Lorenzo Sabatelli, Jerusha T Achterberg, Jonathan D Sugimoto, Ira M Longini, Christopher Dye, and M Elizabeth Halloran. Epidemiological benefits of more-effective tuberculosis vaccines, drugs, and diagnostics. *Proceedings of the National Academy of Sciences*, 106(33):13980–13985, 2009.

[56] Iñaki Comas, Mireia Coscolla, Tao Luo, Sonia Borrell, Kathryn E Holt, Midori Kato-Maeda, Julian Parkhill, Bijaya Malla, Stefan Berg, Guy Thwaites, et al. Out-of-africa migration and neolithic coexpansion of mycobacterium tuberculosis with modern humans. *Nature genetics*, 45 (10):1176, 2013.

[57] Leonard G Wilson. Commentary: Medicine, population, and tuberculosis. *International Journal of Epidemiology*, 34(3):521–524, 2004.

[58] Martien W Borgdorff, Katherine Floyd, and Jaap F Broekmans. Interventions to reduce tuberculosis mortality and transmission in low-and middle-income countries. *Bulletin of the World Health Organization*, 80(3):217–227, 2002.

[59] Christian Lienhardt, Philippe Glaziou, Mukund Uplekar, Knut Lönnroth, Haileyesus Getahun, and Mario Raviglione. Global tuberculosis control: lessons learnt and future prospects. *Nature Reviews Microbiology*, 10(6):407, 2012.

[60] Marie-Claude Boily, C Lowndes, and M Alary. The impact of HIV epidemic phases on the effectiveness of core group interventions: insights

from mathematical models. *Sexually transmitted infections*, 78(suppl 1):i78–i90, 2002.

[61] Joaquín Sanz, Cheng-Yi Xia, Sandro Meloni, and Yamir Moreno. Dynamics of interacting diseases. *Physical Review X*, 4(4):041005, 2014.

[62] Eline L Korenromp, Fabio Scano, Brian G Williams, Christopher Dye, and Paul Nunn. Effects of human immunodeficiency virus infection on recurrence of tuberculosis after rifampin-based treatment: an analytical review. *Clinical Infectious Diseases*, pages 101–112, 2003.

[63] Haileyesus Getahun, Christian Gunneberg, Reuben Granich, and Paul Nunn. HIV infection—associated tuberculosis: the epidemiology and the response. *Clinical Infectious Diseases*, 50(Supplement_3):S201–S207, 2010.

[64] Paul Nunn, Brian Williams, Katherine Floyd, Christopher Dye, Gijs Elzinga, and Mario Raviglione. Tuberculosis control in the era of HIV. *Nature Reviews Immunology*, 5(10):819, 2005.

[65] Peter A Selwyn, Diana Hartel, Victor A Lewis, Ellie E Schoenbaum, Sten H Vermund, Robert S Klein, Angela T Walker, and Gerald H Friedland. A prospective study of the risk of tuberculosis among intravenous drug users with human immunodeficiency virus infection. *New England journal of medicine*, 320(9):545–550, 1989.

[66] Giorgio Antonucci, Enrico Girardi, Mario C Raviglione, Giuseppe Ippolito, Paolo Almi, Gioacchino Angarano, Orlando Armignacco, Sergio Babudieri, Nazario Bevilacqua, Alessandra Bini, et al. Risk Factors for Tuberculosis in HIV-Infected Persons: A Prospective Cohort Study. *Jama*, 274(2):143–148, 1995.

[67] Christopher Dye and Brian G Williams. Criteria for the control of drug-resistant tuberculosis. *Proceedings of the National Academy of Sciences*, 97(14):8180–8185, 2000.

[68] N Sarita Shah, Abigail Wright, Gill-Han Bai, Lucia Barrera, Fadila Boulahbal, Nuria Martín-Casabona, Francis Drobniewski, Chris Gilpin, Marta Havelková, Rosario Lepe, et al. Worldwide emergence of extensively drug-resistant tuberculosis. *Emerging infectious diseases*, 13(3): 380, 2007.

[69] Neel R Gandhi, Paul Nunn, Keertan Dheda, H Simon Schaaf, Matteo Zignol, Dick Van Soolingen, Paul Jensen, and Jaime Bayona.

Multidrug-resistant and extensively drug-resistant tuberculosis: a threat to global control of tuberculosis. *The Lancet*, 375(9728):1830–1843, 2010.

[70] Christopher Dye. Doomsday postponed? Preventing and reversing epidemics of drug-resistant tuberculosis. *Nature Reviews Microbiology*, 7(1):81, 2009.

[71] WHO. Global tuberculosis report 2016. 2016.

[72] Paul EM Fine. Variation in protection by BCG: implications of and for heterologous immunity. *The Lancet*, 346(8986):1339–1345, 1995.

[73] B Bourdin Trunz, PEM Fine, and C Dye. Effect of BCG vaccination on childhood tuberculous meningitis and miliary tuberculosis worldwide: a meta-analysis and assessment of cost-effectiveness. *The Lancet*, 367 (9517):1173–1180, 2006.

[74] Laura C Rodrigues, Vinod K Diwan, and Jeremy G Wheeler. Protective effect of BCG against tuberculous meningitis and miliary tuberculosis: a meta-analysis. *International journal of epidemiology*, 22(6):1154–1158, 1993.

[75] A Roy, M Eisenhut, RJ Harris, LC Rodrigues, S Sridhar, S Habermann, L Snell, P Mangtani, I Adetifa, A Lalvani, et al. Effect of BCG vaccination against mycobacterium tuberculosis infection in children: systematic review and meta-analysis. *Bmj*, 349:g4643, 2014.

[76] Punam Mangtani, Ibrahim Abubakar, Cono Ariti, Rebecca Beynon, Laura Pimpin, Paul EM Fine, Laura C Rodrigues, Peter G Smith, Marc Lipman, Penny F Whiting, et al. Protection by BCG vaccine against tuberculosis: a systematic review of randomized controlled trials. *Clinical infectious diseases*, 58(4):470–480, 2013.

[77] Mauricio L Barreto, Susan M Pereira, Daniel Pilger, Alvaro A Cruz, Sergio S Cunha, Clemax Sant'Anna, Maria Y Ichihara, Bernd Genser, and Laura C Rodrigues. Evidence of an effect of BCG revaccination on incidence of tuberculosis in school-aged children in brazil: second report of the bcg-revac cluster-randomised trial. *Vaccine*, 29(31):4875–4877, 2011.

[78] Sergio Arregui, Joaquín Sanz, Dessislava Marinova, Carlos Martín, and Yamir Moreno. On the impact of masking and blocking hypotheses for

measuring the efficacy of new tuberculosis vaccines. *PeerJ*, 4:e1513, 2016.

[79] Chia-Lin Tseng, Olivia Oxlade, Dick Menzies, Anne Aspler, and Kevin Schwartzman. Cost-effectiveness of novel vaccines for tuberculosis control: a decision analysis study. *BMC public health*, 11(1):55, 2011.

[80] EM Agger and P Andersen. A novel TB vaccine; towards a strategy based on our understanding of BCG failure. *Vaccine*, 21(1-2):7–14, 2002.

[81] Aeras website. www.aeras.org, Accessed: 10-08-2018.

[82] Dessislava Marinova, Jesus Gonzalo-Asensio, Nacho Aguilo, and Carlos Martin. Recent developments in tuberculosis vaccines. *Expert review of vaccines*, 12(12):1431–1448, 2013.

[83] Ian M Orme. Vaccine development for tuberculosis: current progress. *Drugs*, 73(10):1015–1024, 2013.

[84] Wenping Gong, Yan Liang, and Xueqiong Wu. The current status, challenges, and future developments of new tuberculosis vaccines. *Human vaccines & immunotherapeutics*, pages 1–20, 2018.

[85] Carlos Martin, Nacho Aguilo, and Jesus Gonzalo-Asensio. Vaccination against tuberculosis. *Enfermedades infecciosas y microbiologia clinica*, 2018.

[86] Rebecca C Harris, Tom Sumner, Gwenan M Knight, and Richard G White. Systematic review of mathematical models exploring the epidemiological impact of future TB vaccines. *Human vaccines & immunotherapeutics*, 12(11):2813–2832, 2016.

[87] Christopher Dye, Philippe Glaziou, Katherine Floyd, and Mario Raviglione. Prospects for tuberculosis elimination. *Annual review of public health*, 34:271–286, 2013.

[88] Charles S Revelle, Walter R Lynn, and Floyd Feldmann. Mathematical models for the economic allocation of tuberculosis control activities in developing nations. *American review of respiratory disease*, 96(5):893–909, 1967.

[89] Carlos Castillo-Chavez and Zhilan Feng. Global stability of an age-structure model for TB and its applications to optimal vaccination strategies. *Mathematical biosciences*, 151(2):135–154, 1998.

[90] Christopher JL Murray and Joshua A Salomon. Modeling the impact of global tuberculosis control strategies. *Proceedings of the National Academy of Sciences*, 95(23):13881–13886, 1998.

[91] CP Bhunu, W Garira, Z Mukandavire, and G Magombedze. Modelling the effects of pre-exposure and post-exposure vaccines in tuberculosis control. *Journal of theoretical biology*, 254(3):633–649, 2008.

[92] Ted Cohen, Caroline Colijn, and Megan Murray. Modeling the effects of strain diversity and mechanisms of strain competition on the potential performance of new tuberculosis vaccines. *Proceedings of the National Academy of Sciences*, 105(42):16302–16307, 2008.

[93] Christopher Dye and Brian G Williams. Eliminating human tuberculosis in the twenty-first century. *Journal of the Royal Society Interface*, 5(23):653–662, 2008.

[94] M Gabriela M Gomes, Ana O Franco, Manuel C Gomes, and Graham F Medley. The reinfection threshold promotes variability in tuberculosis epidemiology and vaccine efficacy. *Proceedings of the Royal Society of London B: Biological Sciences*, 271(1539):617–623, 2004.

[95] Paula Rodrigues, Alessandro Margheri, Carlota Rebelo, and M Gabriela M Gomes. Heterogeneity in susceptibility to infection can explain high reinfection rates. *Journal of theoretical biology*, 259(2): 280–290, 2009.

[96] Douglas Young and Christopher Dye. The development and impact of tuberculosis vaccines. *Cell*, 124(4):683–687, 2006.

[97] Elad Ziv, Charles L Daley, and Sally Blower. Potential public health impact of new tuberculosis vaccines. *Emerging infectious diseases*, 10 (9):1529, 2004.

[98] M Gabriela M Gomes, Paula Rodrigues, Frank M Hilker, Natalia B Mantilla-Beniers, Marion Muehlen, Ana Cristina Paulo, and Graham F Medley. Implications of partial immunity on the prospects for tuberculosis control by post-exposure interventions. *Journal of theoretical biology*, 248(4):608–617, 2007.

[99] C Dye. Tuberculosis 2000–2010: control, but not elimination [the comstock lecture]. *The International Journal of Tuberculosis and Lung Disease*, 4(12):S146–S152, 2000.

[100] Christopher Dye. Making wider use of the world's most widely used vaccine: Bacille calmette–guérin revaccination reconsidered. *Journal of the Royal Society Interface*, 10(87):20130365, 2013.

[101] Gwenan M Knight, Ulla K Griffiths, Tom Sumner, Yoko V Laurence, Adrian Gheorghe, Anna Vassall, Philippe Glaziou, and Richard G White. Impact and cost-effectiveness of new tuberculosis vaccines in low-and middle-income countries. *Proceedings of the National Academy of Sciences*, 111(43):15520–15525, 2014.

[102] T Lietman and SM Blower. Potential impact of tuberculosis vaccines as epidemic control agents. *Clinical Infectious Diseases*, 30 (Supplement_3):S316–S322, 2000.

[103] Elsje Pienaar, Aaron M Fluitt, Scott E Whitney, Alison G Freifeld, and Hendrik J Viljoen. A model of tuberculosis transmission and intervention strategies in an urban residential area. *Computational biology and chemistry*, 34(2):86–96, 2010.

[104] Jared B Ditkowsky and Kevin Schwartzman. Potential cost-effectiveness of a new infant tuberculosis vaccine in south africa-implications for clinical trials: a decision analysis. *PloS one*, 9(1): e83526, 2014.

[105] Liezl Channing and Edina Sinanovic. Modelling the cost-effectiveness of a new infant vaccine to prevent tuberculosis disease in children in south africa. *Cost Effectiveness and Resource Allocation*, 12(1):20, 2014.

[106] Mahbubur Rahman, Miho Sekimoto, Isamu Takamatsu, Kenji Hira, Takuro Shimbo, Kyoichiro Toyoshima, and Tsuguya Fukui. Economic evaluation of universal BCG vaccination of Japanese infants. *International journal of epidemiology*, 30(2):380–385, 2001.

[107] Thomas R Hawn, Tracey A Day, Thomas J Scriba, Mark Hatherill, Willem A Hanekom, Thomas G Evans, Gavin J Churchyard, James G Kublin, Linda-Gail Bekker, and Steven G Self. Tuberculosis vaccines and prevention of infection. *Microbiology and Molecular Biology Reviews*, 78(4):650–671, 2014.

[108] Christopher Dye, Geoffrey P Garnett, Karen Sleeman, and Brian G Williams. Prospects for worldwide tuberculosis control under the WHO DOTS strategy. *The Lancet*, 352(9144):1886–1891, 1998.

[109] Jacco Wallinga, Peter Teunis, and Mirjam Kretzschmar. Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents. *American journal of epidemiology*, 164(10):936–944, 2006.

[110] Jonathan M Read, Ken TD Eames, and W John Edmunds. Dynamic social networks and the implications for the spread of infectious disease. *Journal of The Royal Society Interface*, 5(26):1001–1007, 2008.

[111] Ken TD Eames, Natasha L Tilston, Ellen Brooks-Pollock, and W John Edmunds. Measured dynamic social contact patterns explain the spread of H1N1v influenza. *PLoS computational biology*, 8(3):e1002425, 2012.

[112] Andrea Apolloni, Chiara Poletto, and Vittoria Colizza. Age-specific contacts and travel patterns in the spatial spread of 2009 H1N1 influenza pandemic. *BMC infectious diseases*, 13(1):176, 2013.

[113] Giancarlo De Luca, Kim Van Kerckhove, Pietro Coletti, Chiara Poletto, Nathalie Bossuyt, Niel Hens, and Vittoria Colizza. The impact of regular school closure on seasonal influenza epidemics: a data-driven spatial transmission model for Belgium. *BMC infectious diseases*, 18 (1):29, 2018.

[114] Giorgio Guzzetta, Marco Ajelli, Zhenhua Yang, Stefano Merler, Cesare Furlanello, and Denise Kirschner. Modeling socio-demography to capture tuberculosis transmission dynamics in a low burden setting. *Journal of theoretical biology*, 289:197–205, 2011.

[115] M Pai et al. Tuberculosis. *Nature Reviews Disease Primers*, 2(16076), 2016.

[116] Sam Abbott, Hannah Christensen, Maeve K Lalor, Dominik Zennor, Colin NJ Campbell, Mary E Ramsay, and Ellen Brooks-Pollock. Exploring the effects of BCG vaccination in patients diagnosed with tuberculosis: observational study using the Enhanced Tuberculosis Surveillance system. *bioRxiv*, page 366476, 2018.

[117] Dessislava Marinova, Jesus Gonzalo-Asensio, Nacho Aguilo, and Carlos Martín. MTBVAC from discovery to clinical trials in tuberculosis-endemic countries. *Expert review of vaccines*, 16(6):565–576, 2017.

[118] Daniel Elias, Hanna Akuffo, and Sven Britton. PPD induced in vitro interferon gamma production is not a reliable correlate of protection

against Mycobacterium tuberculosis. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 99(5):363–368, 2005.

[119] Hans-Willi Mittrücker, Ulrich Steinhoff, Anne Köhler, Marion Krause, Doris Lazar, Peggy Mex, Delia Miekley, and Stefan HE Kaufmann. Poor correlation between BCG vaccination-induced T cell responses and protection against tuberculosis. *Proceedings of the National Academy of Sciences*, 104(30):12434–12439, 2007.

[120] Benjamin MN Kagina, Brian Abel, Thomas J Scriba, Elizabeth J Hughes, Alana Keyser, Andreia Soares, Hoyam Gamieldien, Mzwandile Sidibana, Mark Hatherill, Sebastian Gelderbloem, et al. Specific T cell frequency and cytokine expression profile do not correlate with protection against tuberculosis after bacillus Calmette-Guerin vaccination of newborns. *American journal of respiratory and critical care medicine*, 182(8):1073–1079, 2010.

[121] Peter Andersen and T Mark Doherty. The success and failure of BCG—implications for a novel tuberculosis vaccine. *Nature Reviews Microbiology*, 3(8):656, 2005.

[122] Michele Tizzoni, Paolo Bajardi, Chiara Poletto, José J Ramasco, Duygu Balcan, Bruno Gonçalves, Nicola Perra, Vittoria Colizza, and Alessandro Vespignani. Real-time numerical forecast of global epidemic spreading: case study of 2009 A/H1N1pdm. *BMC medicine*, 10(1):165, 2012.

[123] Alessia Melegaro, Mark Jit, Nigel Gay, Emilio Zagheni, and W John Edmunds. What types of contacts are important for the spread of infections? using contact survey data to explore european mixing patterns. *Epidemics*, 3(3):143–151, 2011.

[124] Colin J Worby, Sandra S Chaves, Jacco Wallinga, Marc Lipsitch, Lyn Finelli, and Edward Goldstein. On the relative role of different age groups in influenza epidemics. *Epidemics*, 13:10–16, 2015.

[125] Pejman Rohani, Xue Zhong, and Aaron A King. Contact network structure explains the changing epidemiology of pertussis. *Science*, 330 (6006):982–985, 2010.

[126] Luigi Marangi, Grazina Mirinaviciute, Elmira Flem, Gianpaolo Scalia Tomba, Giorgio Guzzetta, Birgitte Freiesleben De Blasio, and Piero

Manfredi. The natural history of varicella zoster virus infection in norway: Further insights on exogenous boosting and progressive immunity to herpes zoster. *PloS one*, 12(5):e0176845, 2017.

[127] Simon Cauchemez, Alain-Jacques Valleron, Pierre-Yves Boëlle, Antoine Flahault, and Neil M Ferguson. Estimating the impact of school closure on influenza transmission from sentinel data. *Nature*, 452(7188):750–754, 2008.

[128] Niel Hens, Girma Minalu Ayele, Nele Goeyvaerts, Marc Aerts, Joel Mossong, John W Edmunds, and Philippe Beutels. Estimating the impact of school closure on social mixing behaviour and the transmission of close contact infections in eight european countries. *BMC infectious diseases*, 9(1):187, 2009.

[129] Sergio Arregui, María José Iglesias, Sofía Samper, Dessislava Marinova, Carlos Martin, Joaquín Sanz, and Yamir Moreno. Data-driven model for the assessment of mycobacterium tuberculosis transmission in evolving demographic structures. *Proceedings of the National Academy of Sciences*, 2018. ISSN 0027-8424. doi: 10.1073/pnas.1720606115. URL `http://www.pnas.org/content/early/2018/03/20/1720606115`.

[130] Jonathan M Read, Justin Lessler, Steven Riley, Shuying Wang, Li Jiu Tan, Kin On Kwok, Yi Guan, Chao Qiang Jiang, and Derek AT Cummings. Social mixing patterns in rural and urban areas of southern china. *Proceedings of the Royal Society of London B: Biological Sciences*, 281(1785):20140268, 2014.

[131] Guillaume Béraud, Sabine Kazmercziak, Philippe Beutels, Daniel Levy-Bruhl, Xavier Lenne, Nathalie Mielcarek, Yazdan Yazdanpanah, Pierre-Yves Boëlle, Niel Hens, and Benoit Dervaux. The french connection: the first large population-based contact survey in france relevant for the spread of infectious diseases. *PloS one*, 10(7):e0133203, 2015.

[132] Kathy Leung, Mark Jit, Eric HY Lau, and Joseph T Wu. Social contact patterns relevant to the spread of respiratory infectious diseases in hong kong. *Scientific reports*, 7(1):7974, 2017.

[133] Yoko Ibuka, Yasushi Ohkusa, Tamie Sugawara, Gretchen B Chapman, Dan Yamin, Katherine E Atkins, Kiyosu Taniguchi, Nobuhiko Okabe, and Alison P Galvani. Social contacts, vaccination decisions and influenza in Japan. *J Epidemiol Community Health*, pages jech–2015, 2015.

[134] Moses Chapa Kiti, Timothy Muiruri Kinyanjui, Dorothy Chelagat Koech, Patrick Kiio Munywoki, Graham Francis Medley, and David James Nokes. Quantifying age-related rates of social contact using diaries in a rural coastal population of Kenya. *PloS one*, 9(8): e104786, 2014.

[135] Marco Ajelli and Maria Litvinova. Estimating contact patterns relevant to the spread of infectious diseases in Russia. *Journal of Theoretical Biology*, 419:1–7, 2017.

[136] Olivier le Polain de Waroux, Sandra Cohuet, Donny Ndazima, Adam Kucharski, Aitana Juan-Giner, Stefan Flasche, Elioda Tumwesigye, Rinah Arinaitwe, Juliet Mwanga-Amumpaire, Yap Boum, et al. Characteristics of human encounters and social mixing patterns relevant to infectious diseases spread by close contact: A survey in Southwest Uganda. *bioRxiv*, page 121665, 2017.

[137] Alessia Melegaro, Emanuele Del Fava, Piero Poletti, Stefano Merler, Constance Nyamukapa, John Williams, Simon Gregson, and Piero Manfredi. Social Contact Structures and Time Use Patterns in the Manicaland Province of Zimbabwe. *PloS one*, 12(1):e0170459, 2017.

[138] UN. Population division database. *http://esa.un.org/unpd/wpp/index.htm*, 2016.

[139] Maria A Riolo and Pejman Rohani. Combating pertussis resurgence: One booster vaccination schedule does not fit all. *Proceedings of the National Academy of Sciences*, 112(5):E472–E477, 2015.

[140] Ana I Bento, Aaron A King, and Pejman Rohani. A simulation study on the relative role of age groups under differing pertussis transmission scenarios. *bioRxiv*, page 247007, 2018.

[141] Maria A Riolo, Aaron A King, and Pejman Rohani. Can vaccine legacy explain the british pertussis resurgence? *Vaccine*, 31(49):5903–5908, 2013.

[142] Kiesha Prem, Alex R Cook, and Mark Jit. Projecting social contact matrices in 152 countries using contact surveys and demographic data. *PLoS computational biology*, 13(9):e1005697, 2017.

[143] Mark EJ Newman. Mixing patterns in networks. *Physical Review E*, 67(2):026126, 2003.

[144] Rachel Byng-Maddick and Mahdad Noursadeghi. Does tuberculosis threaten our ageing populations? *BMC Infect Dis*, 16(1):119, 2016.

[145] Mark N Lobato, Kate Cummings, Don Will, and Sarah Royce. Tuberculosis in children and adolescents: California, 1985 to 1995. *Pediatr Infect Dis J*, 17(5):407–411, 1998.

[146] Sara Y Del Valle, James M Hyman, Herbert W Hethcote, and Stephen G Eubank. Mixing patterns between age groups in social networks. *Soc Networks*, 29(4):539–554, 2007.

[147] Elizabeth Miller, Katja Hoschler, Pia Hardelid, Elaine Stanford, Nick Andrews, and Maria Zambon. Incidence of 2009 pandemic influenza A H1N1 infection in England: a cross-sectional serological study. *Lancet*, 375(9720):1100–8, 2010.

[148] Paul J Birrell, Georgios Ketsetzis, Nigel J Gay, Ben S Cooper, Anne M Presanis, Ross J Harris, André Charlett, Xu-Sheng Zhang, Peter J White, Richard G Pebody, et al. Bayesian modeling to unmask and predict influenza A/H1N1pdm dynamics in London. *Proc Natl Acad Sci USA*, 108(45):18238–43, 2011.

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

[150] Peter R Donald, Ben J Marais, and Clifton E Barry III. Age and the epidemiology and pathogenesis of tuberculosis. 2010.

[151] Peter J Dodd, Elizabeth Gardiner, Renia Coghlan, and James A Seddon. Burden of childhood tuberculosis in 22 high-burden countries: a mathematical modelling study. *Lancet Glob Health*, 2(8):e453–e459, 2014.

[152] Peter J Dodd, Charalambos Sismanidis, and James A Seddon. Global burden of drug-resistant tuberculosis in children: a mathematical modelling study. *Lancet Infect Dis*, 16(10):1193–1201, 2016.

[153] Andrea T Cruz and Jeffrey R Starke. Clinical manifestations of tuberculosis in children. *Paediatr Respir Rev*, 8(2):107–117, 2007.

[154] Ben J Marais, Robert P Gie, H Simon Schaaf, Nulda Beyers, Peter R Donald, and Jeff R Starke. Childhood pulmonary tuberculosis: old wisdom and new challenges. *American journal of respiratory and critical care medicine*, 173(10):1078–1090, 2006.

[155] LJ Nelson and CD Wells. Global epidemiology of childhood tuberculosis [childhood tb]. *Int J Tuberc Lung Dis*, 8(5):636–47, 2004.

[156] Pedro Dornelles Picon, Sergio Luiz Bassanesi, Maria Luiza Avancini Caramori, Roberto Luiz Targa Ferreira, Carla Adriane Jarczewski, and Patrícia Rodrigues Vieira. Risk factors for recurrence of tuberculosis. *J Bras Pneum*, **33**(**5**):572–8, 2007.

[157] Robyn S Lee, Jean-François Proulx, Dick Menzies, and Marcel A Behr. Progression to tuberculosis disease increases with multiple exposures. *Eur Respir J*, pages ERJ–00893, 2016.

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

[159] Stephen J Millen, Pieter W Uys, John Hargrove, Paul D Van Helden, and Brian G Williams. The effect of diagnostic delays on the drop-out rate and the total delay to diagnosis of tuberculosis. *PLoS One*, **3**(**4**): e1933, 2008.

[160] Jorge J Moré. The Levenberg-Marquardt algorithm: implementation and theory. In *Numerical analysis*, pages 105–16. Springer, 1978.

[161] David W Dowdy, Jonathan E Golub, Richard E Chaisson, and Valeria Saraceni. Heterogeneity in tuberculosis transmission and the role of geographic hotspots in propagating epidemics. *Proceedings of the National Academy of Sciences*, 109(24):9557–9562, 2012.

[162] Mukund Uplekar, Sachin Atre, William A Wells, Diana Weil, Rafael Lopez, Giovanni Battista Migliori, and Mario Raviglione. Mandatory tuberculosis case notification in high tuberculosis-incidence countries: policy and practice. *European Respiratory Journal*, pages ERJ–00956, 2016.

[163] David W Dowdy, Christopher Dye, and Ted Cohen. Data needs for evidence-based decisions: a tuberculosis modeler's 'wish list'. *Int J Tuberc Lung Dis*, 17(7):866–77, 2013.

[164] Jan AC Hontelez, Mark N Lurie, Marie-Louise Newell, Roel Bakker, Frank Tanser, Till Bärnighausen, Rob Baltussen, and Sake J de Vlas. Ageing with HIV in South Africa. *AIDS (London, England)*, 25(13), 2011.

[165] Jamie T Griffin, Neil M Ferguson, and Azra C Ghani. Estimates of the changing age-burden of plasmodium falciparum malaria disease in sub-Saharan Africa. *Nat Commun*, 5, 2014.

[166] Mauricio L Barreto, Daniel Pilger, Susan M Pereira, Bernd Genser, Alvaro A Cruz, Sergio S Cunha, Clemax Sant'Anna, Miguel A Hijjar, Maria Y Ichihara, and Laura C Rodrigues. Causes of variation in BCG vaccine efficacy: examining evidence from the BCG REVAC cluster randomized trial to explore the masking and the blocking hypotheses. *Vaccine*, 32(30):3759–3764, 2014.

[167] PEM Fine and LC Rodrigues. Mycobacterial diseases. *The Lancet*, 335 (8696):1016–1020, 1990.

[168] Barry R Bloom and Paul EM Fine. The BCG experience: implications for future vaccines against tuberculosis. In *Tuberculosis*, pages 531–557. American Society of Microbiology, 1994.

[169] T Oettinger, M Jørgensen, A Ladefoged, K Hasløv, and P Andersen. Development of the mycobacterium bovis BCG vaccine: review of the historical and biochemical evidence for a genealogical tree. *Tubercle and lung disease*, 79(4):243–250, 1999.

[170] Timothy F Brewer. Preventing tuberculosis with bacillus calmette-guerin vaccine: a meta-analysis of the literature. *Clinical Infectious Diseases*, 31(Supplement_3):S64–S67, 2000.

[171] Sanjay P Zodpey, Sunanda N Shrikhande, et al. The geographic location (latitude) of studies evaluating protective effect of BCG vaccine and it's efficacy/effectiveness against tuberculosis. *Indian journal of public health*, 51(4):205, 2007.

[172] Paul EM Fine and Emilia Vynnycky. The effect of heterologous immunity upon the apparent efficacy of (eg BCG) vaccines. *Vaccine*, 16(20): 1923–1928, 1998.

[173] P D'Arcy Hart and Ian Sutherland. BCG and vole bacillus vaccines in the prevention of tuberculosis in adolescence and early adult life. *Br Med J*, 2(6082):293–295, 1977.

[174] Isabel Miceli, Isabel N De Kantor, Diana Colaiácovo, Graciela Peluffo, Irene Cutillo, Roberto Gorra, Roberto Botta, Silvia Hom, and HG Ten Dam. Evaluation of the effectiveness of BCG vaccination using the

case control method in buenos aires, argentina. *International journal of epidemiology*, 17(3):629–634, 1988.

[175] Feisal Abdullah al Kassimi, Mohamed Saleh al Hajjaj, Ibrahim Othman al Orainey, and Elijah Afolabi Bamgboye. Does the protective effect of neonatal BCG correlate with vaccine-induced tuberculin reaction? *American journal of respiratory and critical care medicine*, 152 (5):1575–1578, 1995.

[176] Sotiros D Chaparas, Clifford J Maloney, and Sally R Hedrick. Specificity of tuberculins and antigens from various species of mycobacteria. *American Review of Respiratory Disease*, 101(1):74–83, 1970.

[177] M Harboe, RN Mshana, O Closs, G Kronvall, and NH Axelsen. Cross-reactions between mycobacteria: Ii. crossed immunoelectrophoretic analysis of soluble antigens of BCG and comparison with other mycobacteria. *Scandinavian journal of immunology*, 9(2):115–124, 1979.

[178] Gillian F Black, Hazel M Dockrell, Amelia C Crampin, Sian Floyd, Rosemary E Weir, Lyn Bliss, Lifted Sichali, Lorren Mwaungulu, Huxley Kanyongoloka, Bagrey Ngwira, et al. Patterns and implications of naturally acquired immune responses to environmental and tuberculous mycobacterial antigens in northern Malawi. *The Journal of infectious diseases*, 184(3):322–329, 2001.

[179] RE Weir, PEM Fine, B Nazareth, S Floyd, GF Black, E King, C Stanley, L Bliss, K Branson, and HM Dockrell. Interferon-$\gamma$ and skin test responses of schoolchildren in southeast England to purified protein derivatives from Mycobacterium tuberculosis and other species of mycobacteria. *Clinical & Experimental Immunology*, 134(2):285–294, 2003.

[180] Ana Paula Ferreira, Antônio S Aguiar, Maycron WB Fava, José Otávio A Corrêa, Francisco M Teixeira, and Henrique C Teixeira. Can the efficacy of bacille calmette-guerin tuberculosis vaccine be affected by intestinal parasitic infections? *The Journal of infectious diseases*, 186(3):441–442, 2002.

[181] Sian Floyd, Jorg M Pönnighaus, Lyn Bliss, Prince Nkhosa, Lifted Sichali, Glyn Msiska, and Paul EM Fine. Kinetics of delayed-type hypersensitivity to tuberculin induced by bacille calmette-guerin vaccination in northern malawi. *The Journal of infectious diseases*, 186 (6):807–814, 2002.

[182] Wouter Hoefsloot, Jakko Van Ingen, Claire Andrejak, Kristian Ängeby, Rosine Bauriaud, Pascale Bemer, Natalie Beylis, Martin J Boeree, Juana Cacho, Violet Chihota, et al. The geographic diversity of nontuberculous mycobacteria isolated from pulmonary samples: an NTM-NET collaborative study. *European Respiratory Journal*, 42(6):1604–1613, 2013.

[183] Carroll E Palmer and Mary W Long. Effects of infection with atypical mycobacteria on BCG vaccination and tuberculosis. *American review of respiratory disease*, 94(4):553–568, 1966.

[184] Lise Brandt, Joana Feino Cunha, Anja Weinreich Olsen, Ben Chilima, Penny Hirsch, Rui Appelberg, and Peter Andersen. Failure of the mycobacterium bovis BCG vaccine: some species of environmental mycobacteria block multiplication of BCG and induction of protective immunity to tuberculosis. *Infection and immunity*, 70(2):672–678, 2002.

[185] Rogelio Hernandez-Pando, LARR Pavön, Kutzi Arriaga, Hector Orozco, Vicente Madrid-Marina, and Graham Rook. Pathogenesis of tuberculosis in mice exposed to low and high doses of an environmental mycobacterial saprophyte before infection. *Infection and immunity*, 65 (8):3317–3327, 1997.

[186] Caroline Demangel, Thierry Garnier, Ida Rosenkrands, and Stewart T Cole. Differential effects of prior exposure to environmental mycobacteria on vaccination with mycobacterium bovis BCG or a recombinant BCG strain expressing rd1 antigens. *Infection and immunity*, 73(4): 2190–2196, 2005.

[187] Sarah L Young, Lynn Slobbe, Rachel Wilson, Bryce M Buddle, Geofferey W de Lisle, and Glenn S Buchan. Environmental strains of mycobacterium avium interfere with immune responses associated with mycobacterium bovis BCG vaccination. *Infection and immunity*, 75 (6):2833–2840, 2007.

[188] NB Mantilla-Beniers and MGM Gomes. Mycobacterial ecology as a modulator of tuberculosis vaccine success. *Theoretical population biology*, 75(2-3):142–152, 2009.

[189] Mauricio L Barreto, Laura C Rodrigues, Sergio S Cunha, Susan Pereira, Miguel A Hijjar, Maria Yury Ichihara, Silvana C de Brito,

and Ines Dourado. Design of the brazilian bcg-revac trial against tuberculosis: a large, simple randomized community trial to evaluate the impact on tuberculosis of BCG revaccination at school age. *Controlled clinical trials*, 23(5):540–553, 2002.

[190] Laura C Rodrigues, Susan M Pereira, Sergio S Cunha, Bernd Genser, Maria Yury Ichihara, Silvana C de Brito, Miguel A Hijjar, Alvaro A Cruz, Clemax Sant'Anna, Ana Luiza Bierrenbach, et al. Effect of BCG revaccination on incidence of tuberculosis in school-aged children in brazil: the bcg-revac cluster-randomised trial. *The Lancet*, 366(9493): 1290–1295, 2005.

[191] Kenneth F Wallis. The two-piece normal, binormal, or double gaussian distribution: its origin and rediscoveries. *Statistical Science*, pages 106–112, 2014.

[192] Ahmet Soysal, Kerry A Millington, Mustafa Bakir, Davinder Dosanjh, Yasemin Aslan, Jonathan J Deeks, Serpil Efe, Imogen Staveley, Katie Ewer, and Ajit Lalvani. Effect of BCG vaccination on risk of mycobacterium tuberculosis infection in children with household tuberculosis contact: a prospective community-based study. *The Lancet*, 366(9495): 1443–1451, 2005.

[193] Jason R Andrews, Mark Hatherill, Hassan Mahomed, Willem A Hanekom, Monica Campo, Thomas R Hawn, Robin Wood, and Thomas J Scriba. The dynamics of QuantiFERON-TB gold in-tube conversion and reversion in a cohort of South African adolescents. *American journal of respiratory and critical care medicine*, 191(5):584–591, 2015.

[194] Helen McShane. Editorial commentary: understanding BCG is the key to improving it, 2013.

[195] WHO. *Global Tuberculosis Report 2015*. World Health Organization, 2015.

[196] Ann M Ginsberg, Morten Ruhwald, Helen Mearns, and Helen McShane. TB vaccines in clinical development. *Tuberculosis*, 99:S16–S20, 2016.

[197] Lew Barker, Luc Hessel, and Barry Walker. Rational approach to selection and clinical development of TB vaccine candidates. *Tuberculosis*, 92:S25–S29, 2012.

[198] Helen A Fletcher. Correlates of immune protection from tuberculosis. *Current molecular medicine*, 7(3):319–325, 2007.

[199] Kamlesh Bhatt, Sheetal Verma, Jerrold J Ellner, and Padmini Salgame. Quest for correlates of protection against tuberculosis. *Clinical and Vaccine Immunology*, 22(3):258–266, 2015.

[200] Michele D Tameris, Mark Hatherill, Bernard S Landry, Thomas J Scriba, Margaret Ann Snowden, Stephen Lockhart, Jacqueline E Shea, J Bruce McClain, Gregory D Hussey, Willem A Hanekom, et al. Safety and efficacy of MVA85A, a new tuberculosis vaccine, in infants previously vaccinated with BCG: a randomised, placebo-controlled phase 2b trial. *The Lancet*, 381(9871):1021–1028, 2013.

[201] Elisa Nemes, Hennie Geldenhuys, Virginie Rozot, Kathryn T Rutkowski, Frances Ratangee, Nicole Bilek, Simbarashe Mabwe, Lebohang Makhethe, Mzwandile Erasmus, Asma Toefy, et al. Prevention of M. tuberculosis Infection with H4: IC31 Vaccine or BCG Revaccination. *New England Journal of Medicine*, 379(2):138–149, 2018.

[202] M Elizabeth Halloran, Kari Auranen, Sarah Baird, Nicole E Basta, Steven E Bellan, Ron Brookmeyer, Ben S Cooper, Victor DeGruttola, James P Hughes, Justin Lessler, et al. Simulations for designing and interpreting intervention trials in infectious diseases. *BMC medicine*, 15(1):223, 2017.

[203] James M Trauer, Justin T Denholm, and Emma S McBryde. Construction of a mathematical model for tuberculosis transmission in highly endemic regions of the asia-pacific. *Journal of theoretical biology*, 358: 74–84, 2014.

[204] Dany Pascal Moualeu, Martin Weiser, Rainald Ehrig, and Peter Deuflhard. Optimal control for a tuberculosis model with undetected cases in cameroon. *Communications in Nonlinear Science and Numerical Simulation*, 20(3):986–1003, 2015.

[205] Peter R Cox. *Life tables*. Wiley Online Library, 1972.

[206] David Schoenfeld. Partial residuals for the proportional hazards regression model. *Biometrika*, 69(1):239–241, 1982.

[207] Yosef Hochberg and Yoav Benjamini. More powerful procedures for multiple significance testing. *Statistics in medicine*, 9(7):811–818, 1990.

[208] WHO. Tuberculosis database. *http://apps.who.int*, 2016.

[209] Herbert W Hethcote. The mathematics of infectious diseases. *SIAM review*, 42(4):599–653, 2000.

[210] Ben Bolker et al. bbmle: Tools for general maximum likelihood estimation, 2010.

[211] David Diez. Survival analysis in R. *OpenIntro. org*, 2013.

[212] Johanneke Kleinnijenhuis, Jessica Quintin, Frank Preijers, Leo AB Joosten, Daniela C Ifrim, Sadia Saeed, Cor Jacobs, Joke van Loenhout, Dirk de Jong, Hendrik G Stunnenberg, et al. Bacille Calmette-Guerin induces NOD2-dependent nonspecific protection from reinfection via epigenetic reprogramming of monocytes. *Proceedings of the National Academy of Sciences*, 109(43):17537–17542, 2012.

[213] Eva Kaufmann, Joaquin Sanz, Jonathan L Dunn, Nargis Khan, Laura E Mendonça, Alain Pacis, Fanny Tzelepis, Erwan Pernet, Anne Dumaine, Jean-Christophe Grenier, et al. BCG educates hematopoietic stem cells to generate protective innate immunity against tuberculosis. *Cell*, 172(1):176–190, 2018.

[214] Stefan HE Kaufmann, Sarah Fortune, Ilaria Pepponi, Morten Ruhwald, Lewis K Schrager, and Tom HM Ottenhoff. TB biomarkers, TB correlates and human challenge models: New tools for improving assessment of new TB vaccines. *Tuberculosis*, 99:S8–S11, 2016.

[215] Michele Tameris, Hennie Geldenhuys, Angelique KanyKany Luabeya, Erica Smit, Jane E Hughes, Samantha Vermaak, Willem A Hanekom, Mark Hatherill, Hassan Mahomed, Helen McShane, et al. The candidate TB vaccine, MVA85A, induces highly durable Th1 responses. *PloS one*, 9(2):e87340, 2014.

[216] A Williams, NP Goonetilleke, H McShane, Simon O Clark, Graham Hatch, SC Gilbert, and AVS Hill. Boosting with poxviruses enhances Mycobacterium bovis BCG efficacy against tuberculosis in guinea pigs. *Infection and immunity*, 73(6):3814–3816, 2005.

[217] Nilu P Goonetilleke, Helen McShane, Carolyn M Hannan, Richard J Anderson, Roger H Brookes, and Adrian VS Hill. Enhanced immunogenicity and protective efficacy against Mycobacterium tuberculosis of bacille Calmette-Guerin vaccine using mucosal administration and

boosting with a recombinant modified vaccinia virus Ankara. *The Journal of Immunology*, 171(3):1602–1609, 2003.

[218] Marta Romano, Sushila D'Souza, Pierre-Yves Adnet, Rachid Laali, Fabienne Jurion, Kamiel Palfliet, and Kris Huygen. Priming but not boosting with plasmid DNA encoding mycolyl-transferase Ag85A from Mycobacterium tuberculosis increases the survival time of Mycobacterium bovis BCG vaccinated mice against low dose intravenous challenge with M. tuberculosis H37Rv. *Vaccine*, 24(16):3353–3364, 2006.

[219] Frank AW Verreck, Richard AW Vervenne, Ivanela Kondova, Klaas W van Kralingen, Edmond J Remarque, Gerco Braskamp, Nicole M van der Werff, Ariena Kersbergen, Tom HM Ottenhoff, Peter J Heidt, et al. MVA. 85A boosting of BCG and an attenuated, phoP deficient M. tuberculosis vaccine both show protective efficacy against tuberculosis in rhesus macaques. *PloS one*, 4(4):e5264, 2009.

[220] H Martin Vordermeier, Bernardo Villarreal-Ramos, Paul J Cockle, Martin McAulay, Shelley G Rhodes, Tyler Thacker, Sarah C Gilbert, Helen McShane, Adrian VS Hill, Zhou Xing, et al. Viral booster vaccines improve Mycobacterium bovis BCG-induced protection against bovine tuberculosis. *Infection and immunity*, 77(8):3364–3373, 2009.

[221] Elma Z Tchilian, Christiane Desel, Emily K Forbes, Silke Bandermann, Clare R Sander, Adrian VS Hill, Helen McShane, and Stefan HE Kaufmann. Immunogenicity and protective efficacy of prime-boost regimens with recombinant ΔureC hly+ Mycobacterium bovis BCG and modified vaccinia virus Ankara expressing M. tuberculosis antigen 85A against murine tuberculosis. *Infection and immunity*, 77(2):622–631, 2009.

[222] Peter A Briss, Lance E Rodewald, Alan R Hinman, Abigail M Shefer, Raymond A Strikas, Roger R Bernier, Vilma G Carande-Kulis, Hussain R Yusuf, Serigne M Ndiaye, Sheree M Williams, et al. Reviews of evidence regarding interventions to improve vaccination coverage in children, adolescents, and adults12. *American journal of preventive medicine*, 18(1):97–140, 2000.

[223] JAC Sterne, LC Rodrigues, and IN Guedes. Does the efficacy of BCG decline with time since vaccination? *The international journal of tuberculosis and lung disease*, 2(3):200–207, 1998.

[224] Sergio Arregui, Alberto Aleta, Joaquin Sanz, and Yamir Moreno. Projecting social contact matrices to different demographic structures. *bioRxiv*, page 343491, 2018.

[225] Marcel Behr, Ruth Ellis, Joel Ernst, Tom Evans, JoAnne Flynn, Mark Hatherill, Daniel Hoft, Ajit Lalvani, Helen McShane, Edward Nardell, et al. Developing vaccines to prevent sustained infection with Mycobacterium tuberculosis. *Vaccine*, 33(26):3056–3064, 2015.

[226] Jason R Andrews, Farzad Noubary, Rochelle P Walensky, Rodrigo Cerda, Elena Losina, and C Robert Horsburgh. Risk of progression to active tuberculosis following reinfection with Mycobacterium tuberculosis. *Clinical infectious diseases*, 54(6):784–791, 2012.

[227] Christopher JL Murray, Alan D Lopez, et al. The global burden of disease, 1996.

[228] Christopher JL Murray and Arnab K Acharya. Understanding DALYs. *Journal of health economics*, 16(6):703–730, 1997.

[229] JA Fox Rushby and Kara Hanson. Calculating and presenting disability adjusted life years (DALYs) in cost-effectiveness analysis. *Health policy and planning*, 16(3):326–331, 2001.

[230] Marthe R Gold, David Stevenson, and Dennis G Fryback. HALYS and QALYS and DALYS, Oh My: similarities and differences in summary measures of population Health. *Annual review of public health*, 23(1): 115–134, 2002.

[231] Christopher JL Murray, Ryan M Barber, Kyle J Foreman, Ayse Abbasoglu Ozgoren, Foad Abd-Allah, Semaw F Abera, Victor Aboyans, Jerry P Abraham, Ibrahim Abubakar, Laith J Abu-Raddad, et al. Global, regional, and national disability-adjusted life years (DALYs) for 306 diseases and injuries and healthy life expectancy (HALE) for 188 countries, 1990–2013: quantifying the epidemiological transition. *The Lancet*, 386(10009):2145–2191, 2015.

[232] Nicholas J Kassebaum, Megha Arora, Ryan M Barber, Zulfiqar A Bhutta, Jonathan Brown, Austin Carter, Daniel C Casey, Fiona J Charlson, Matthew M Coates, Megan Coggeshall, et al. Global, regional, and national disability-adjusted life-years (DALYs) for 315 diseases and injuries and healthy life expectancy (HALE), 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *The Lancet*, 388(10053):1603–1658, 2016.

[233] Joshua A Salomon, Theo Vos, Daniel R Hogan, Michael Gagnon, Mohsen Naghavi, Ali Mokdad, Nazma Begum, Razibuzzaman Shah, Muhammad Karyana, Soewarta Kosen, et al. Common values in assessing health outcomes from disease and injury: disability weights measurement study for the Global Burden of Disease Study 2010. *The Lancet*, 380(9859):2129–2143, 2012.

[234] Robin Wood, Hua Liang, Hulin Wu, Keren Middelkoop, Tollulah Oni, Molebogeng X Rangaka, Robert J Wilkinson, Linda-Gail Bekker, and Stephen D Lawn. Changing prevalence of tuberculosis infection with increasing age in high-burden townships in South Africa. *The international journal of tuberculosis and lung disease*, 14(4):406–412, 2010.