Mathematical modelling of epidemiological systems:
from data and equations to public health implications

IFT - UNESP
INSTITUTO DE FÍSICA TEÓRICA

Caroline Franco

IFT-T.008/2021

Franco, Caroline.
F825m Mathematical modelling of epidemiological systems: from data and equations to public health implications / Caroline Franco. – São Paulo, 2021
121 f.

Tese (doutorado) – Universidade Estadual Paulista (Unesp), Instituto de Física Teórica (IFT), São Paulo
Orientador: Roberto André Kraenkel

1. Modelos matemáticos. 2. Doenças transmissíveis. 3. Biologia - Modelos matemáticos. 4. COVID-19 (Doença). 5. Malária. I. Título

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca do Instituto de Física Teórica (IFT), São Paulo. Dados fornecidos pelo autor(a).

Mathematical modelling of epidemiological systems:
from data and equations to public health implications

This thesis has received financial support from CAPES and FAPESP (Brazil)

Mathematical modelling of epidemiological systems:
from data and equations to public health implications

THESIS

submitted to the Institute of Theoretical Physics
São Paulo State University, Brazil

as a partial fulfilment of the requirements for the degree of

Doctor of Science

December 2021

Caroline Franco

Roberto André Kraenkel (supervisor)

Examination Committee

Prof. Dr. Roberto André Kraenkel (supervisor)
Instituto de Física Teórica, Universidade Estadual Paulista

Prof. Dr. Cláudia Codeço
Programa de Computação Científica, Fundação Oswaldo Cruz

Prof. Dr. Edgardo Brigatti
Instituto de Física, Universidade Federal do Rio de Janeiro

Prof. Dr. Marcelo Ferreira da Costa Gomes
Programa de Computação Científica, Fundação Oswaldo Cruz

Prof. Dr. Suani Pinho
Instituto de Física, Universidade Federal da Bahia

Instituto de Física Teórica - UNESP
R. Dr. Bento Teobaldo Ferraz 271, bloco II
01140-070 São Paulo, Brasil

“To Summarise briefly: A white rabbit is pulled out of a top hat. Because it is an extremely large rabbit, the trick takes many billions of years. All mortals are born at the very tip of the rabbit’s fine hairs. where they are in a position to wonder at the impossibility of the trick. But as they grow older they work themselves even deeper into the fur. And there they stay. They become so comfortable they never risk crawling back up the fragile hairs again. Only philosophers embark on this perilous expedition to the outermost reaches of language and existence. Some of them fall off, but others cling on desperately and yell at the people nestling deep in the snug softness, stuffing themselves with delicious food and drink.

‘Ladies and gentlemen,’ they yell, ‘we are floating in space!’ but none of the people down there care.

‘What a bunch of troublemakers!’ they say. And they keep on chatting: Would you pass the butter, please? How much have our stocks risen today? What is the price of tomatoes? Have you heard that Princes Di is expecting again? ”

Jostein Gaarder, Sophie’s World

Abstract

The mathematical modelling of infectious diseases is an extremely interdisciplinary field, where we need to amalgamate different areas of expertise, cultures and scientific backgrounds. Based on the available biological and epidemiological data, compartmental models can be tailored to investigate different research questions regarding a disease spread dynamics within a population. In our COVID-19 and malaria modelling exercises, we used different model structures to address different issues.

Through the CoMo model framework, we have produced a robust model to compare non-pharmaceutical intervention scenarios and assess their relative effects on reducing cases and deaths, during the COVID-19 pandemic. This model was broadly applied to inform policy making at national and international levels. Taking into account inter-household connectivity heterogeneities, we have also contributed towards bridging the gap between population and household-level models for communicable diseases, using insights from network percolation theory.

Within the COVID-19 modelling context, we have also developed other model structures to address different research questions. Simplified age-structuring and intervention implementations were used to explore and optimise the outcome of single interventions (*viz.* school reopening and vaccination) or estimate epidemiologically relevant parameters (*viz.* transmission rates for a new variant).

In the context of endemic malaria, we presented single-species models that can and have been useful to inform health policymakers. Nevertheless, since multiple species can coexist in some parts of the world, it is valuable to evaluate their combined burden while taking into account their interactions. In this context, a novel multi-species modelling framework is proposed and its applicability to weak interaction regimes is proven analytically. Such framework provided theoretical support for model structures already being applied to support health policy making.

Our contributions encompass both the development of theoretical methodologies to support more robust models and the development of such models, aiming to support health decision making.

Keywords: mathematical modelling; compartmental models; infectious diseases; SARS-CoV-2; malaria.

Areas: Physics; Complex systems; Mathematical Biology.

Resumo

A modelagem matemática de doenças infecciosas é um campo extremamente interdisciplinar no qual se faz necessário combinar conhecimentos obtidos a partir de diferentes especialidades, culturas e disciplinas científicas. Com base em dados biológicos e epidemiológicos disponíveis, modelos compartimentais podem ser construídos sob medida visando responder a diferentes questões científicas referentes à dinâmica de propagação de uma doença dentro de uma dada população. Em nossos modelos descrevendo a dinâmica de COVID-19 e malária aqui apresentados, utilizamos diferentes estruturas para solucionar diferentes problemas.

Através da abordagem introduzida pelo modelo CoMo, produzimos um modelo robusto para simular o efeito de diferentes cenários de intervenções não-farmacêuticas na redução de casos e mortes, durante a pandemia de COVID-19. Esse modelo foi utilizado para apoiar a implementação de políticas públicas a níveis nacional e internacional. Levando em conta heterogeneidades na rede de contatos entre domicílios, também contribuímos ao preencher a lacuna metodológica existente entre modelos a nível populacional e domiciliar, usando idéias extraídas da teoria de percolação em redes.

Ainda no contexto de modelos para COVID-19, desenvolvemos outras estruturas visando responder diferentes questões científicas. Simplificações na estrutura etária e na implementação de intervenções foram utilizadas para que se pudesse melhor avaliar e otimizar o efeito de intervenções únicas (*viz.* reabertura de escolas e vacinação) ou estimar parâmetros de relevância epidemiológica (*viz.* transmissibilidade de novas variantes).

No contexto de malária endêmica, descrevemos modelos independentes para diferentes espécies, que são atestadamente úteis para informar tomadas de decisão no contexto de saúde pública. Visto que múltiplas espécies coexistem no meio ambiente, também levamos em conta a avaliação de sua carga combinada e o efeito de possíveis interações interespecíficas. Neste contexto, propusemos uma nova abordagem metodológica que viabiliza o desenvolvimento de modelos incluindo múltiplas espécies e provamos analiticamente sua validade em regimes de interações fracas. Tal abordagem proporcionou o devido suporte teórico para estruturas de

modelos já sendo adotadas para informar tomadas de decisão em saúde pública.

Nossas contribuições englobam tanto o desenvolvimento de metodologias teóricas que fundamentam a construção de modelos mais robustos, quanto o desenvolvimento de tais modelos visando apoiar tomadas de decisão em saúde pública.

Palavras Chaves: modelos matemáticos; modelos compartimentais; doenças infecciosas; SARS-CoV-2; malária.

Áreas do conhecimento: Física; Sistemas complexos; Biologia matemática.

Contents

Acknowlegments

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001 and the São Paulo Research Foundation (FAPESP) grant numbers 2019/26310-2 and 2017/26770-8.

Throughout the writing of this dissertation I have received a great deal of support and encouragement.

I would first like to thank Prof. Roberto Kraenkel, my supervisor in the São Paulo State University, who accepted me as a PhD student in 2018 with a bold research proposal, supported me and guided me through the evolving challenges and opportunities that came with it.

I also thank Prof. Lisa J. White, my supervisor during my period in the University of Oxford, for all the brilliant opportunities to learn from and engage with world-class researchers in modelling for global health.

Both acted as mentors and guides, supporting me with their expertise and advising me through important decisions regarding my career. I feel very lucky to have had the opportunity to work with and learn from such role models.

I thank my research group colleagues in São Paulo and Oxford for the many eye-opening discussions, long meetings and intense coding sessions. Without our collaborative and friendly environment, this thesis would not be half of what it is.

Finally, I am thankful for my family and friend's constant source love, encouragement and support throughout all the moving, travelling and moments of uncertainty. Each of you know your role and I am grateful for all inputs, as they helped me go further towards chasing my dreams.

Preface

> *"I simply wish that, in a matter which so closely concerns the wellbeing of the human race, no decision shall be made without all the knowledge which a little analysis and calculation can provide."*
>
> (Daniel Bernoulli, 1760)

Mathematical modelling can be described as a translation of a real-world problem into a mathematical formulation with the aim to represent, analyse, make predictions or gain insight onto the investigated phenomena. In the context of infectious disease modelling, such formulations can be useful tools to understand the underlying biological mechanisms, support public health decision-making or forecast. A limitless variety of models can be developed and used to represent a single epidemiological system, depending of the mathematical approach selected to represent it.

Despite its usefulness and potential impact as a tool for stakeholders, epidemiological models only started to be more systematically studied very recently. Bernoulli's 1760 analysis on the benefits of universal vaccination against smallpox [16; 17] was an isolated early exception. After that, only at the beginning of the 20th century Hamer and Ross would start investigating, respectively, measles and malaria epidemiological dynamics through mathematical modelling [79; 128], followed by Kermack and McKendrick analysis on the necessary conditions for an epidemic to occur [89; 32].

Note that, even though these seminal models were able to elegantly explain some observed epidemiological patterns, they carried many underlying assumptions regarding the population affected, the disease spread and the recovery mechanisms. Hence, the need to develop more complex models became swiftly prominent, either

with the aim of relaxing some of the assumptions, or being more specific to the biology of certain diseases.

“All models are wrong, some are useful” is a common maxim, generally attributed to the statistician George Box. A corollary interpretation of this aphorism is that, even though there is no right model, there are good modelling approaches. These are the ones informed by our aims, research questions and available data. Simple models can be useful if they provide relevant information for a particular purpose, but the complexity of a model will increase together with the number of functions it performs. The challenge lies in keeping a balance between biological realism and model intricacy, since it might not always be worth adding extra degrees of freedom if they come with additional uncertainties. In a global health context, useful models are the ones validated by epidemiological data and able to inform decision-making.

Having worked closely with other experienced modellers, biologists, epidemiologist, public health experts and policy makers, we have tailored each model framework presented here to address specific public health issues or answer specific research questions. Some of these questions questions were inspired by an eagerness to better understand biological and epidemiological mechanisms, but most of them were simply induced by demands to support health decision making. Note that this thesis and many of the models hereby described have been written during the COVID-19 pandemic, the yet most significant and disrupting health crisis seen in the XXI century. Such shock has therefore reshaped this thesis outcomes and impact, as well as it has shakensw the whole world.

Here, we broach a few case studies where compartmental models were developed to address different research questions in the public health context. We structure the content in 9 main chapters, divided in 3 independent parts: one introductory part and two parts comprising scientific contributions where mathematical modelling was used in the contexts of an emerging pandemic and an ongoing endemic. Parts II and III contain their own introduction and closely follow the structure of our published papers or manuscripts in preparation.

Part I provides a theoretical background on mathematical modelling of infectious diseases. Chapter 1 introduces some key concepts and basic model compartmental structures that we will later build on, when developing more complex models for specific diseases. Chapter 2 reports our experience collecting and processing epidemiological data, whilst also exploring a few of the general challenges involved with data wrangling.

Part II describes our experience using mathematical models to inform decision making at public health level in the context of the progressing COVID-19 pandemic. Chapter 3 delineates some of the key events throughout the evolution of the COVID-19 pandemics and reports the development participatory modelling networks to support policy making in Brazil and other countries. Chapter 4 details the theoretical framework we used to model the COVID-19 epidemiological

dynamics, including our contribution in developing a modelling approach to bridge the gap between compartmental and household-level models. Chapter 5 expands on the applications of our mathematical models on informing policy-making in Brazil, including the use of models to analyse the transmissibility of a new variant, school reopening strategies and vaccination scenarios. We also include considerations about lessons learned on model communication to stakeholders.

Part III describes the development of a multi-species malaria model. Chapter 6 contextualises the malaria endemic panorama in Brazil, which is concentrated in the Amazon region. Chapter 7 introduces typical mathematical models to describe the disease dynamics for the two most common species of *Plasmodium*, focusing on its application to model malaria in the Brazilian Amazon region. Chapter 8 describes a general theoretical framework developed to model multi-species epidemiological dynamics, which is an approximation technique that can be used to simplify any multi-species infectious disease dynamic model, including malaria.

Part I

Introduction

1

Compartmental models for infectious diseases

In this chapter, we will introduce some of the building blocks used in mathematical modelling of infectious diseases, including paradigmatic toy models that can be used as a starting point for building more complex models, such as the ones presented in parts II and III of this thesis.

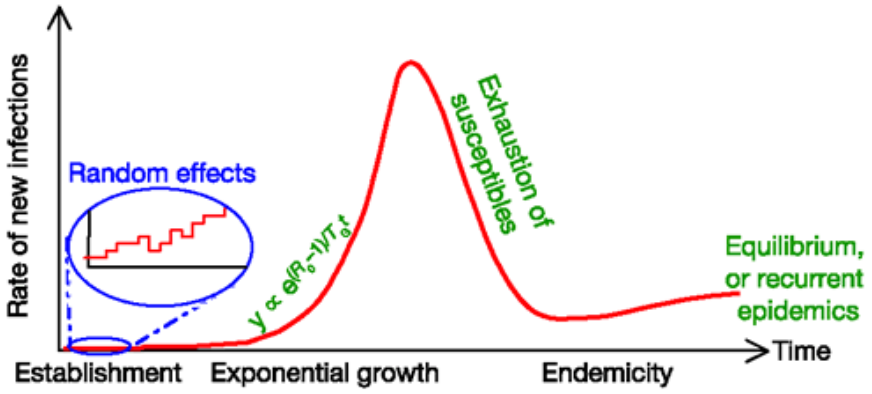


Figure 1.1: Classic epidemic profile. (Source: Ferguson, N., Keeling, M., John Edmunds, W. et al. [66])

Looking at a typical epidemic curve (figure 1.1), we can identify a few distinctive phases:

- Epidemic establishmentment: a small number of cases, subject to stochastic fluctuations, is introduced in a fully susceptible population.
- Exponential growth: a rapid rise in the number of cases that can usually be described by an exponential growth function.
- Depletion of susceptibles: with the a decreasing susceptible population pool, the exponential description for the infected population growth is no longer accurate.

- Post-epidemic phase: after peaking, due to depletion of susceptibles, the disease may disappear, become endemic or recur periodically.

At the onset of an epidemic, when a few infectious individuals are introduced into a fully susceptible population, it is reasonable to assume that the disease spread is not limited by the size of the pool of susceptible individuals. Hence, the number of infectious cases can be approximately described by a linear ordinary differential equation such as

$$\frac{dI}{dt} = \lambda I \tag{1.1}$$

with solution

$$I = I_0 \exp^{\lambda t} \tag{1.2}$$

where I_0 is the number of infected cases at $t = 0$ and λ is the transmission rate, a quantity which can be written as a function of other system parameters, such as its *basic reproduction number* (more on that in section 1.2).

Nonetheless, with the exhaustion of the susceptible population, such single-variable exponential approximation becomes inaccurate whereas the interaction between susceptible, infectious and recovered populations becomes increasingly relevant. Hence, more complex models are required to represent the system dynamics. For instance, one can use a susceptible-infected-removed (SIR) model, such as the one described in section 1.1. This is an example of a *compartmental model*, where sub-groups within a population are represented by model *compartments*. Each compartment is translated, in mathematical terms, as a dynamical variable, which in turn has its dynamics described by a differential equation.

1.1 SIR models

One of the simplest compartmental models one can write to simulate a disease propagation is the susceptible-infected-removed (SIR) system, introduced by Kermack and McKendrick [89]. Three variables (S, I and R) represent the status of the individuals in the susceptible, infected and removed (which includes recovered/immune to reinfection and deceased) sub-populations, referred as the model's compartments (figure 1.2).

The flow between compartments can be modelled by the multiplication between a rate and the outcoming compartment state variable at time t. Each state variable (compartment) can then be associated with a differential equation, where such flows between compartments are encoded.

Note that, though this flow rates can be constants parameter, they can also be functions of time. In dynamic models, the rate at which susceptible individuals become infected (namely, the system's *force of infection*, $\lambda(t)$) depends on the

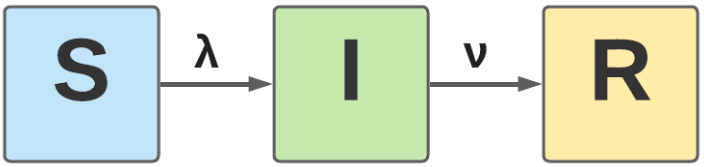


Figure 1.2: Diagram representing a susceptible-infected-recovered model compartments (S, I and R) and flows between compartments. λ and ν represent the force of infection and the recovery rate, respectively.

proportion of the population that is infected per unit time (I/N, N being the total population).

We can write the following system of equations to represent a SIR system with dynamic transmission:

$$\frac{dS}{dt} = -\lambda S \tag{1.3}$$

$$\frac{dI}{dt} = \lambda S - \nu I \tag{1.4}$$

$$\frac{dR}{dt} = \nu I \tag{1.5}$$

where

$$\lambda = \beta \frac{I}{N} \tag{1.6}$$

is the system's *force of infection* and $N = S + I + R$ is the total population, assumed constant. Here we also introduce β as the disease's *transmission rate* and ν as its *recovery rate*.

Inspecting equation 1.4, we can note that whenever $(\frac{\beta S}{N} - \nu) < 0$, the rate of change of the the infected sub-population is negative [10]. At this point it is useful to define a new quantity: the system's basic reproduction number, $R_0 = \frac{\beta}{\nu}$, which will be further discussed in the following section.

1.2 Basic reproduction number (R_0)

An epidemic's *basic reproduction number* (R_0) is an important quantity that translates how fast an epidemic develops within a population. It can be defined as the *average number of new infections that result from each single infection* introduced into a totally susceptible population. Mathematically, we can write

$$R_0 = \beta T_D \,, \tag{1.7}$$

where T_D is the duration of infection and β is the *effective contact rate per unit time*, an important quantity often referred as *transmission rate*.

In the case of *contagious transmission* (i.e., direct or indirect contact between an infected person and a susceptible person is required for transmission), the transmission rate can be written as

$$\beta = pc\,, \tag{1.8}$$

where p is the probability of infection given a contact and c is the number of contacts per unit time.

If we also introduce $\nu = 1/T_D$ as the disease's recovery rate, we can write, on the one hand,

$$R_0 = \frac{\beta}{\nu}\,. \tag{1.9}$$

On the other hand, we can also write the system's transmission rate, λ, as

$$\lambda = \beta - \nu\,. \tag{1.10}$$

Hence,

$$\lambda = (R_0 - 1)/T_D \tag{1.11}$$

and equation 1.2 becomes

$$I = I_0 \exp^{(R_0-1)/T_D t}\,. \tag{1.12}$$

Note that R_0, as defined in equation 1.7, explicitly depends on transmission rates which can only be defined at populations level. Hence, R_0, and therefore the epidemic's exponential growth rate, is not only defined by a disease's biology, but also by cultural and demographic characteristics of each specific population affected by a disease.

1.3 Effective reproduction number (R_t)

As an epidemic unfolds and the proportion of susceptible individuals in a population decreases, R_0 can no longer represent the average number of new cases generated by each previous infection. Therefore, we need to define a new quantity useful at any stage during the epidemic. The *effective reproduction number* (R_t) is, hence, defined as *the average number of secondary cases per infectious case in a population containing both susceptible and non-susceptible individuals* and can be calculated as:

$$R_t = \beta(t) T_D \frac{S(t)}{N}\,. \tag{1.13}$$

Note that R_t is an implicit function of time, not only because $S(t)$ is dynamic, but also because we allow $\beta(t)$ to change in time (which can happen when the disease's transmissibility changes in time, e.g. due to interventions).

Finally, note that, at $t = 0$, $S(0) = N$ and $\beta(0)$ are constants. Hence,

$$R(0) = \beta(0)DS(0)/N = \beta(0)D\,, \tag{1.14}$$

which is, by definition, R_0. Furthermore, if we consider $\beta(t) = \beta$ constant, we can write

$$R(t) = R_0\frac{S(t)}{N}\,. \tag{1.15}$$

1.3.1 Herd Immunity Threshold

R_t is a useful definition because it indicates whether the incidence of infections is increasing ($R_t > 1$) or decreasing ($R_t < 1$). Identifying when a population is under each of these regimes it is essential for decision making. Moreover, it can be useful to determine when the population switches between these two regimes (or when the epidemic saturates).

Let x^* be the critical proportion of susceptibles reached when $R_t = 1$. Thus, using equation 1.15,

$$R_0x^* = 1 \implies x^* = \frac{1}{R_0}\,. \tag{1.16}$$

It can be even more critical to determine a threshold value for the proportion of individuals that need to be immune before transmission drops to a level where incidence starts decreasing. More strictly, the *herd immunity threshold* (HIT) can be defined as the critical proportion of *non-susceptible* individuals in a population so that the epidemic saturates, being evaluated as

$$HIT = 1 - \frac{1}{R_0}\,. \tag{1.17}$$

In one hand, if the epidemic develops naturally (no interventions) the population would be protected from new outbreaks above the HIT. Yet a proportion of $1 - \frac{1}{R_0}$ of the individuals in the population would need to be infected or recovered to reach this regime (note that the HIT is defined in terms of population susceptibility, not strictly immunity). In the other hand, considering, for instance, a mass vaccination scenario, this number can be used to inform the minimum number of individuals that need to be immunised to control the spread of an epidemic.

Finally, note that the definition in equation 1.17 is only valid if we assume that natural recovering leads to long-term immunity and that there is no antigenic variation (which could lead to recurring outbreaks). This would imply that we are assuming an underlying susceptible-infected-removed dynamics, which we will explore in details in the following section.

R_0, HIT and stability analysis for the SIR model

For R_0 as defined for the system 1.3-1.5, a major epidemic will only occur if $R_0 \frac{S}{N} \geq 1$, which implies that the critical proportion of susceptibles, $x^* = \frac{S}{N}$, needs to be

$$x^* \geq \frac{1}{R_0} \tag{1.18}$$

for the epidemic to occur; a result consistent with what we explored in section 1.3.1. From the perspective of introducing a few infected individuals into a fully susceptible population, i.e. $S_0 \approx N$, the condition for the epidemic to develop within this fully susceptible population is simply

$$R_0 > 1\,. \tag{1.19}$$

This result can also be obtained from a stability analysis of the system's equilibria. Note that this is a non-linear set of equations, as the dynamic definition for λ introduces non-linear terms to the differential equations, but it is still simple enough so it is possible to analytically perform a stability analysis. The system 1.3-1.5 has at least two fixed point solutions. One of them represents a disease-free equilibrium, i.e.

$$\mathbf{X_1} = \{S_1, I_1, R_1\} = \{N, 0, 0\}\,, \tag{1.20}$$

The other one represents a state where the disease spread throughout the population, so all individual got infected and recovered:

$$\mathbf{X_2} = \{S_2, I_2, R_2\} = \{0, 0, N\}\,. \tag{1.21}$$

As R can always be obtained by $N - S - I$ and equations for S and I do not depend on R, we can analyse the system's stability by looking at the two eigenvalues obtained when the sub-system 1.3-1.4 is linearised, which are given by

$$\lambda_A = -\frac{\beta I}{N}\,, \tag{1.22}$$

$$\lambda_B = \frac{\beta S}{N} - \nu\,. \tag{1.23}$$

For the disease-free equilibrium, we have

$$\lambda_A|_{X_1} = 0\,, \tag{1.24}$$

$$\lambda_B|_{X_1} = \beta - \nu\,, \tag{1.25}$$

which means that X_1 is stable (though not asymptotically stable along the S axis) for $\beta - \nu < 0$, which is equivalent to the condition

$$R_0 < 1. \tag{1.26}$$

For the fully recovered population equilibrium, we have

$$\lambda_A|_{X_2} = 0\,, \tag{1.27}$$
$$\lambda_B|_{X_2} = -\nu\,, \tag{1.28}$$

which is stable (but not asymptotically stable) for all values of ν.

1.3.1. *Theorem (Threshold for SIR Epidemic). The disease-free steady state is stable (but not asymptotically stable) if $R_0 < 1$, so that the disease dies out; unstable if $R_0 > 1$, so that an epidemic may potentially occur [33].*

Summarising, as presented in Britton [33], there are two potentially stable fixed points and the steady-state solution will depend both on R_0 and the initial conditions.

The baseline disease dynamic represented by the system 1.3-1.5 is a depiction of systems with a clear removed sub-population, like most childhood diseases, such as measles [33]. Another way to represent recovery, when immunity usually does not last (e.g. micro-parasitic infections like tuberculosis, meningitis or gonorrhoea), would be through a susceptible-infected-susceptible (SIS) structure. In such a system, the recovering flow (νI) leaves the infected compartment and enters the susceptible compartment. In this case, R_0 represents a threshold that determines whether the disease dies out or becomes endemic. In other words, it determines whether the disease-free or the endemic equilibrium will be stable.

If recovery or removal is not a relevant part of the disease's dynamics (e.g. AIDS), one can use a susceptible-infectious (SI) model, which has an even simpler structure. It can be seen as a sub-model of the SIR model when $\nu = 0$. Note that, in this case, there is no recovery and the disease-free equilibrium is always unstable, so the disease always spreads throughout a population once it has been introduced.

The discussed model structures can be further built on as more mechanisms become relevant or assumptions need to be relaxed. For example, the *closed population* assumption might not be good to represent an endemic disease, which is prevalent in a population for an extended amount of time. In fact, it can be shown that adding births and deaths to the SIR dynamics introduces a stable endemic (i.e., coexistence) steady state with $S^* = N/R_0$ (where $R_0 = \frac{\beta}{\nu+\mu}$, with μ being the death rate) to the system [33]. Other useful model structure will be presented in the next section, where we refine the baseline SIR model introducing extra flows and compartments.

1.4 SEIRS models

In this section, we will build on the original SIR model structure, so we can represent a system with a latent period, waning immunity and population dynamics.

It is appropriate to take a latent period into account when it takes a significant amount of time for the initially inoculated pathogens to reproduce, before active transmission occurs [88], i.e., there are individuals that are *infected* but not yet *infectious*: the *exposed* sub-population. The addition of a exposed (E) compartment to the model structure is equivalent to introducing a time delay that slows the dynamics of a SIR system, with equivalent stability properties.

Another refinement to our model structure would be to consider waning immunity, which means assuming that immunity lasts for a limited period of time before a recovered individual is once again susceptible.

Finally, we will include human demographics terms, which are relevant when the time-frame of the disease dynamics is comparable to that of human demographics. This leads to a SEIRS model structure such as the one depicted in figure 1.3.

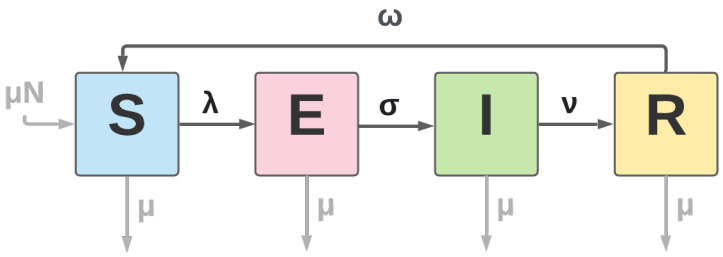


Figure 1.3: Diagram representing a susceptible-exposed-infected-recovered model compartments (S, E, I and R) with waning immunity and human demographics (birth and death rates represented by the rate μ). λ represents the force of infection, ν is the recovery rate, $1/\sigma$ is the latent period and ω is the immunity-loss rate.

If β, λ and ν are defined as in the previous section, we can write a SEIRS dynamics as:

$$\frac{dS}{dt} = \mu N - \lambda S + \omega R - \mu S \tag{1.29}$$

$$\frac{dI}{dt} = \lambda S - \sigma E - \mu E \tag{1.30}$$

$$\frac{dI}{dt} = \sigma E - \nu I - \mu I \tag{1.31}$$

$$\frac{dR}{dt} = \nu I - \omega R - \mu R \tag{1.32}$$

where

$$\lambda = \beta \frac{I}{N} \tag{1.33}$$

is the system's *force of infection* and $N = S + E + I + R$ is assumed constant, as the birth and death rates are assumed to be equal to μ. Note that here we also assume no vertical transmission (as all newborns are enter the susceptible compartment). $\frac{1}{\sigma}$

is the latent period and ω is the rate at which immunity is lost. Note that the SEIRS structure can be reduced to a SEIR if $\omega = 0$ (no immunity loss) and waning immunity does not affect the endemic steady state existence and stability, which can be analysed analytically in this case.

From equations 1.29-1.30 and $N = S + E + I + R$, we can obtain that, for the endemic equilibrium $\mathbf{X}^* = \{S^*, E^*, I^*, R\}$ where $I^* \neq 0$,

$$S^* = \frac{N(\nu+\mu)(\sigma+\mu)}{\beta\sigma} \equiv \frac{N}{R_0}\,, \tag{1.34}$$

$$E^* = \frac{(\mu+\nu)}{\sigma} I^*\,, \tag{1.35}$$

$$I^* = \frac{N\sigma(\mu+\omega)(R_0-1)}{\beta\sigma + \omega R_0(\nu+\mu)}\,, \tag{1.36}$$

$$R^* = N - S^* - E^* - I^*\,. \tag{1.37}$$

Above, we have also introduced the expression for the basic reproduction number, which is now

$$R_0 = \frac{\beta\sigma}{(\nu+\mu)(\sigma+\mu)}\,, \tag{1.38}$$

as it represents the product between the *transmission rate* (β), *duration of infection* ($\frac{1}{\nu+\mu}$) and *probability of surviving the exposed stage* ($\frac{\sigma}{\sigma+\mu}$).
Note the steady state solution X^* only exists for $R_0 > 1$. Another fixed point for the system wold correspond to the trivial disease-free state: $\mathbf{X_0} = \{N, 0, 0, 0\}$. Though its existence does not depend on the value of R_0, its stability will.

1.4.1. *Theorem (Threshold for SEIRS Endemic). The disease-free steady state is asymptotically stable if $R_0 \leq 1$, so that the disease dies out; unstable if $R_0 > 1$. In the latter case, the endemic steady state is feasible and asymptotically stable [88].*

Once again, R_0 acts as a determinant for the system's steady state dynamics. Note that the introduction of demographics into the model structure allows for existence and stability of an endemic steady state.

1.5 Outputs of interest

After implementing a compartmental model to simulate a system of interest and numerically integrating it, we can straight-forwardly output time-series for the predicted number of susceptible, infectious and recovered individuals. However, in an epidemiological context, such figures might not necessarily relate to quantities measured in the real world, such as *Prevalence* and *Incidence*.

One can define the *"true" prevalence* as the percentage of infectious individuals among the total *population at risk*, i.e.,

$$Prev(t) = 100 \cdot \frac{I}{P} \quad [\%]. \tag{1.39}$$

Nevertheless, it is not usually reasonable to assume that all infectious cases have been detected and reported to the local health system. Hence, it might be more useful to define a *test positive prevalence*, such as

$$Prev_{test}(t) = 100 \cdot \frac{\rho I}{P} \quad [\%], \tag{1.40}$$

where ρ is a parameter that can take into account reporting rates, testing rates and/or their sensitivity.

Another epidemiological output of interest would be the disease's *incidence* in a population, which is a measure of the number of new cases reported within a certain time range. For instance, the World Health Organization defines the malaria case incidence as the number of *cases per 1000 population at risk per year* [158], i.e.,

$$Inc(t) = 1000 \cdot \frac{\rho I}{P} \quad \left[\frac{cases\,/\,1000\,people}{year}\right], \tag{1.41}$$

where ρ is defined as above.

1.6 Conclusions

In this chapter we introduced a few relevant compartmental model structures used to represent infectious disease dynamics, the qualitative analysis techniques that can be applied to such systems and relevant quantities of interest.

We have shown that, evaluating the basic reproduction number for a disease spreading in a particular population, it is possible to anticipate whether an epidemic will occur in a population. Using stability analysis, we can determine whether an endemic state exists and its corresponding prevalence level. Such approach can also be helpful in determining whether disease eradication is possible.

The mathematical modelling frameworks presented here can be good simplifications of real-life disease dynamics, as long as fact-based assumptions lead to the development of a relevant model structure. Heterogeneities such as population structuring and spatial aspects can be taken into account when relevant, pondering the extra complexity added by additional degrees of freedom. For instance, we chose to use a SEIR-based structure to represent the COVID-19 dynamics, explored in part II of this thesis. For COVID-19, considering the latent period was essential, hence we build on the SEIRS structure, adding age heterogeneities (which very characteristics for SARS-CoV-2) and extra infectious compartments.

Eradication, testing and other control strategies can also be included in such models, for instance, by considering their effect onto the system's force of infection or by adding additional compartments for the treated sub-populations. In part III, we use a SIRS-like model structure to represent the epidemiological dynamics of malaria, a vector-borne disease. In that case, the force of infection is a function of parameters related to the mosquito population dynamics and the vector control strategies. Extra compartments are also included both to take into account asymptomatic and undetectable individuals, but also to account for treated individuals.

2

Wrangling epidemiological data

" Our key message is: test, test, test "
(Tedros Adhanom Ghebreyesus
- World Health Organisation Director-General, 2020)

The word *epidemiology* comes from the Greek *epi* (upon), *dēmos* (people) and *logos* (the study of). It literally means the study of what befalls upon a population. In the public health context, it can be defined as "the study of the distribution and determinants of health-related states or events in specified populations, and the application of this study to control of health problems" [94].

Both in epidemiological modelling and epidemiology as a whole, we focus on studying health-related conditions at population or society-level, as opposed to individual-level. Being a quantitative discipline, epidemiology is concerned with finding determinants and patterns of health events in a population using systematic and data-driven approaches [36]. Hence, it is heavily dependent on data availability, as facilitated by public health surveillance systems. Surveillance should be timely, representative, sensitive and specific. This means that case definitions should be clear, information should be accurate and notification responses should be fast and frequent.

Notifiable diseases are conditions which reporting is required by law. For instance, in Brazil, suspected cases of dengue and malaria in the amazon region should be compulsorily reported on a weekly bases, whereas severe acute respiratory infection cases, including SARS-CoV-2 cases, should be reported immediately (within 24h) [105]. Nevertheless, even for such notifiable diseases we might encounter challenges with data analysis and acquisition at population level.

Access to information is a fundamental right for Brazilian citizens [21] and all information produced by the Brazilian State should be public [20], but most

epidemiological databases are classified under *passive transparency*, requiring filling of a formal request through the *Serviço de Informação ao Cidadão (eSIC)* [23]. This represents a first challenge in the context of timely data acquisition, since each request takes up to two months to be processed and responded to. Also, requests need to be very specific, mentioning the data structure, variables and granularity you expect. Hence, previous knowledge of the database structure is key.

Fortunately, during the the COVID-19 pandemic the *OpenDataSUS* [22] and *LocalizaSUS* [24] platforms were created by the Brazilian Ministry of Health as a response to academia and society's pressure towards data openness. Such databases comprehend information on coronavirus disease 2019 (COVID-19) and severe acute respiratory infection (SARI), funding, infrastructure, hospital bed availability and vaccination coverage [31]. Nevertheless, even in a scenario where there is access to daily or weekly updated databases, reporting can still suffer from delays due to the time lag between seeking treatment, testing, filling forms and reporting.

Another challenge is the quality and relevance of data. For instance, regarding quality, *active case detection* is not a rule, even for notifiable diseases such as the *coronavirus disease 2019* (COVID-19) [28]. This means that reported cases only account for clinically ill individuals that sought treatment through registered health services, got tested and the test result was positive. Clearly, that only reflects a portion of the actual infectious cases in the population. Regarding relevance of the available data, we could mention the subjectivity involved in different possible case definitions; some of them being very useful for surveillance, but not so much as input for compartmental models.

In the next few sessions, we will discuss how to deal with these challenges while preparing epidemiological data to be used in a mathematical modelling context. Even though many times this part of the process is left-out for not being as academically thought-provoking, it is an extremely important and strenuous part of the modelling work, sometimes involving political issues and conflicts of interest.

2.1 Dealing with delayed reporting of notifiable diseases

During an epidemic, the total case count generated by a health surveillance system *in the present* underestimates the real number of infectious cases in a population due to reporting delays and under-reporting. These two underestimation sources have different causes and, hence, there are different ways to deal with them.

Reporting delays comprehend several stages, from seeking treatment, collecting samples, processing a test, returning a positive result to reporting and uploading a case into the relevant database. For instance, during the COVID-19 pandemic in Brazil, the time lag comprising all these stages could add up to a several days

between symptom onset and a confirmed case being effectively registered in the *Sistema de Informação de Vigilância Epidemiológica da Gripe (SIVEP-Gripe)* database [26; 127].

To correct for reporting delays, we can use techniques such as the Beyesian *nowcasting* proposed by McGough et al. [103], which essentially estimates more accurate *real-time* case counts of a disease by bringing forward future official notifications by the mean time between symptom onset and hospitalisation. Obtaining *"nowcasted" time-series data* is not only useful in terms of providing a more accurate snapshot of the current epidemic situation, but it is also essential to allow relevant and timely predictions (*forecasting*) for the disease's incidence and mortality.

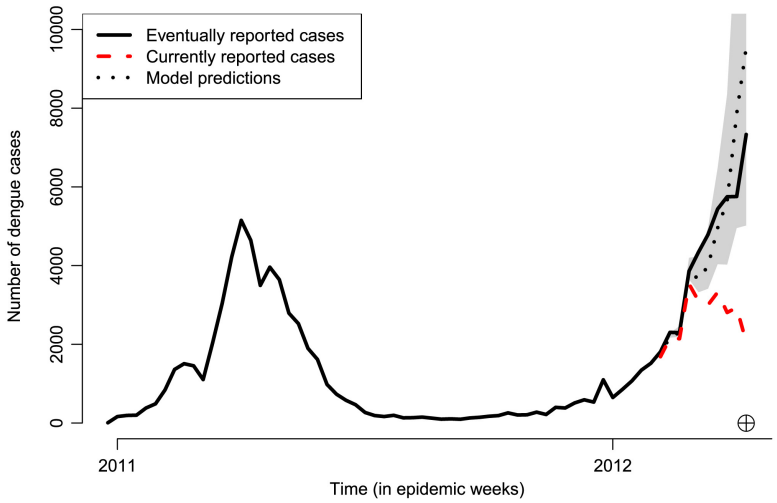


Figure 2.1: Example of a system where a *nowcasting* algorithm was used to obtain timely case estimates (black doted line, with 95% C.I in grey), improving the accuracy of the case count before the database was consolidated. Weekly time-series of initially reported (red dashed line) and eventually reported (solid black line) dengue cases in Rio de Janeiro, Brazil, between January 2011 and April 2012. (Source: Bastos et al. [14])

For the case of under-reporting, the issue consists in part of the positive cases not being registered due to insufficient testing kits, difficult access to health systems, limited front-line health staff, lack of treatment seeking in case of mild symptoms or even planned imperfect detection.

The solution for the under-reporting problem, at a public health policy level, would be the implementation of extensive and accessible testing; not only to confirm mild and severe clinical cases, but also to actively detect asymptomatic cases. If that is not a viable option, corrected figures could be obtained if *reporting rates* estimates are available.

2.2 Processing epidemiological data sets

Given that one has access to a relevant epidemiological data set, a number of processing steps need to be followed so that such data is presented in a format that can be used for analysis and model fitting. In this context, it is usual to represent incidence data as aggregated daily, weekly our yearly time-series. To be able do that, we should be mindful about case definitions; relevant dates for case data aggregation; epidemiological week/year definition for the relevant setting or health system.

Deciding the best way to manipulate data will be informed both by its intended used and information found in data dictionaries.

Case definitions

In a anonymised case-by-case epidemiological database, different variables might be found defining varied case outcome or classification. For instance, a single database might comprise both suspected, untested and confirmed cases within a disease; both mild and severe (hospitalised) cases of a syndrome; different strains of one disease or even different diseases. When aggregating time-series for use in epidemiological data analysis, we need to be sure to be sub-setting the database for the clinical outcome of interest (whether it is an specific strain, the sub-set of clinically severe cases or only test-positive confirmed cases).

Date for case data aggregation

In such databases, one can also encounter different recorded dates referring to each case and it can confusing to select which should be used when aggregating cases as a time-series. We might usually have dates for symptom onset, testing, treatment seeking and filling reported. The best practice is to aggregate cases by the date of symptom onset (when such information can be attained), as using this procedure we will not be accumulating systematic errors and delays.

Epidemiological week definitions

An epidemiological week is an standardised way to count weeks year after year, for the purpose of epidemiological surveillance. The division of the 365 days of the year in 52 or 53 epidemiological weeks (*epiweeks*) generates a calendar that is used as a basis to group cases, deaths or other epidemiological events.

The World Health Organization (WHO) uses the following definition:
"epidemiological weeks start on a Sunday and end on a Saturday; the first epidemiological week of the year ends, by definition, on the first Saturday of January,

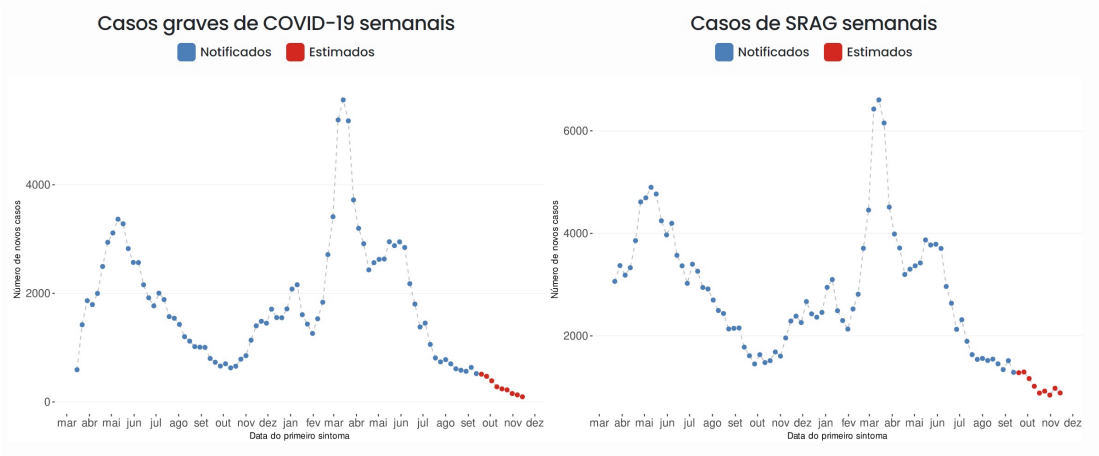


Figure 2.2: Historical time-series for SARI (*SRAG*) and COVID-19 cases in São Paulo city, as presented in the *Observatório COVID-19 BR* website for the SIVEP-Gripe database from 15/11/2021. The blue points represent reported cases, whereas red points represent case estimates, as predicted by a now-casting algorithm [103]. The Observatório COVID-19 BR team maintains these plots updated as new versions of the SIVEP-Gripe database become available. More details on the Observatório COVID-19 BR collaboration can be found in chapter 3. (Source: Observatório COVID-19 BR [3])

as long as it falls at least four days into the month, even if it means that this first week starts in December" [157].

For the International Organization for Standardization (ISO), in the other hand, *all weeks in the ISO-8601 week-based calendar start on a Monday; "the first calendar week of a calendar year includes up to three days from the previous calendar year; the last calendar week of a calendar year includes up to three days from the following calendar year"* [84].

Each country's Ministry of Health can have its own definition for the first epidemiological day of the week or the first epidemiological day of the year, hence it is worth checking which standard is the most relevant in each setting.

2.3 SARS-CoV-2 data

The severe SARS-CoV-2 case data used to calibrate the models described in part II of this thesis comes from the *SIVEP-Gripe* database [26; 25; 127], which comprises all severe acute respiratory infection (SARI) cases admitted in hospitals and reported in the country. SARI is a notifiable clinical presentation caused by varied etiological agents, including viruses (influenza viruses, adenovirus, hantavirus and coronaviruses) and bacteria (pneumococci, leptospira, etc).

The Brazilian Ministry of Health considers as a SARI case any presentation including fever over 38°C, cough and dyspnea; with or without a sore throat and gastrointestinal manifestations. The COVID-19 presentation can include some of

these symptoms, hence severe hospitalised COVID-19 cases and deaths are reported as SARI cases [31; 105].

When an antigen or any molecular biological test is performed on a SARS-CoV-2 hospitalised patient, the test result information is stored in the SIVEP-Gripe database. Nevertheless, tests were scarce and SARS-CoV-2 testing protocols were inconsistent throughout the first few months of the pandemic in Brazil, with not all severe/hospitalised COVID-19 cases being tested [87]. Hence the reported numbers for *confirmed* COVID-19 hospitalised cases and deaths could not be considered so representative in Brazil. Complementary, the *eSUS-VE* database [27] comprises data from localised contact tracing efforts and reported COVID-19 mild cases, which though still technically notifiable [28], are far less representative as there was even less testing.

As we consistently monitored databases for both SARS and confirmed COVID-19 cases (Figure 2.2), it was clear that SARS-CoV-2 severe cases comprised the great majority of SARI notifications since the beginning of 2020, hence we decided to fit our dynamic models to the totality of SARI cases and deaths. The data used consisted on time series of hospitalisations and deaths of SARI in São Paulo, Brazil, reported in the 2020 SIVEP-Gripe database [22] and openly made available in the *Observatório COVID-19 BR* website repository [113].

2.4 Malaria data

Malaria is a notifiable disease with compulsory case reporting and data made available under passive transparency, meaning the submission of a formal request through the *Serviço de Informação ao Cidadão (eSIC)* [23] is required to access the relevant databases.

Throughout the time, different notification systems and databases were used to report malaria cases in Brazil. Up to 1995, all reported cases were registered in the *Sistema de Informação Série Histórica de Malária (SHM)* [134]. After 1996, since the majority of cases occurred in the Amazon region (see figure 2.3), reporting was divided between the Brazilian Amazon (AC, AM, AP, MA, MT, PA, RO, RR and TO states) and extra-Amazon (AL, BA, CE, DF, ES, GO, MG, MS, PB, PE, PI, PR, RJ, RN, RS, SC, SE and SP states) regions.

Malaria cases reported in the Amazon region, between 1996 and 2002, were registered in the *Sistema de Informação do Programa Nacional de Controle da Malária (SISMAL)* and, from 2003 on, in the *Sistema de Informação da Vigilância Epidemiológica - Malária (Sivep-Malária).* In the extra-Amazon region, cases reported between 1996 and 2003 were registered in *SISMAL* and, from 2004 on, they were registered in the *Sistema de Informação de Agravos de Notificação (SINAN)* [29].

For the modelling exercises presented in part III, individual case notification data

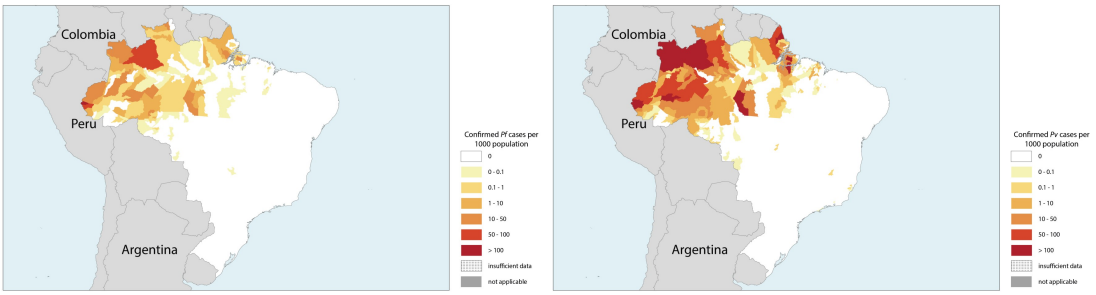


Figure 2.3: Spatial distribution of *P. falciparum* (left) and *P. vivax* (right) confirmed cases in Brazil in 2017. (Source: WHO Global Health Observatory (GHO) data [155])

was obtained from the SISMAL, SINAN and SIVEP databases, comprising the period between 1998 and 2018, being processed and aggregated as specie-specific daily and weekly time-series by municipality and state. The outputted time-series were made available in a online data repository [68].

2.5 Conclusions

Each epidemiological data set comes with its own particularities and challenges. Different diseases have different testing and reporting protocols, depending on the available knowledge and case definitions. Likewise, different states, municipalities or health districts might have access different resource levels and follow their own procedures, in a non homogeneous sensitivities to a disease's real incidence [42]. These aspects can introduce not only uncertainties, but spatial heterogeneity and biases to a data set.

Part II

Modelling an emergent pandemic disease

3 The SARS-CoV-2 pandemic

> ***PANDEMIC*** *An epidemic occurring worldwide, or over a very wide area, crossing international boundaries, and usually affecting a large number of people.*
>
> (John M. Last, *A Dictionary of Epidemiology*, 2001, p. 131)

On the 31st of December 2019 the World Health Organization (WHO) China Country Office registered cases of a pneumonia of unknown etiology detected in Wuhan City, Hubei Province. Patients were reported presenting fever, with a few of them having difficulty in breathing, and chest radiographies showing invasive lesions of both lungs[161]. From that date, cases rose continuously without having their causal agent identified, until the first occurrences were linked with exposures in one seafood marked in Wuhan City. At that time, a new type of coronavirus was isolated, the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [148], causing the disease named coronavirus disease (COVID-19) [170]. On the 11th of January 2020 the novel coronavirus was sequenced, on the same day the first related death was reported[160]. A few days later, exported cases started being detected in other Chinese provinces, as well as in Thailand, Japan and the Republic of Korea, leading to the implementation of screening and track-and-trace measures. On the 30th of January 2020, with a total of 98 cases being reported in over 18 countries outside China, the WHO's Director-General declared the novel coronavirus outbreak a public health emergency of international concern (PHEIC), WHO's highest level of alarm.

Though reverse real-time reverse transcriptase-polymerase chain reaction (RT-PCR) tests were rapidly developed [148; 43; 45], in a matter of 3 months the coronavirus disease (COVID-19) outbreaks were being reported all around the world, affecting people of all ages, with alarming levels of spread and severity. WHO's worldwide recommendation was for countries to immediately activate the highest levels of

national response protocols taking a whole-of-government, whole-of-society approach.

On Wednesday the 11th of March the World Health Organization declared COVID-19 a global *pandemic*, as the novel coronavirus had already been detected in 113 countries and territories outside China, with 118319 cases and 4292 deaths confirmed globally [162; 94]. At the time, there were no pharmaceutical interventions commercially available to treat the disease, prevent infection or curb the epidemic. Testing rates were poor in many countries [159; 87; 85] across the globe and, as control strategies, most countries relied on behaviour change and non-pharmaceutical interventions (NPI) including self-isolation of symptomatic individuals; increased hand hygiene; mask wearing; social distancing; working from home where possible; and school closure amongst others. Containment efforts primarily aimed at preventing infection and reducing onward transmission. However, most countries were already experiencing locally acquired cases with untraceable contacts, meaning that community transmission had been established. Local authorities were hence mobilised towards mitigation strategies with the aim of delaying the peak and flattening the epidemic curve to reduce the daily demand on health care resources, thus preserving hospital capacity.

While borders where closed, schools were shut down and families were separated, virus mutations arose as new variants of concern (VOC) started circulating, with higher transmissibility or severity [53; 55; 54; 168], making stronger surveillance and the scientific study of the VOC also priorities. Meanwhile, a global race towards developing the first safe and effective vaccine or treatment was already going on and on the 31st of December 2020 the first emergency use validation for a COVID-19 vaccine was released [120; 110; 164]. Other vaccines soon started to be validated and produced in large scale to be distributed around the globe [9; 163; 147; 165; 130; 166], while large collaborations accelerate its deployment and accessibility [111; 167].

As of the 24th August 2021, over 211 million cases and 4.4 million deaths had been reported globally over 192 countries [169], still with millions of new cases being reported each day. Four VOC had been circulating (Alpha, Beta, Gamma and Delta) without much general consensus on the extent to which these mutations affect disease severity and vaccine effectiveness [44]. Though a wide-reaching decrease in disease morbidity and mortality was expected after mass vaccination efforts and naturally acquired immunity, vaccine hesitancy and their uncertain effect against new VOC are ongoing challenges. Even as higher percentages of the population become immunised, COVID-19 related cases and deaths are still a major issue worldwide (see the mortality curve in Fig. 3.1).

Furthermore, by the end of 2021, as this thesis is written, equitable access to vaccination is still a challenge and outbreaks are still being registered worldwide, either . In such scenario, granted plenty of access to treatment and prophylaxis, we can start to glimpse a future where life finally resembles the longed "normality".

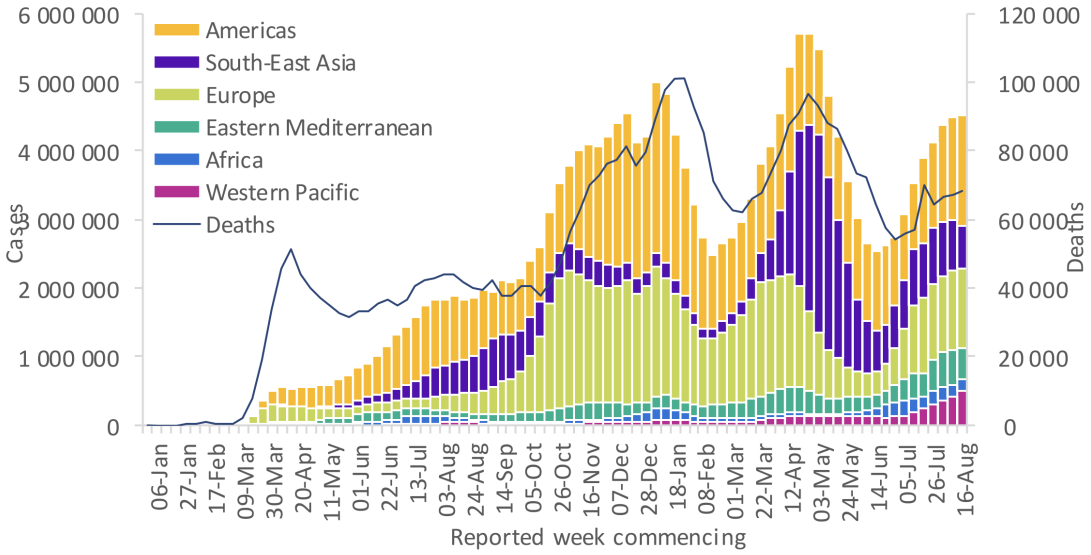


Figure 3.1: COVID-19 cases reported weekly by WHO Region and global deaths, as of 22 August 2021. (Source: World Health Organization [169])

3.1 Mathematical modelling during the COVID-19 pandemic

"The more certain someone is about covid-19, the less you should trust them", as written by George Davey Smith, Michael Blastland and Marcus Munafo in October 2019 in a BMJ editorial [47]. Since the emergence of SARS-CoV-2, mathematical modelling has become an important tool to support strategic planning towards combating the spread of the disease and making decisions at public health level [4; 5; 11; 116]. In this context, epidemiological mathematical models have played an increasingly prominent role advising policymakers [108; 72; 4; 5; 11; 116; 49; 65; 67; 124; 104; 129; 138; 146; 13; 40; 61; 112]. A proliferation of SEIR-like compartmental models assessing the effect of different intervention scenarios have been proposed [49; 65; 67], aiming to understand and make predictions regarding spread of SARS-CoV-2. As simplified representations of reality, these models come with assumptions regarding the nature of the underlying network of interactions [124]. Each with its own strengths, they were more or less fit to address different questions, depending on how they were designed.

As the first vaccines against COVID-19 started to be developed, similar models have also been used to address questions regarding vaccine deployment strategies and efficacy at populational level, specially because most vaccines require two doses to reach peak efficacy [147; 120]. Using mathematical models to assess optimal vaccination strategies in a context of limited supply was not a new idea, though. Previously to SARS-CoV-2 emergence, this was done for a variety of communicable

diseases [12; 56; 57; 75; 171]. For instance, a study on influenza viruses, Matrajt et al. [101] had already explored how two different vaccination strategies (single dose or two dose) could be integrated into pandemic control plans. The best strategy is shown to be dependent on the level of partial protection introduced by each dose.

In the SARS-CoV-2 context, mass vaccination deployments have been the main approach tackling the spread of the disease, whilst complementing or replacing non-pharmaceutical interventions (NPIs). SIR-like models have pointed to the possibility of securely relaxing NPIs months after vaccination campaigns, depending on the rate of vaccination [91] but investigations regarding how to better distribute the doses within a population were inherent.

3.2 Observatório COVID-19 BR and COVID-19 Modelling (CoMo) Consortium contributions

The SARS-CoV-2 pandemic has not only largely overwhelmed health systems, but it has also created tremendous challenges for decision-makers, who have had to implement highly disruptive containment measures with very little empirical scientific evidence to support their decision-making process. Given this lack of data [47; 46], predictive mathematical models have played an increasingly prominent role [5; 173; 72; 4; 11; 116; 49; 65; 67; 112], even in low- and middle-income countries, such as Brazil [5; 100; 44].

The need to inform policy making brought scientists to work together in broad collaborative initiatives across the globe, increasing the pace at which such evidence was published and communicated.

Two successful scientific collaborations that acted primarily on using mathematical models towards informing public health decision making at national and international level were *Observatório COVID-19 BR* [2] and *COVID-19 Modelling (CoMo) Consortium* [1], respectively.

The Observatório COVID-19 BR started as an independent initiative between our modelling group in the Institute of Theoretical Physics and close collaborators, being founded in March 2021 by Professors Roberto Kraenkel (UNESP), Paulo Inácio Prado (USP) e Renato Coutinho (UFABC) It then rapidly caught the attention of the media [99; 39] and other scientists, rapidly growing into a collaboration consisting of over 80 voluntary researchers throughout Brazil and abroad, aiming to contribute to the dissemination of reliable information based on detailed data and scientifically grounded analysis, both to the scientific community and the wider population [70; 60; 59]. Simulating the outcomes of different NPI strategies scenarios, summed with several other modelling and data analysis initiatives, we were able to communicate with policymakers such as Health Secretaries or Funerary Services in the state of São Paulo, being able to contribute in the decision making debate

towards mitigating the spread of the disease. During this process we connected and worked unprecedentedly closely with fellow modellers throughout Brazil and abroad.

Part of our group has also joined forces with the COVID-19 Modelling (CoMo) Consortium, an international group of infectious disease modellers and public health experts funded by Prof. Lisa J. White (University of Oxford) that collaborated on the creation of a modelling interface to help simulating the effect of different non-pharmaceutical interventions on mitigating the epidemic [5; 173]. This modelling interface has been used to support decision making in numerous countries and it is continuously developing, whereas new information regarding the SARS-CoV-2 dynamics is unveiled, new pharmaceutical interventions are developed and new questions are raised.

In Chapters 4–5, we will present some of the modelling work we developed in collaboration with Observatório COVID-19 BR and CoMo Consortium, to support decision making and understanding SARS-CoV-2 spreading mechanisms. We have developed different models to address different questions within different epidemiological settings.

We introduce, in Chapter 4, the modelling framework resulting from CoMo's participatory approach to COVID-19 dynamics modelling, which aimed to circumvent the modelling translational capacity gap between high and middle income countries by developing an interactive modelling interface and regularly meeting with in-country policy makers and health experts. The basis of such framework has already been published in Aguas et al. [5] but it is continuously developing, according to the needs of policy makers and as new interventions become pertinent. For instance, it has been used to explore the outcomes of using dexamethazone to reduce mortality in COVID-19 hospitalised patients in another publication by Águas et al. [173] and consortium members. We then built on CoMo's original age-structured SEIR extended model as we adapted it to the Brazilian hospital compartmentalisation system and proposed a correction to the household contact matrices, using network percolation theory. This theoretical part of the work has been published in Franco et al. [69].

In Chapter 5, we present some applications of the improved CoMo model and the results of other modelling exercises using slightly different frameworks. The modelling framework described in Chapter 4 is applied to investigate school reopening scenarios in major urban centers in Brazil, their consequences and how a proper contact tracing strategy could be used to mitigate transmission in the school setting. This modelling work is resultant from a collaboration between the Observatório COVID-19 BR modelling group and the CNPQ collaborative project *Modelling SARS-CoV-2 transmission dynamics in Brazil: Real time science to subside evidence based policy making* , coordinated by Prof. Cristiana Toscano (UFG). The results from this collaboration have been documented as a manuscript which is

currently accessible as a preprint in Borges et al. [19].

Loooking from a different and perhaps more theoretical perspective, we also developed a simpler model to investigate optimal vaccination strategies in a limited supply setting, such as Brazil. We use a SEIR model that includes a two-dose vaccination schedule with a between-doses delay modelled through delay differential equations and linear optimisation of vaccination rates, finding that the best strategy depends on an trade-off between the vaccine production rate and the relative efficacy of the first dose. This work resulted from a collaboration within some of Observatório COVID-19 BR member, having been submitted for peer-reviewing and being currently accessible as a preprint in Souto Ferreira et al. [137].

Finally, since in November 2020 the P.1 SARS-CoV-2 variant of concern emerged in the city of Manaus leading to a collapse of the local health system in early 2021, we used another SEIR model to evaluate the transmissibility and reinfection rates of this variant. In in Chapter 5 we show that, from our estimates, the P.1 VOC posed a global threat due to its very high transmissibility, highlighting the need to urgently monitor and contain its spread. These results have been published in Coutinho et al. [44].

We conclude discussing the impact of our results and the challenges faced when trying to communicate them to policymakers.

4 A COVID-19 pandemic modelling framework

The CoMo Consortium SARS-CoV-2 model builds on the SEIR model structure, as presented in chapter 1. In the early stages of the pandemic, it became clear that older individuals were disproportionately affected by severe disease and mortality [62; 37], hence, introducing age-structuring in our modelling framework was essential. Furthermore, with the aim to take into account the demand for hospital resources, the preliminary *infectious* compartment was sub-divided into seven infected compartments characterised by symptoms, severity and treatment seeking. The progression of individuals through the infection life cycle is represented by the diagram in Figure 4.1.

As individuals are infected, they transit through an incubation phase (or exposed state, E), at the end of which they can display different symptomatology, with some

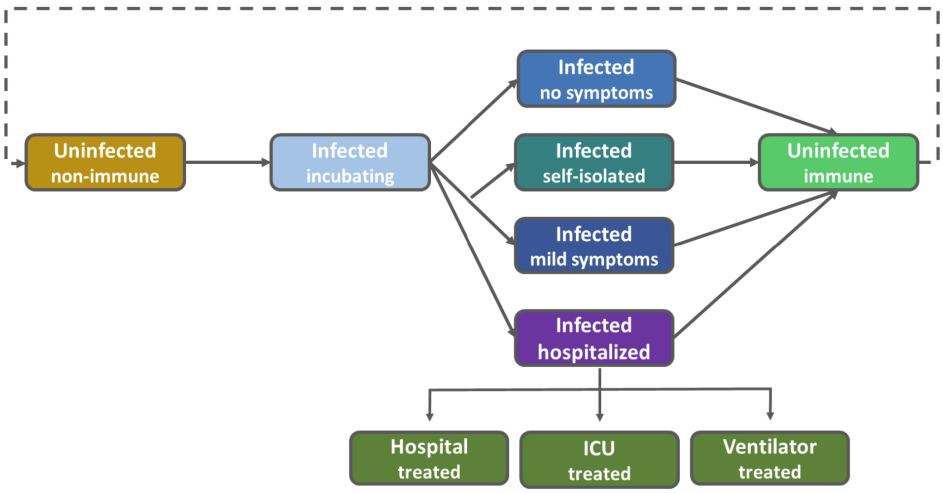


Figure 4.1: A diagram of the baseline CoMo model structure representing the unmitigated epidemic spread scenario. (Source: Aguas et al. [5])

never having any symptoms, while others require hospitalisation. We attribute different flow rates between the exposed state and each of the possible infectious states based on incidence data and previous clinical and epidemiological studies [96; 90; 18; 106; 50; 143; 145; 62; 149; 118; 95; 126].

If a person is hospitalised, the model tracks their health care requirements (surge care bed, ICU bed, ventilator). Disease-induced mortality rates are heavily dependent on how severe the infection outcome is and whether individuals can receive the appropriate treatment. The model assumes that those who recover from infection will become immune.

Since early studies, it has been shown that COVID-19 severity and mortality is strongly age dependent [145]. Hence, each model compartment is further stratified by age classes, corresponding to each country's demographic classifications. Transmission between each pair of age classes is evaluated using contact matrices estimated for different settings (school, work, home, and others) by Prem et al. [122]. Empirical social contact matrices data has always been scarce, with the exception is the POLYMOD study, a extensive survey that described social mixing patterns in eight European countries during 2005/2006, measuring contacts of more than 7000 participants across eight European countries [107]. Prem et al. [122] used this data combined with health surveys and demographic data as input to a Bayesian hierarchical model to project contact matrices for 144 other countries, including Brazil.

The resulting CoMo model interface can be found in https://comomodel.net/ [1], where publicly available country-specific data on cases and mortality for COVID-19 is used for visual calibration of model parameters to user-selected non-pharmaceutical interventions scenarios. These scenarios include the implementation of combinations of NPI that modify the contact matrices proportionally to their respective coverage and efficacies. The initially implemented interventions consisted of self-isolation, screening, social distancing, handwashing and mask wearing, working from home, school closures, shielding the elderly, travel ban, home quarantining and lockdown, and an extensive description of each of them can be found in the Supplementary Material of Aguas et al. [5]. Since the publication of this work, more interventions have been implemented into the modelling interface, following a development workflow informed by the needs of policy-makers and newly unfolded information about the disease spread dynamics (Figure 4.2).

A detailed description of the mathematical structure underlying this modelling framework will be featured in the following sections, where we thoroughly describe two nested versions of this model adapted for modelling the SARS-CoV-2 pandemic spread in Brazil, developed in a collaboration between CoMo Consortium and Observatório COVID-19 BR. The standard version closely follows the CoMo original model introduced in Aguas et al. [5], with a few modifications in the infected

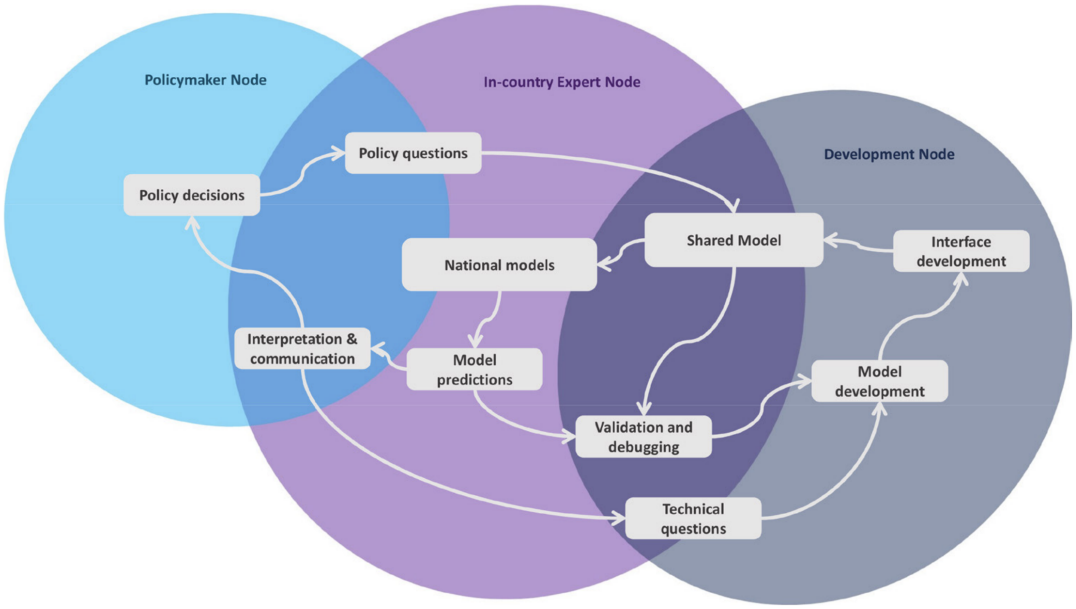


Figure 4.2: CoMo Consortium outlook and interaction flows. This diagram illustrates how the different partners interact in order to digest policy questions into model simulations through the in-country expert node and the development node and ultimately result in actionable predictions informing policy decisions. CoMo, COVID-19 Modelling. (Source: Aguas et al. [5])

hospitalized compartments accounting for the Brazilian health system particularities. In a second version (model with percolation), we implement a correction that modulates the home contact matrices through a percolation correction function, motivated by the fact that the mitigation strategies for COVID-19 affect the connections between different households, which is not explicitly taken into account in the standard model version.
Finally, we show that our model with percolation effects was better supported by the data than the same model without such effects. By allowing a more reliable assessment of the impact of NPIs, our improved model provides a better description of the epidemiological dynamics.

4.1 Bridging the gap between population and household-level models

As introduced in the previous chapter, mathematical modelling has become an important tool for strategic planning since the emergence of COVID-19 outbreaks around the world. Hence, many SEIR-like compartmental models started to be developed, like CoMo model, described above. As simplified representations of reality, all models come with assumptions regarding the nature of the underlying network of interactions [124]. More explicitly, most SEIRS compartmental models

assume that the population is homogeneously-mixed, meaning that every individual in each compartment has the same probability of contacting others, regardless of spatial distribution.

Though these underlying assumptions might be appropriate to make robust predictions for well-mixed populations, they may not be suitable in a context with significant contact network heterogeneity. Therefore, several methods have been proposed to account for contact structure heterogeneities in compartmental models [104; 129; 138]. Nevertheless, they usually consider static underlying network structures, which are only good approximations for slow epidemic spreads, in which contact rates do not change significantly over time [146; 13]. This is not the case in a pandemic such as the one caused by SARS-CoV-2, where governments were compelled to impose wide contact and movement restrictions that essentially caused a broad reduction of connectivity between individuals. The need for a more flexible and dynamic approach became evident, hence motivating the present study.

In this context, network theory provides a picture of a homogeneously mixed population as a (highly connected) regular random network of individuals (vertices) connected through possible disease transmission contacts (edges) with long-range connectivity properties [13]. During an outbreak, the disease would spread across these links. Social distancing non-pharmaceutical interventions (NPIs) could, therefore, be thought of as modulators of the strength and even persistence of such links. One of the main results of network theory is that, as contact networks become more connected, a critical transition occurs, and a once fragmented and disconnected (or weakly connected) set sub-networks becomes a highly connected component. This phenomenon is known as percolation [40; 61], and the existence of a critical percolation threshold in the mean number of contacts is established for many kinds of networks. Across this threshold, even small changes in the number of contacts can lead to large changes in the connectivity of the network.

In compartmental models, SARS-CoV-2 mitigation strategies are modelled by altering contact rates, thus changing the force of infection. This can be done at different degrees of granularity, depending on the level of detail of contact matrices. For instance, many works [5; 112; 48] have used the categorisation employed by Prem et al. [122], which divides contacts into four settings, namely home, work, school, and other. In this way, the effectiveness of different NPIs are reflected in reductions in the contact rates for each setting, depending on the nature of the intervention – e.g. school closure reduces contact rates at school. Since the contact matrix is a linear combination of the contact matrices for each setting, with coefficients dependent on the NPIs' coverages and efficacies, each NPI contributes linearly to reduce the infection force. By itself, this change in contacts among individuals does not affect the relationship between the mean number of contacts in the underlying network and the force of infection of the compartmental model. This is adequate if the structure of the network is not greatly affected. However, when social distancing NPIs are applied with high coverages and connections between

individuals are continuously being removed, this approximation is prone to break down, and the compartmental model may no longer provide a satisfactory portrayal of the epidemiological dynamics.

To account for this change in network structure and resultant reduction in the force of infection, here we propose to modify how the effect of NPIs are modelled in compartmental models. We integrate results from percolation theory into an age-structured SEIRS model by using a non-linear correction between the strength of NPIs and the resulting contact matrix, thus directly affecting the force of infection. In the following sections, we build on the previously implemented and widely used COVID-19 Modelling Consortium (CoMo) model [5], adapting its compartmentalisation to the Brazilian hospital system. Then, we present a nested model version that takes into account the loss of long-range connectivity (percolation) effect. Finally, we fit both model versions to hospitalisation and mortality data for SARS-CoV-2 during 2020 in São Paulo, Brazil, the country's most populous city, with over 12 million inhabitants, and the first to detect COVID-19 cases. Through model comparison using the Akaike Information Criteria (AIC) [34] we find that the data strongly supports the model incorporating percolation effect.

4.2 Standard model structure

To model the epidemiological dynamics of COVID-19 in São Paulo, we apply an age-structured SEIRS model with infected compartments stratified by severity of symptoms, treatment requirement and accessibility to healthcare. The main framework for this SARS-CoV-2 epidemic model was developed in collaboration with the CoMo Consortium [5] with slightly different treatment seeking compartments, an adaptation applied to better represent the Brazilian organisation of hospital beds. The progression of individuals through the infection cycle, for this version of the CoMo model, is represented by the diagram in Figure 8.1. Each model compartment depicted in Figure 8.1 is divided into 19 sub-compartments, comprising all 19 age classifications from the Brazilian Institute of Geography and Statistics (IBGE) [81]. Transmission between each pair of age classes is therefore evaluated based on the contact rates between them, which are the contact matrices estimated for different settings (school, work, home, and others) in Brazil by Prem et al. [122]. Let us note here that this is already a limitation for this class of models, as these matrices consist of country level projections, as will be discussed in section 4.6

We considered non-pharmaceutical intervention (NPI) scenarios including social distancing, work from home, and school closure. Here, the effect of mitigation strategies enter as linear corrections to the contact matrices, since these interventions aim to reduce contact rates or the risk of contagious at each possible contact (mask use, for example). For instance, a certain coverage on the school closure NPI would decrease contacts between all age classes in the school setting,

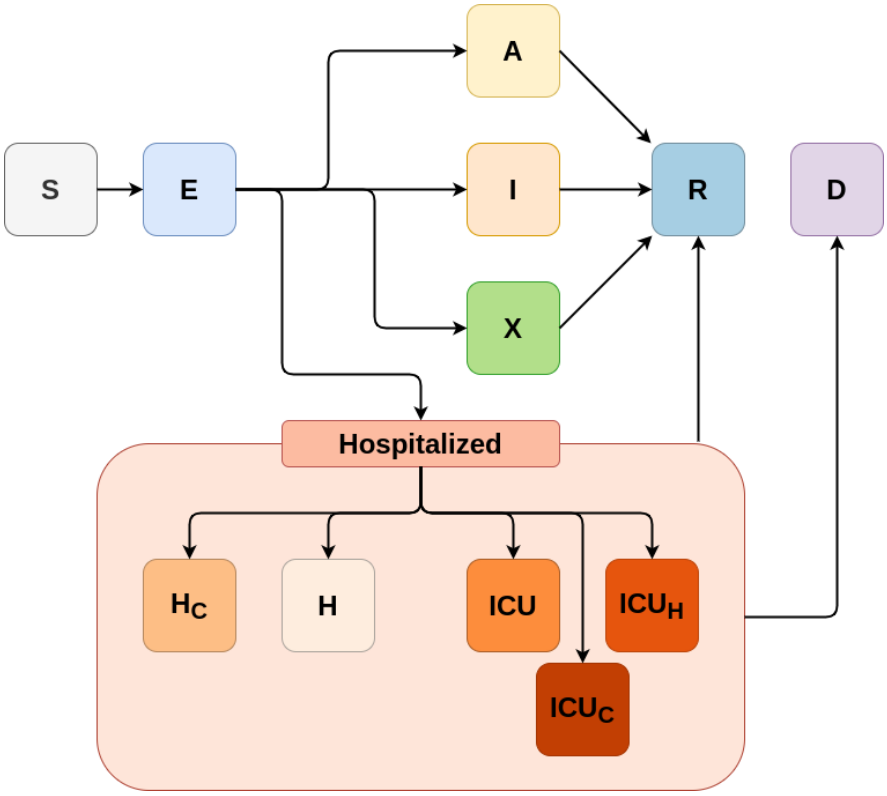


Figure 4.3: A diagram of the baseline model structure for SARS-CoV-2 in Brazil, representing the unmitigated epidemic spread scenario. The variables in the compartments represent individuals S: susceptible, E: presymptomatic infectious, A: asymptomatic infectious, I: infectious with mild symptoms, X: infectious with mild symptoms and self-isolated, H_C: infectious, requiring hospital treatment but denied, H: infectious and hospitalised, ICU: infectious, receiving intensive care, ICU_C: infectious, requiring intensive care but denied, ICU_H: infectious, requiring intensive care but being admitted in regular hospital bed, R: Recovered, D: Deceased. All compartments are further subdivided into 5-year age classes from 0 to 90+ years old. (Source: Franco et al. [69])

which is modelled as a proportional decrease in contact rates in school contact matrices.

4.2.1 Model equations

The model consists in an expanded SEIRS model to account for asymptomatic individuals and detailed structure of the Brazilian health system. We write such epidemiological dynamics as the following system of 11 equations:

$$\frac{d\mathbf{S}}{dt} = -\lambda\mathbf{S} + \omega\mathbf{R} + \bar{\mathbf{A}}\mathbf{g}\cdot\mathbf{S} + \mu_b - \mu_d\mathbf{S} \tag{4.1}$$

$$\frac{d\mathbf{E}}{dt} = \lambda\mathbf{S} - \gamma\mathbf{E} + \bar{\mathbf{A}}\mathbf{g}\cdot\mathbf{E} - \mu_d\mathbf{E} \tag{4.2}$$

$$\frac{d\mathbf{A}}{dt} = \gamma(1-P_{clin})(1-IHR)\mathbf{E} - \nu_i\mathbf{A} + \bar{\mathbf{A}}\mathbf{g}\cdot\mathbf{A} - \mu_d\mathbf{A} \tag{4.3}$$

$$\frac{d\mathbf{I}}{dt} = \gamma P_{clin}(1-P_{selfis})(1-IHR)\mathbf{E} - \nu_i\mathbf{I} + \bar{\mathbf{A}}\mathbf{g}\cdot\mathbf{I} - \mu_d\mathbf{I} \tag{4.4}$$

$$\frac{d\mathbf{X}}{dt} = \gamma P_{selfis}P_{clin}(1-IHR)\mathbf{E} - \nu_i\mathbf{X} + \bar{\mathbf{A}}\mathbf{g}\cdot\mathbf{X} - \mu_d\mathbf{X} \tag{4.5}$$

$$\frac{d\mathbf{H}}{dt} = \gamma IHR(1-P_{icu})(1-H_c)\mathbf{E} - \nu_s\mathbf{H} + \bar{\mathbf{A}}\mathbf{g}\cdot\mathbf{H} - \mu_d\mathbf{H} \tag{4.6}$$

$$\frac{d\mathbf{HC}}{dt} = \gamma IHR(1-P_{icu})H_c\mathbf{E} - \nu_{sc}\mathbf{HC} + \bar{\mathbf{A}}\mathbf{g}\cdot\mathbf{HC} - \mu_d\mathbf{HC}$$

$$\frac{d\mathbf{ICU}}{dt} = \gamma IHRP_{icu}(1-ICU_c)\mathbf{E} - \nu_{icu}\mathbf{ICU} + \bar{\mathbf{A}}\mathbf{g}\cdot\mathbf{ICU} - \mu_d\mathbf{ICU} \tag{4.7}$$

$$\frac{d\mathbf{ICUH}}{dt} = \gamma IHR\ P_{icu}ICU_c(1-ICU\ H_c)\mathbf{E} - \nu_{icuh}\mathbf{ICUH} + \bar{\mathbf{A}}\mathbf{g}\cdot\mathbf{ICUH} - \mu_d\mathbf{ICUH} \tag{4.8}$$

$$\frac{d\mathbf{ICUC}}{dt} = \gamma IHR\ P_{icu}ICU_cICU\ H_c\mathbf{E} - \nu_{icuc}\mathbf{ICUC} + \bar{\mathbf{A}}\mathbf{g}\cdot\mathbf{ICUC} - \mu_d\mathbf{ICUC} \tag{4.9}$$

$$\begin{aligned}\frac{d\mathbf{R}}{dt} = {} & \nu_i\mathbf{A} - \omega\mathbf{R} + \nu_i\mathbf{X} + \nu_i\mathbf{I} + \bar{\mathbf{A}}\mathbf{g}\cdot\mathbf{R} - \mu_d\mathbf{R} \\ & + \nu_s(1-P_dIHFR_h)\mathbf{H} + \nu_{icu}(1-P_{dicu}IHFR_{icu})\mathbf{ICU} \\ & + \nu_{icuc}(1-P_{dicuc}IHFR_{icu})\mathbf{ICUC} + \nu_{sc}(1-P_{dhc}IHFR_h)\mathbf{HC} \\ & + \nu_{icuh}(1-P_{dicuh}IHFR_{icu})\mathbf{ICUH}\end{aligned} \tag{4.10}$$

For the baseline scenario (no NPIs in place), the system's force of infection can be written:

$$\begin{aligned}\lambda = {} & p\hat{c}[\rho\mathbf{E} + \mathbf{A} + \mathbf{I} + mean\_imports + \rho_s(\mathbf{H} + \mathbf{ICU} + \mathbf{ICUH})]/P \\ & + p(\hat{c}_{home} + \hat{c}_{other})(\mathbf{X} + \mathbf{HC} + \mathbf{ICUC})/P,\end{aligned} \tag{4.11}$$

where $\hat{c}$ represent the sum of each setting contact matrices ($\hat{c}_{home}$, $\hat{c}_{work}$, $\hat{c}_{school}$ and $\hat{c}_{other}$). In the following subsections a full expression for λ, including the effect of all NPIs, will be derived and all parameters will be defined. To represent the reporting

dynamics, we can also write the cumulative cases and deaths, respectively, as:

$$\frac{d\mathbf{C}}{dt} = r\gamma(1-IHR)(1-P_{clin})\mathbf{E} + r_c\gamma(1-IHR)P_{clin}\mathbf{E} + r_h\gamma IHR\mathbf{E} \quad (4.12)$$

$$\begin{aligned}\frac{d\mathbf{D}}{dt} = {} & \nu_s P_{dh} IHFR_h \mathbf{H} + \nu_{sc} P_{dhc} IHFR_h \mathbf{HC} + \nu_{icu} P_{dicu} IHFR_{icu} \mathbf{ICU} \\ & + \nu_{icuc} P_{dicuc} IHFR_{icu} \mathbf{ICUC} + \nu_{icuh} P_{dicuh} IHFR_{icu} \mathbf{ICUH} \\ & + \mu_d(\mathbf{H} + \mathbf{HC} + \mathbf{ICU} + \mathbf{ICUC} + \mathbf{ICUH} + \mathbf{I} + \mathbf{X})\end{aligned} \quad (4.13)$$

where each of the dynamic variables (defined in Table 4.1) is further subdivided in 19 age classes consisting of 5 years age bins (0-4,5-9, up to 90+). Thereby, each of the parameters written in the model, aside from **A** (ageing matrix), should be thought of as diagonal matrices containing parameter values corresponding to each age class. Take, as an example, the natural mortality rate, given by

$$\hat{\mu_d} = \mathrm{diag}(\mu_{d1}, \mu_{d2}, ..., \mu_{dD}) = \mathrm{diag}(\vec{\mu_d}). \quad (4.14)$$

Code	Equations	Description
S	**S**	Susceptible population
E	**E**	Infected and presymptomatic population
I	**A**	Infected population, asymptomatic and not isolated
CL	**I**	Infected population, mildly symptomatic and not isolated
X	**X**	Infected population, mildly symptomatic and self-isolated at home
H	**H**	Infected population, hospitalized in simple bed.
HC	**HC**	Infected population that require hospital treatment but but are denied, due to healthcare system overload
ICU	**ICU**	Infected population, hospitalised in Intensive Care Units (ICU).
ICUH	**ICUH**	Infected population that require ICU but are hospitalised in simple beds, due to unavailability in ICU beds.
ICUC	**ICUC**	Infected population that require ICU but are denied both an ICU or hospital simple bed, due to healthcare system overload.
R	**R**	Recovered population
C	**C**	Cumulative reported cases
CM	**D**	Cumulative reported deaths

Table 4.1: List of model variables definitions and notations both for the equations and the R code.

Note that, in the system of equations presented above, we suppress the hat or bold from all diagonal matrices to avoid an overloaded notation, but we choose to keep all dynamic variables in bold as a reminder that each of them actually represents a vector of dimension $D = 19$. Hence, the dimension of the ordinary differential equation system that needs to be solved is $D \cdot 11$, where 11 is the number of

Code	Equation	Description	Value	Source
lam	λ	force of infection	Variable	Eq. (4.28)
mort	μ_d	natural mortality ($days^{-1}$)	Age dependent	Instituto Brasileiro de Geografia e Estatísti
ageing	$\bar{\mathbf{A}}\bar{\mathbf{g}}$	speed of population ageing ($days^{-1}$)	-	-
birth	μ_b	birth rate ($days^{-1}$)	-	Instituto Brasileiro de Geografia e Estatísti
gamma	γ	Inverse of incubation period ($days^{-1}$)	$1/5.8$	Wei et al. [153]
ihr	IHR	Infection hospitalisation rate	Age dependent	Salje et al. [131]
omega	ω	Rate of which recovered people become susceptible again ($days^{-1}$)	0	Assumed
rho	ρ	Relative infectiousness of presymptomatic individuals	0.105	Wei et al. [152]
rhos	ρ_s	Relative infectiousness of hospitalised individuals (reduced due to hospitalisation)	0.10	Assumed
pclin	P_{clin}	Proportion of symptomatic individuals	0.30 (0-19) 0.56 (20-59) 0.69 (60+)	SMSSP Sun et al. [140] Sun et al. [140]
selfis	P_{selfis}	Proportion of symptomatic individuals who self-isolate	Variable	Assumed
prob_icu	P_{icu}	Proportion of hospitalised individuals who need ICU beds	Age dependent	SIVEP [22]
critH	H_c	Proportion of hospitalised individuals who have not received attendance	Variable	Section 4.2.2
critICU	ICU_c	Proportion of hospitalised individuals who need ICU beds and have not received one	Variable	Section 4.2.2
critICUH	ICU_h	Proportion of hospitalised individuals who need ICU beds and have not received one and also not have received simple beds	Variable	Section 4.2.2
nui	ν_i	Recovery rate of mild symptomatic/asymptomatic individuals ($days^{-1}$)	1/9	Cevik et al. [38]
nus	ν_s	Recovery/death rate of hospitalised individuals ($days^{-1}$)	1/8.3	SIVEP [22]
nusc	ν_{sc}	Recovery/death rate of hospitalised individuals who have not received attendance ($days^{-1}$)	1/11	Assumed
nu_icu	ν_{icu}	Recovery/death rate of hospitalised individuals in ICU beds ($days^{-1}$)	1/14.7	SIVEP [22]
nu_icuh	ν_{icuh}	Recovery/death rate of hospitalised individuals who need ICU beds but received simple beds ($days^{-1}$)	1/11	Assumed
nu_icuc	ν_{icuc}	Recovery/death rate of hospitalised individuals who need ICU beds and have not received attendance ($days^{-1}$)	1/11	Assumed
ifr[,3]	$IHFR_{icu}$	In hospital fatality rate (ICU required)	Age dependent	Portella et al. [121]
ifr[,4]	$IHFR_h$	In hospital fatality rate (common bed)	Age dependent	Portella et al. [121]
pdeath_h	P_d	Maximum probability of death for a hospitalised infection requiring common bed	45.9	SIVEP [22]
pdeath_icu	P_{dicu}	Maximum probability of death for a hospitalised infection requiring ICU	69	SIVEP [22]
pdeath_hc	P_{dhc}	Maximum probability of death for a hospitalised infection requiring common bed but not receiving attendance	80	Assumed
pdeath_icuh	P_{dicuh}	Maximum probability of death for a hospitalised infection requiring ICU but receiving common bed attendance	97	Assumed
pdeath_icuc	P_{dicuc}	Maximum probability of death for a hospitalised infection requiring ICU but not receiving attendance	99	Assumed
report	r	Report rate of asymptomatic cases	0.00	Assumed
reportc	r_c	Report rate of symptomatic cases	0.01	Assumed
reporth	r_h	Report rate of hospitalized cases	0.95	Assumed
give	q	Threshold of occupancy for loss of health system efficiency	0.65	Assumed

Table 4.2: List of baseline model parameters used in the system 4.1-4.10.

compartments in the system. A description of each parameter from the model is available at table 4.2.

The full model has over 200 equations, but the main mechanisms regarding infection and non-pharmaceutical intervention (NPI) effects are encoded in its force of infection, λ. The implemented NPI, their effect on contact matrices and, finally, their effect on λ will be described in the following subsections.

Non-pharmaceutical interventions

We have implemented a varied set of NPI that affected the age-dependent contact rates encoded by the setting-specific contact matrices. Here we describe each of the seven NPI implemented in our model:

- *Self-Isolation*: Symptomatic individuals that do not require hospitalization voluntarily isolate themselves during the time of infection and reduce the chance of infecting others. The beginning and end period of this intervention is defined by $\theta_{selfis}(t)$ and represents the days t when the population adheres to this behavior. The impact of this NPI depends on its adherence to self-isolation $selfis_{cov}$ and estimated reduction in contacts by self-isolation $selfis_{eff}$ values, where

$$P_{selfis} = selfis_{cov}(t) selfis_{eff} \theta_{selfis}(t) \tag{4.15}$$

- *Social Distancing*: the population avoids or reduces contacts in the community setting ($\hat{c}_{com}$). This intervention comprises reduction of contacts on churches, markets, social events and gatherings, shopping activities, gyms, and others. The beginning and end period of this intervention is defined by $\theta_{dist}(t)$. The impact of this NPI depends on its adherence to social distancing at community level $(dist_{cov})$ and reduction of contacts in the community among those adhering to social distancing $(dist_{eff})$ values, where:

$$dist(t) = dist_{cov}(t) dist_{eff} \theta_{dist}(t); \tag{4.16}$$

- *Use of masks*: This intervention comprises individual protection measures, given by the adoption of mask usage. The beginning and end period of this intervention is defined by $\theta_{mask}(t)$. The impact of this NPI depends on its adherence to mask usage $(mask_{cov})$ and effectiveness $(mask_{eff})$, where

$$mask(t) = mask_{cov}(t) mask_{eff} \theta_{mask}(t); \tag{4.17}$$

- *Work from home*: This intervention reduces contacts in the work environment ($\hat{c}_{work}$) as workers perform their activities from their home. The beginning and end period of this intervention is defined by $\theta_{work}(t)$. The impact of this

NPI depends on the adherence to home-office ($work_{cov}$) and reduction of contacts at work among those adhering to home-office ($work_{eff}$), where:

$$work(t) = work_{cov}(t) work_{eff} \theta_{work}(t); \tag{4.18}$$

- *School closure*: This intervention reduces the contacts in the school setting ($\hat{c}_{school}$) due to limitation of in-school activities or school closures. The beginning and end period of this intervention is defined by $\theta_{school}(t)$. The effectiveness of this NPI depends on the adherence to online (not in-person) school activities ($school_{cov}$) and the reduction of contacts in school upon school closure ($school_{eff}$), where:

$$school(t) = school_{cov}(t) school_{eff} \theta_{school}(t); \tag{4.19}$$

  Note that in the main text, $school_{cov}$ is also referred as PCS (potential contacts in school).

- *Cocoon elderly*: This intervention reduces the contacts to a proportion of the older adult population, given a minimum age $D^{\dagger}$. The beginning and end period of this intervention is defined by $\theta_{cocoon}(t)$.The effectiveness of this NPI depends on the adherence to cocooning of older adults ($cocoon_{cov}$) and reduction of contacts with older adults in all settings as a results of cocooning older adults ($cocoon_{eff}$).

- *Travel ban*: This intervention models the interruption of travel flow from outside the city and the isolation of cases coming from outside, which reduces or eliminate import cases. This intervention is given by:

$$imports = (1 - travel_{eff}) mean\_imports \tag{4.20}$$

  where ($mean\_imports$) is the mean value of imported cases, $travel_{eff}$ the effectiveness of this intervention, and $imports$ the number of new cases that are added to the population per day.

A thorough description of the NPI related parameters can be found in tables 4.3 and 4.4 (with table 4.4 presenting values valid for standard and percolation model values, while table 4.5 has values valid for a third model version with increase NPI coverge, as described in section 4.4.2). Figure 4.4 shows coverages ($NPI_{cov}\theta_{NPI}(t)$), efficiencies (NPI_{eff}) and their corresponding product $NPI_{cov}NPI_{eff}\theta_{NPI}(t)$ over time (valid for the standard and percolation model versions).

Effect of NPI on contact matrixes

Assuming that a Susceptible individual in the n-th age class, S_n, can be in contact with any of the infected groups among all age classes, we have that

$$\frac{dS_n}{dt} \propto -S_n \sum_j c_{n,j}(a_j^A A_j + a_j^E E_j + a_j^I I_j + ...) \tag{4.21}$$

Code	Equation	Description	Value	Source
mask_cov	$mask_{cov}$	Adherence to mask usage	See Table 4.4	[74; 141; 144]
mask_eff	$mask_{eff}$	Estimated reduction of contact due to mask use	0.85	Chu et al. [41]
selfis_cov	$selfis_{cov}$	Adherence to self-isolation	See Table 4.4	[74; 141; 144]
selfis_eff	$selfis_{eff}$	Estimated reduction of contact due to self-isolation if symptomatic	0.80	Assumed
dist_cov	$dist_{cov}$	Adherence to social distancing in community level	See Table 4.4	[74; 141; 144]
dist_eff	$dist_{eff}$	Reduction of contacts in the community among those adhering to social distancing	0.95	Assumed
work_cov	$work_{cov}$	Adherence to work from home policies	See Table 4.4	[74; 141; 144]
work_eff	$work_{eff}$	Reduction of contacts at work among those adhering to work from home policies	0.95	Assumed
school_cov	$school_{cov}$	Adherence to online (not in-person) school activities	See Table 4.4	[74; 141; 144]
school_eff	$school_{eff}$	Reduction of contacts in school upon school closure	1.00	Assumed
cocoon_cov	$cocoon_{cov}$	Adherence to cocooning of older adults	See Table 4.4	[74; 141; 144]
cocoon_eff	$cocoon_{eff}$	Reduction of contacts with older adults in all settings as a result of cocooning older adults	0.95	Assumed
travel_eff	$travel_{eff}$	Effectiveness of travel interruption policies	See Table 4.4	Assumed
mean_imports	$mean\_imports$	Mean number of infected individuals that travel to the study site	0.2	Assumed

Table 4.3: Brief description of intervention parameters, together with values (if not time-dependent) and sources. In the case of time-dependent parameters, we supply additional tables with dates of interventions for the models standard and percolation (Table 4.4) and standard +30% NPIs (Table 4.5).

Parameter	Start date	End date	Value
$selfis_{cov}$	2020-03-24	2020-08-31	0.70
$dist_{cov}$	2020-03-18	2020-05-31	0.70
$dist_{cov}$	2020-06-01	2020-06-30	0.59
$dist_{cov}$	2020-07-01	2020-08-31	0.45
$school_{cov}$	2020-03-21	2020-08-31	0.95
$mask_{cov}$	2020-03-19	2020-05-31	0.20
$mask_{cov}$	2020-06-01	2020-06-30	0.35
$mask_{cov}$	2020-07-01	2020-08-31	0.42
$work_{cov}$	2020-03-16	2020-05-31	0.60
$work_{cov}$	2020-06-01	2020-06-30	0.48
$work_{cov}$	2020-07-01	2020-08-31	0.36
$cocoon_{cov}$	2020-03-14	2020-05-31	0.10
$cocoon_{cov}$	2020-06-01	2020-06-30	0.40
$cocoon_{cov}$	2020-07-01	2020-07-31	0.50
$cocoon_{cov}$	2020-08-01	2020-08-31	0.60
$travel_{eff}$	2020-02-19	2020-03-18	0.00
$travel_{eff}$	2020-03-19	2020-08-31	0.70

Table 4.4: Values of time-dependent interventions used for model fitting in the case of Standard and Percolation model. The patterns observed in [74; 141; 144] were used as qualitative proxies for intervention coverage values.

Parameter	Start date	End date	Value
$selfis_{cov}$	2020-03-24	2020-08-31	0.910
$dist_{cov}$	2020-03-18	2020-05-31	0.910
$dist_{cov}$	2020-06-01	2020-06-30	0.767
$dist_{cov}$	2020-07-01	2020-08-31	0.585
$school_{cov}$	2020-03-21	2020-08-31	1.000
$mask_{cov}$	2020-03-19	2020-05-31	0.258
$mask_{cov}$	2020-06-01	2020-06-30	0.458
$mask_{cov}$	2020-07-01	2020-08-31	0.552
$work_{cov}$	2020-03-16	2020-05-31	0.780
$work_{cov}$	2020-06-01	2020-06-30	0.624
$work_{cov}$	2020-07-01	2020-08-31	0.468
$cocoon_{cov}$	2020-03-14	2020-05-31	0.130
$cocoon_{cov}$	2020-06-01	2020-06-30	0.520
$cocoon_{cov}$	2020-07-01	2020-07-31	0.650
$cocoon_{cov}$	2020-08-01	2020-08-31	0.780
$mean\_imports$	2020-02-19	2020-08-31	0.140
$travel_{eff}$	2020-02-19	2020-03-18	0.000
$travel_{eff}$	2020-03-19	2020-08-31	0.700

Table 4.5: Values of time dependent interventions used for model fitting with stronger interventions (Standard model + 30% NPIs). Here we created an hypothetical intervention scenario with adherence 30% higher than the ones described in Table 4.4.

where $c_{n,j}$ measures the contact strength between people of n-th and j-th age classes, forming the contact matrix $\hat{c}$, of dimension $D \times D$. The $a_j^A, a_j^E, a_j^I, ...$ measures how infectious these different model compartments are. For example, asymptomatic people, **A**, may be more infectious than the symptomatic ones, **I**, since they may not be isolating themselves, given they are unaware of their infectious state. The NPIs are considered as modifications on both $\hat{c}$, reducing contacts between people, and also on the different $a_j^A, a_j^I, ..$, accounting for behavioural aspects, such as increased hand hygiene.

Given the definitions above, we are finally able to breakdown the general structure of $\hat{c}$. It is mainly composed of 4 matrices, $\hat{c}_{home}$ - which measures the amount of contacts of people at home, $\hat{c}_{work}$ - for contacts at work, $\hat{c}_{school}$ - for contacts at school and $\hat{c}_{other}$ - for other kinds of human interactions, such as going to restaurants, movies and churches. Therefore, in absence of any NPIs, the resultant contacts matrix would be simply

$$\hat{c} = \hat{c}_{home} + \hat{c}_{work} + \hat{c}_{school} + \hat{c}_{other}. \tag{4.22}$$

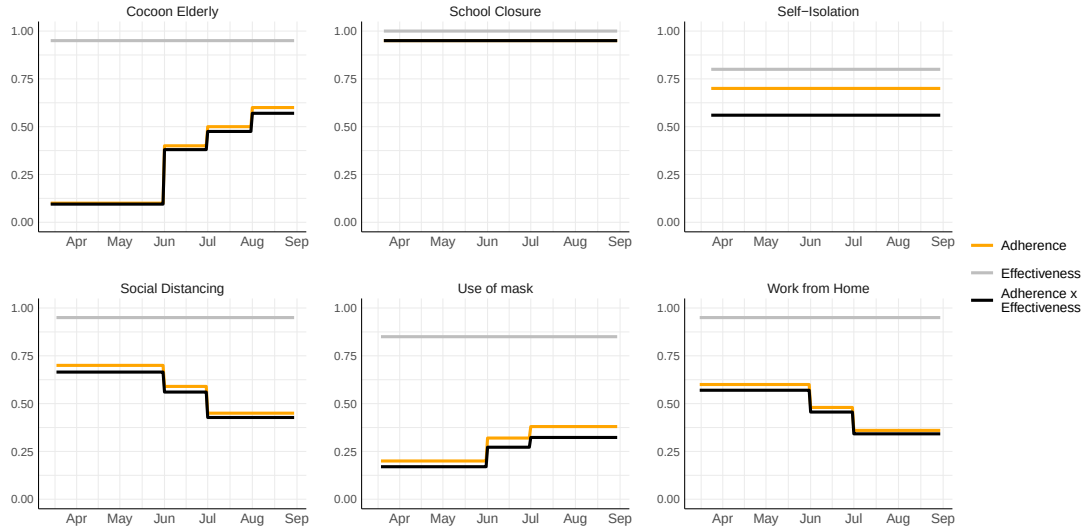


Figure 4.4: Diagram of adherence, efficiency (estimated reduction of contact due to each NPI) and their product over time ($NPI_{cov}NPI_{eff}\theta_{work}(t)$), for each of the non-pharmaceutical interventions considered both in the standard model and the model with percolation. The diagrams for the model with 30% more NPI adherence are similar, only with 30% higher values for adherence and the product. (Source: Franco et al. [69])

As NPIs are inserted, the contact matrix is modified. Suppose the simple case of home-office policies, that is, people should work at home for a period of $work_{dur}$ weeks. A fraction $work_{cov}$ of the population is able to adhere to such policies, and they have an effectiveness $work_{eff}$ in reducing this kind of contact, then we have that the contact matrix becomes

$$\hat{c} = \hat{c}_{home} + (1 - work_{cov}work_{eff}\theta_{work}(t))\hat{c}_{work} + \hat{c}_{school} + \hat{c}_{other}. \tag{4.23}$$

Here, $\theta_{work}(t)$ is a function that measures if either home office policies are being applied or not, and is usually a step function, being 1 during the period $work_{dur}$, defined by starting and finishing dates of such policies, and 0 out of said period. We could apply this same mathematical form for other interventions, such as school closing, commerce and restaurants functioning in reduced periods or not functioning at all, and many other NPIs we have seen being tried out in order to contain virus spread. We would get

$$\begin{aligned} \hat{c} \quad = \quad & \hat{c}_{home} + \\ & (1 - work_{cov}work_{eff}\theta_{work}(t))\hat{c}_{work} + \\ & (1 - school_{cov}school_{eff}\theta_{school}(t))\hat{c}_{school} + \\ & (1 - dist_{cov}dist_{eff}\theta_{dist}(t))\hat{c}_{other} \end{aligned} \tag{4.24}$$

With this final form of contact matrix we can add yet another possible NPI, cocooning the elderly. That means they are more isolated and protected, since they

are one of the most vulnerable to COVID-19 death. When cocooning is applied, the contacts of people above a certain age, let's say age_{cocoon}, are reduced in $cocoon_{cov}cocoon_{eff}$, implying that $\hat{c}$ values from the $D^{\dagger}$ to D lines and rows must be reduced. Here, $D^{\dagger}$ is the index from which cocooning starts, for example, if cocooning is applied in people over 65 years old, then $D^{\dagger} = 13$. Defining $\eta = 1 - cocoon_{cov}cocoon_{eff}\theta_{cocoon}(t)$ we write

$$\hat{g}_{D^{\dagger}}(\eta) = \text{diag}(\vec{1}_{D^{\dagger}}, \eta\vec{1}_{D-D^{\dagger}}) \tag{4.25}$$

and with it, make the final contacts matrix

$$\overline{c} \;=\; \hat{g}_{D^{\dagger}}(\eta)\hat{c}\hat{g}_{D^{\dagger}}(\eta) \tag{4.26}$$

Note that

$$\hat{g}_{D^{\dagger}}(\eta)\hat{c}\hat{g}_{D^{\dagger}}(\eta) = \begin{bmatrix} c_{1,1} & c_{1,2} & \cdots & c_{1,D^{\dagger}-1} & \eta c_{1,D^{\dagger}} & \cdots & \eta c_{1,D} \\ c_{2,1} & c_{2,2} & \cdots & c_{1,D^{\dagger}-1} & \eta c_{1,D^{\dagger}} & \cdots & \eta c_{1,D} \\ \vdots & \vdots & \ddots & & \vdots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots & & \vdots \\ \eta c_{D^{\dagger},1} & \eta c_{D^{\dagger},2} & \cdots & \eta c_{D^{\dagger},D^{\dagger}-1} & \eta^2 c_{D^{\dagger},D^{\dagger}} & \cdots & \eta^2 c_{D^{\dagger},D} \\ \vdots & \vdots & & \vdots & \vdots & \ddots & \vdots \\ \eta c_{D,1} & \eta c_{D,2} & \cdots & \eta c_{D,D^{\dagger}-1} & \eta^2 c_{D,D^{\dagger}} & \cdots & \eta^2 c_{D,D} \end{bmatrix} \tag{4.27}$$

so the elderly are more isolated among themselves, since $\eta < 1 \implies \eta^2 < \eta$, whilst still having reduced contacts with the other age classes.

Note that, in the above implementation, we are assuming it to be reasonable for the contact rates to decrease linearly with increasing adherence to mitigating strategies in schools, workplaces or other settings. Nevertheless, owing to the fact that SEIR-like models assume underlying homogeneously mixed populations, this assumption might not hold for household settings as their connectivity structure is dynamically affected by NPIs in a non-homogeneous fashion. In section 4.3, we discuss an alternative model structure that takes into account such an effect at household-level.

Force of infection

Taking into account that Self-Isolated ($\mathbf{X}$) individuals are only able to infect through home and "other" matrices, we can wrap everything in a force of infection given by:

$$\begin{aligned}\lambda = &(1 - hand(t))p\overline{c}[\rho\mathbf{E} + \mathbf{A} + \mathbf{I} + imports + \rho_s(\mathbf{H} + \mathbf{ICU} + \mathbf{ICUH})]/P \\ &+ (1 - hand(t))p(\overline{c}_{home} + \overline{c}_{other})(\mathbf{X} + \mathbf{HC} + \mathbf{ICUC})/P\end{aligned} \tag{4.28}$$

where $hand(t) = hand_{cov}hand_{eff}\theta_{hand}(t)$.

4.2.2 Hospital burden

We have slightly changed the way hospital burden is implemented, compared to Aguas et al. [5]. We assume that, if the occupation of beds is under some threshold value the health system infrastructure is able to handle correctly any new entrance to the hospital. After this threshold q, some of the patients might not find the needed support due to hospital overload, until full capacity, where patients are not accepted anymore. This is modelled to estimate H_C and ICU_C by the following function:

$$f(x) = \begin{cases} 0, & \text{if } x \leq q, \\ 1 - (x(b - ax) + c), & \text{if } q \leq x \leq 1 \\ 1, & \text{if } x > 1 \end{cases} \tag{4.29}$$

Where x is the ratio between number of patients and available beds and a, b, c are computed in a way to ensure continuity of the function and its first derivative:

$$a = \frac{1}{2q(q-1) + (q^2 - 1)} \tag{4.30}$$

$$b = 2aq \tag{4.31}$$

$$c = a - b \tag{4.32}$$

With this definition, we have ICU_c being computed using
$x \equiv \mathbf{ICU}/\text{number of ICU beds}$, H_c computed using
$x \equiv (\mathbf{ICUH} + \mathbf{H})/\text{number of common beds}$ (as $\mathbf{ICUH}$ uses common beds).
Finally, to model priority of common beds to ICU needing individuals compared to $\mathbf{H}$ we simply assume that $ICU_h = H_c^2$, since these values are between 0 and 1, therefore $ICU_h \leq H_c$. Figure 4.5 shows these values assuming that $q = 0.65$.

4.2.3 Basic Reproduction Number

To calculate the Basic Reproduction Number (R_0) through the Next Generation Matrix (NGM) Method [6, Chapter 6] we need to redefine the model as a system of

differential equations subdivided into two groups: the infected one, $\mathbf{y} = (\mathbf{E}, \mathbf{A}, \mathbf{I}, \mathbf{X}, \mathbf{H}, \mathbf{HC}, \mathbf{ICU}, \mathbf{ICUH}, \mathbf{ICUC})^T$, and the non-infected one, $\mathbf{z} = (\mathbf{S}, \mathbf{R}, \mathbf{D})^T$. We can then write the system as

$$\dot{\mathbf{y}} = F(\mathbf{y}, \mathbf{z}) - G(\mathbf{y}, \mathbf{z}) \tag{4.33}$$

$$\dot{\mathbf{z}} = J(\mathbf{y}, \mathbf{z}) \,, \tag{4.34}$$

where F stands for the transition of Susceptible individuals, $\mathbf{S}$, into infected ones, $\mathbf{y}$, whilst G accounts for transitions within infected classes, from exposed to all other classes in $\mathbf{y}$, as well as recoveries and deceases. The function J accounts for the counterparts of these same effects into the $\mathbf{z}$ equations. Then, to calculate R_0, we linearize the system around disease free equilibrium, where $\mathbf{y} \approx \mathbf{0}$ and $\mathbf{z} \approx (\mathbf{P}(t=0), \mathbf{0}, \mathbf{0})^T$, where $\mathbf{P}(t)$ is the population age distribution at time t. That way, the $\mathbf{y}$ equation becomes

$$\dot{\mathbf{y}} = (\hat{F} - \hat{G})\mathbf{y} \,, \tag{4.35}$$

where $\hat{F}$ and $\hat{G}$ are the linearized matrices that come from the functions F and G, respectively. Noticing that the only entrance of new infected comes from the λS terms in the exposed classes, $\mathbf{E}$. That way, only the first D lines/rows of $\hat{F}$ are not null. That way, defining

$$\hat{\sigma} = \mathrm{diag}(\mathbf{S})\, \bar{c}\, \mathrm{diag}(\mathbf{P}(t=0))^{-1} \tag{4.36}$$

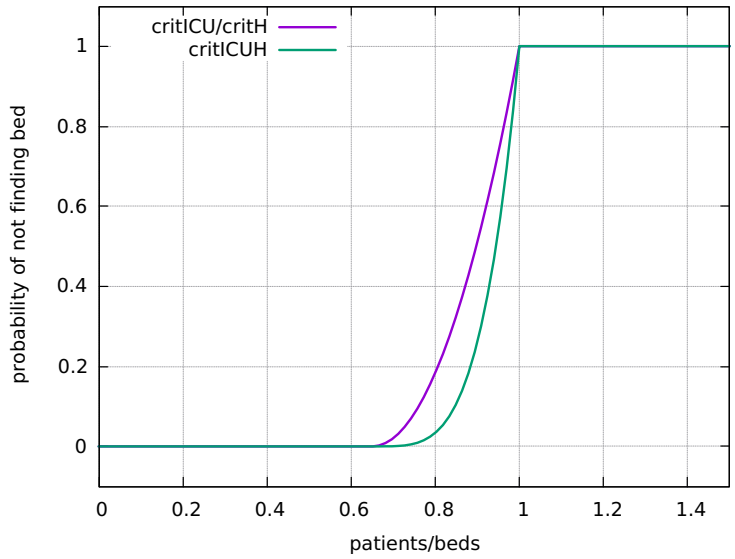


Figure 4.5: Probability of not finding a bed as function of the current occupancy. $q = 0.65$. (Source: Franco et al. [69])

and

$$\hat{\sigma}_{ho} = \text{diag}(\mathbf{S})\ (\bar{c}_{home} + \bar{c}_{other})\ \text{diag}(\mathbf{P}(t=0))^{-1} \tag{4.37}$$

allow us to write

$$\hat{F} = (1 - hand(t))p \begin{bmatrix} \rho\hat{\sigma} & \hat{\sigma} & \hat{\sigma} & \hat{\sigma}_{ho} & \rho_s\sigma & \hat{\sigma}_{ho} & \rho_s\sigma & \rho_s\sigma & \hat{\sigma}_{ho} \\ & & & & \mathbb{0}_{8D,9D} & & & & \end{bmatrix} \tag{4.38}$$

The matrix $\hat{G}$ will have its first D columns as terms of exposed, $\mathbf{E}$, becoming the other infected compartments considered, that is,
$(\mathbf{A}, \mathbf{I}, \mathbf{X}, \mathbf{H}, \mathbf{HC}, \mathbf{ICU}, \mathbf{ICUH}, \mathbf{ICUC})$. It's diagonal blocks are terms of recovery, decease and ageing effects. For the sake of simplifying notation, let
$\hat{\phi}_j = \nu_j - \bar{A}g + \mu_d$, for $j = i,\ s,\ sc,\ icu,\ icuh,\ icuc$, so that the full form of $\hat{G}$ can be given by

$$\hat{G} = \begin{bmatrix} \gamma - \bar{A}g + \mu_d & \mathbb{0} & & & \dots & & & \mathbb{0} \\ -(1-P_{clin})(1-IHR)\gamma & \hat{\phi}_i & & & & & & \\ -P_{clin}(1-P_{selfis})(1-IHR)\gamma & \mathbb{0} & \hat{\phi}_i & & & & & \\ -P_{clin}P_{selfis}(1-IHR)\gamma & & & \hat{\phi}_i & \ddots & & & \vdots \\ -IHR(1-P_{icu})(1-H_c)\gamma & & & & \hat{\phi}_s & & & \\ -IHR(1-P_{icu})H_c\gamma & \vdots & & & \ddots & \hat{\phi}_{sc} & & \\ -IHR\,P_{icu}(1-ICU_c)\gamma & & & & & & \hat{\phi}_{icu} & \\ -IHR\,P_{icu}ICU_c(1-ICUH_c)\gamma & & & & & & & \hat{\phi}_{icuh} & \mathbb{0} \\ -IHR\,P_{icu}ICU_cICUH_c\gamma & \mathbb{0} & & & \dots & & & \mathbb{0} & \hat{\phi}_{icuc} \end{bmatrix} \tag{4.39}$$

Finally, we can define the NGM as in chapter 6 of Allen et al. [6], $\hat{M}_{NGM} = \hat{F}\hat{G}^{-1}$. With that, R_0 is defined as the spectral radius of $\hat{M}_{NGM}$, which in the simplest cases is just its dominant eigenvalue, here calculated with rARPACK R package.

4.3 Percolation model structure

The system described in the previous section introduced a phenomenological implementation that models the NPI mitigating effect through a proportional reduction of the rates in the contact matrices. However, this approach failed to show a substantial decrease the transmission even when considering high coverages of NPIs. For instance, at hypothetical coverages and efficacies of 100% for school closure, work from home, and social distancing, the infection rate would still not go down immediately. Consequently, we found that data fitting was challenging when trying to simulate NPI coverages as closely as possible to the patterns seen in real life.

Moreover, NPIs affected all contact matrices in the standard model, with the exception of the household matrix (which accounts for both contacts within and between households, indistinguishably), implying that the dynamics among households persisted as fully well-mixed, even under high coverage of NPIs. In reality, however, the contact network among households should become more sparse as connecting pathways are removed, in such a way that the "effective" contact rates should be considerably reduced. We draw a comparison with percolation theory to address the fact that, above a certain threshold value of combined NPI adherence, inter-household contacts are much less common.

In a non-NPI scenario, we could picture the whole population as a collection of households assigned as vertices with links between households (created due to interactions at work, schools, or other places) represented by edges. This highly connected network would form one giant connected component, which could therefore be approximated as a homogeneously mixed population [13]. However, by introducing social distancing measures and increasing their coverages, we break connections in the network until long-range connectivity is lost. The critical value of number of connections (or edge density) where this transition happens is the so-called percolation threshold [61]. Above the percolation threshold of contact loss, i.e., in a high coverage social distancing scenario, we would expect to see the emergence of small household clusters that, though well connected within themselves, would be poorly connected between them.

Here we model the percolation effect on home contact matrices assuming that, while contacts outside households are kept above the percolation threshold, the effect of mitigation strategies is less apparent in home settings. On the other hand, the probability of infection of a susceptible individual drops drastically when its mean number of contacts goes below a threshold [117]. Mathematically, we modify the overall contact matrix defined in Equation 4.24, $\hat{c}$, by multiplying the home contact matrix, $\hat{c}_{home}$, by a modulating factor dependetn on all NPI. The resultant contact matrix can be written

$$\begin{aligned} \hat{c} \;=\; & (1 - f_{perc})\hat{c}_{home} + \\ & (1 - work_{cov}work_{eff}\theta_{work}(t))\hat{c}_{work} + \\ & (1 - school_{cov}school_{eff}\theta_{school}(t))\hat{c}_{school} + \\ & (1 - dist_{cov}dist_{eff}\theta_{dist}(t))\hat{c}_{other}, \end{aligned} \tag{4.40}$$

where f_{perc} is the percolation correction function, which is dependent on all NPI coverages and efficacies, and satisfies the following requirements:

- $0 \leq f_{perc} \leq 1$;
- $f_{perc} \to 0$ as NPI are completely lifted; for low adherence to NPI, no connectivity loss should be noticed since different households would still be strongly connected;

- $1 - f_{perc} \ll 1$ for high effectiveness of NPI; as connections between different households are widely severed, so f_{perc} approaches its maximum value.

A hyperbolic functional form, as follows, is proposed to model this effect:

$$f_{perc} = \frac{h_{eff}}{2}\left[1 + \tanh\left(h_{steep}(W_{NPI} - T_{perc})\right)\right] \tag{4.41}$$

where $0 < h_{eff} < 1$ is the maximum reduction in contacts, a limit introduced due to the percolation effect; $0 < T_{perc} < 1$ is the percolation threshold, i.e., the fraction of connections we need to remove so that the network no longer percolates; and h_{steep} is the steepness of the percolation correction function determining how fast the network percolates: a large h_{steep} implies f_{perc} changing abruptly from 0 to h_{eff} when $W_{NPI} \approx T_{perc}$ (see Figure 4.6).

Finally, W_{NPI} is defined as the *combined NPI adherence*, where the resulting reduction of contacts due to each type of intervention are weighted by the age and contact structure of the population they are applied to, namely

$$W_{NPI} = p_{work}a_{work}(t) + p_{school}a_{school}(t) + p_{other}a_{dist}(t) \tag{4.42}$$

where, for instance, work adherence (a_{work}) can be written as the product of the respective effectiveness ($work_{eff}$), coverage ($work_{cov}$) and duration of home-office strategies (translated as the *step function* $\theta_{work}(t)$):

$$a_{work}(t) = work_{eff}work_{cov}\theta_{work}(t), \tag{4.43}$$

The same holds for $school$ and $dist$, forming the set $\{a_j(t)\}$, for $j = \{work,\ school,\ dist\}$, which throughout the text will be referenced as adherences. The weights p_j, $j = \{work,\ school,\ other\}$, are calculated as

$$p_j = \frac{\mathbf{P}(t=0)\hat{c}_j\mathbf{P}^T(t=0)}{\mathbf{P}(t=0)(\hat{c}_{work} + \hat{c}_{school} + \hat{c}_{other})\mathbf{P}^T(t=0)}, \tag{4.44}$$

where $\mathbf{P}(t=0)$ is the initial age distribution of the population and c_j the contact matrices in each setting.

Now, clearly $0 < W_{NPI}(t) < 1$. Note also that $W_{NPI}(t)$ has this specific form to consider population structure, accounting for how much each of these NPIs effectively reduce contacts in each age-class. Moreover, $W_{NPI}(t)$ can vary according to the current implemented interventions, adding flexibility to the model.

Finally, considering all effects due to NPIs and percolation, the effective contact matrix, $\hat{c}$, is written as

$$\begin{aligned} \hat{c} \;=\; & (1 - f_{perc}(t))\hat{c}_{home} + \\ & (1 - a_{work}(t))\hat{c}_{work} + \\ & (1 - a_{school}(t))\hat{c}_{school} + \\ & (1 - a_{dist}(t))\hat{c}_{other}. \end{aligned} \tag{4.45}$$

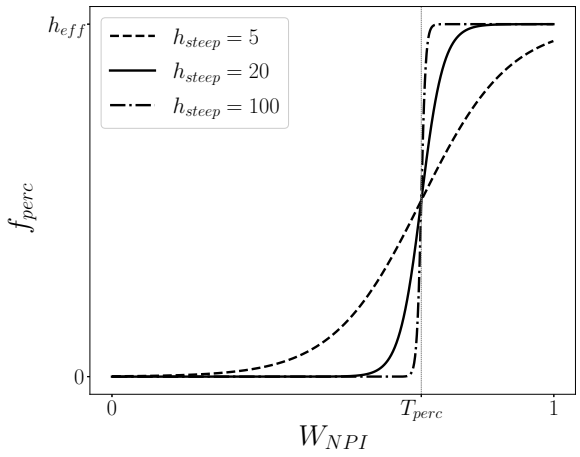


Figure 4.6: Percolation correction as a function of the combined adherence to interventions. We plot the resultant curve for different values of h_{steep}, which is the steepness of the hyperbolic tangent in the definition of f_{perc} (Eq. 4.41). Here we used $T_{perc} = 0.6$ for the percolation threshold, which is indicated as the curve's inflexion point. Note that changing its value simply dislocates the curve along the x-axis. For low W_{NPI}, $f_{perc} \to 0$, meaning no connectivity is lost at low adherence levels. Near $W_{NPI} = T_{perc}$, f_{perc} grows rapidly. At the high W_{NPI} regime, where more connections are lost, $f_{perc} \to h_{eff} \approx 1$. (Source: Franco et al. [69])

This resultant matrix can, therefore, be used to simulate the overall contact rates between age groups, throughout all settings. In the simulations presented here, the populations and setting-specific contact matrices for Brazil were obtained from publicly available data [81; 122]; all a_j are estimated based on the effect of mitigation strategies adopted by the city of São Paulo [144; 141; 74]; whereas h_{steep} and T_{perc} are obtained through fitting to epidemiological data.

4.4 Model comparison

Aiming to systematically compare the data representativeness of each of the previously described nested model structure, we have defined the following three scenarios:

- **Standard model**: derived from the CoMo model [5], without including the percolation effect, which in practice meant setting $T_{perc} = h_{steep} = h_{eff} = 0$ and fitting only p and $startdate$ to the data.
- **Standard model + 30% NPI adherence**: the model equations and parameters are the same as before, but we consider a much greater (30%) adherence to all NPIs. This potentially unrealistic instance was included here to show that an underestimated NPI adherence in the standard version would not by itself favour the model with percolation – given that we already know that it

always reduces the contact rates.

- **Model with percolation**: derived from standard model, including the percolation effect, where T_{perc}, h_{steep}, p and $startdate$ were fitted to the data.

4.4.1 Fitting models to data

We used a Levenberg-Marquardt non-linear least square regression fitting algorithm [58] to minimise the squared residuals of the predicted number of weekly cases with to the observed values. This is equivalent to maximising the likelihood of the model under the assumption that the data point errors are normally distributed (details in SM Section III). With that, we obtained the maximum likelihood estimates (*mle*) for p, $startdate$, h_{steep} and T_{perc}. This procedure was applied to test the three version of th model, as described above.

Fitting algorithm

To perform a nonlinear least squares fitting of the free parameters (p, T_{perc}, h_{steep}, $startdate$) to the data, we used the Levenberg-Marquardt algorithm implemented in the minpack.lm R package [58].

In order to fit both new cases (C) and new deaths (D), we had to account for residuals in different scales. One way to do that was by normalising each of the variables in respect to their total sum. The resulting residual (R) is, therefore:

$$R = \frac{\sum(C_{model} - C_{observed})}{\sum C_{observed}} + \frac{\sum(D_{model} - D_{observed})}{\sum D_{observed}} \tag{4.46}$$

The algorithm minimise the square of this quantity, while evaluating the respective negative log-likelihood and minimising it.

To perform this kind of non-linear optimisation, we need to input the algorithm with a series of initial guesses. We tested a wide range of $startdate$ values (from $2020-01-10$ to $2020-02-24$) and for each one we ran the fitting algorithm using several reasonable initial guesses for the other free parameters. Hence, this method gives us fitted p, T_{perc} and h_{steep} for each $startdate$ considered.

With the goal to find a probability distribution for the fitted parameters [34], we selected the run which returned the lowest residual for each $startdate$, with its respective (p, T_{perc}, h_{steep}) set. We then computed the negative log-likelihood for each start date, L_t:

$$L_t = N \ln\left(\frac{1}{N}\sum_{i=1}^{N} R_{i,t}^2,\right) + \frac{N}{2\ln(2\pi)} \tag{4.47}$$

from which we can therefore derive the probability for each startdate, given by

$$P_t = \frac{exp(-L_t + min(\{L_t\}))}{\sum_t exp(-L_t + min(\{L_t\}))}. \tag{4.48}$$

Note that equations 4.47-4.48 are valid under the simplifying assumption that the data follows a normal distribution.

Finally, maximising the probability (equivalent to minimising the negative log-likelihood), we find sets of best fitted parameters for each of the model versions considered (See Table 4.6).

Data sources

The data used consisted on time-series of total hospitalisations and deaths for severe acute respiratory illness (SARI) in São Paulo, Brazil, reported in the 2020 SIVEP-Gripe database [22], independently of antigen or molecular biological test results. Filtering the data only for test positive cases was not reasonable because, as discussed in section 2.3, testing protocols were inconsistent and COVID-19 cases represented the absolute majority of SARI cases reported in the period. To be certain records had already been consolidated, i.e. did not suffer from delayed reporting (which can be a significant issue in the Brazilian SARI notification system, demanding corrections that take into account delay distributions [103]), we restricted the data to weekly aggregate data points comprising the first 23 weeks of the pandemic in the city, from 15 March to 31 August 2020.

Based on the current literature and proxy data for mobility and coverage of NPIs [74; 144; 141], we set values for all parameters using reasonable data sources or proxies (see SM Tables III-V for all parameter values and references). Thus, there were just a few parameters that could not be inferred from the literature and therefore required to be fitted to epidemiological data. These fitted parameters were the probability of infection given a contact (p), start date of community transmission ($startdate$), and both h_{steep} and T_{perc} from Eq. 4.41 for the model that takes percolation into account.

4.4.2 Comparing model fitness

As the model with percolation simply introduces two extra parameters onto the standard model, maintaining its structure, we can consider them nested models and therefore compare the quality of the fittings using the Akaike Information Criteria (AIC) [63].

Using the results from the best fitting to data, for each nested model version i, we

calculated the respective the AIC_i as

$$AIC_i = 2NLL_{min} + 2n_{par} \tag{4.49}$$

where NLL_{min} is the minimum negative log-likelihood obtained from fitting each model, and n_{par} is the number of fitted parameters (two free parameters for the standard models and four for the model with percolation).

The AIC estimates the loss of information in data of each model, and thus the model with the lowest AIC is the best supported by the data. To compare the models we thus use ΔAIC:

$$\Delta AIC_i = AIC_i - AIC_{min} \tag{4.50}$$

where AIC_{min} is the minimum over all AIC values obtained.

The resultant AIC and ΔAIC obtained for each model version can be found in Table 4.7.

Model	Parameter	Mean	SD	Quantile 2.5%	Quantile 50%	Quantile 97.5%
Percolation	startdate	2020-01-30	-	2020-01-30	2020-01-30	2020-01-30
Percolation	p	0.0461	0.0002	0.0461	0.0461	0.0462
Percolation	T_{perc}	0.516	0.003	0.513	0.516	0.520
Percolation	h_{steep}	4.83	0.02	4.81	4.83	4.84
Standard	startdate	2020-01-15	-	2020-01-15	2020-01-15	2020-01-15
Standard	p	0.0294	≈ 0	0.0294	0.0294	0.0294
Standard + 30% NPIs	startdate	2020-01-30	-	2020-01-30	2020-01-30	2020-02-01
Standard + 30% NPIs	p	0.0402	0.0003	0.0401	0.0401	0.0411

Table 4.6: List of parameters values for the three models with mean, standard deviation (SD), and 2.5th, 50th and 97.5th percentiles.

4.5 Results

A good way to compare the two model implementations, i.e. with and without considering the percolation effect, is by calculating an *NPI-dependent basic reproduction number ($R_0(W_{NPI})$)*, using the next generation matrix (NGM) method described in section 4.2.3. This resultant $R_0(W_{NPI})$ would represent the epidemic basic reproduction number in a hypothetical fully susceptible population under the NPI corresponding to W_{NPI}.

In fact, the basic reproduction number is a quantity determined by the population affected by a pathogen in a specific environment, rather than a variable that is exclusively determined by the biology of a pathogen. Hence, we can interpret this $R_0(W_{NPI})$ as an indicator of the pathogen's transmissibility in an environment subjected to the combined NPI adherence W_{NPI}.

From Fig. 4.7, we can see that $R_0(W_{NPI})$ diverges among the two model versions (standard model and model with percolation) as the combined NPI adherence approaches the fitted percolation threshold (indicated as T_{perc}). Percolation implies considerably lower R_0 values for greater NPI adherence, which is consistent with the sharp decrease in the effect of NPIs near the percolation threshold modelled by the percolation correction function (Eq. 4.41). In summary, this result shows that, near to and abovethe percolation threshold (T_{perc}), i.e. for high combined NPI adherence and effectiveness, the standard model will always overestimate transmission rates.

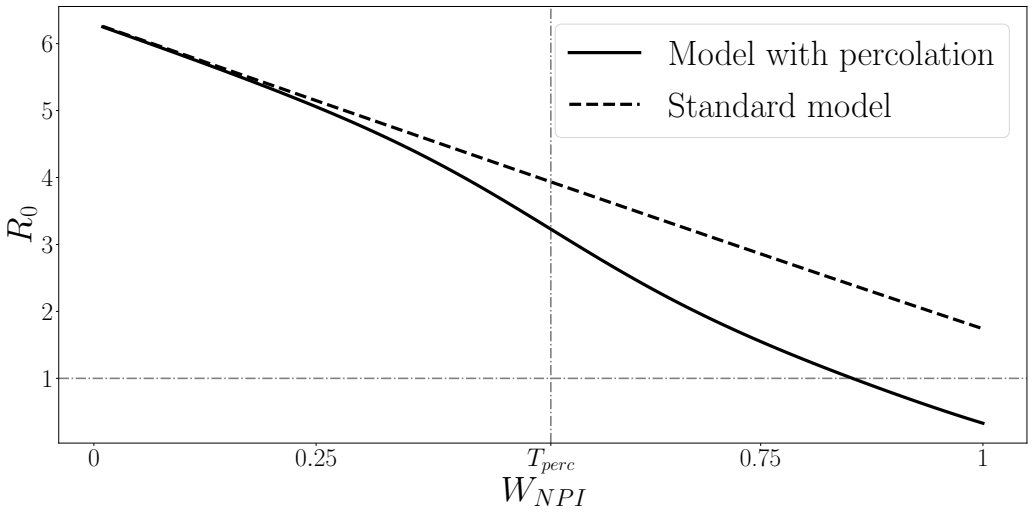


Figure 4.7: *NPI-dependent epidemic basic reproduction number* ($R_0(W_{NPI})$) as a function of the combined NPI adherence for both model versions. Note the sharp divergence near the percolation threshold ($T_{perc} = 0.514$, as a result of fitting the model with percolation to data). (Source: Franco et al. [69])

After implementing both model versions with equivalent fixed parameter sets, and also with increased NPIs in one scenario (see Tables III-V in the SM), we fitted the free parameters (among T_{perc}, h_{steep}, p and $startdate$) in each model to weekly new cases and new deaths from SARI data for São Paulo during 2020.

Fig. 4.8 shows the resulting curves for all fitted models, with confidence intervals assuming that fitted parameters follow a multivariate normal distribution. These confidence intervals are estimated by parametric bootstrap using the covariance matrix from the Levenberg-Marquardt algorithm (details in SM Section III). The comparison between the two standard model curves shows that increasing intervention coverages (even to the point of assuming unrealistic values) can modulate the epidemic curve to a certain extent, but still does not result in a good fit to data.

From the computed AIC for each model version (table 4.7) we conclude that, even though in figure 4.8 the curves for the *standard model + 30% NPI* and the *model with percolation* might look somewhat close, the AIC clearly shows a much higher level of empirical support for the model version with percolation.

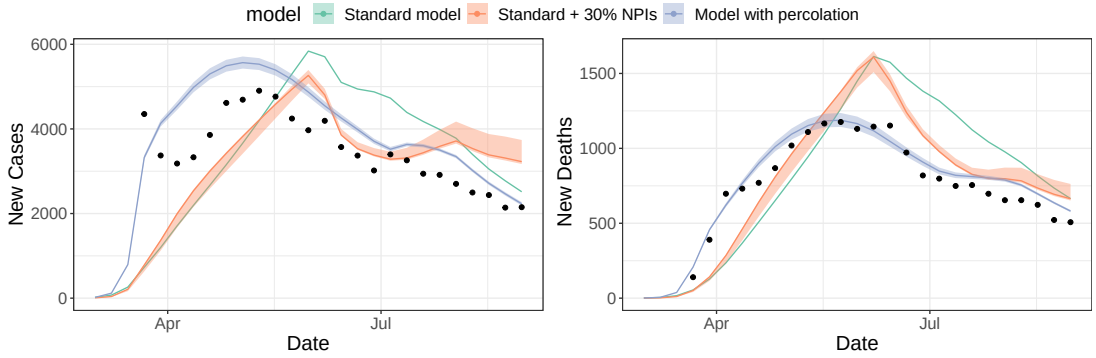


Figure 4.8: Results from simultaneous model fitting to SARI new cases and new deaths data, for each of the proposed model versions: the model with percolation (blue) and the standard model (green), using the same parameter sets, apart from the fitted parameters (p, $startdate$, T_{perc}, h_{steep}), and the standard model version with additional 30% adherence to all NPIs (red) (also fitted to data); dots represent data points from the SIVEP-Gripe database (15 March to 31 August 2020) [22]; and shaded areas represent bootstrapped confidence intervals based on parameter uncertainty from the fitting (too thin to see for the green curve). (Source: Franco et al. [69])

Model	p	$startdate$	h_{steep}	T_{perc}	AIC	ΔAIC
Standard	0.0294	2020-01-15	-	-	-756	191
Standard + 30% NPI adherence	0.0401	2020-01-30	-	-	-793	154
Percolation	0.0461	2020-01-30	4.83	0.516	-947	0

Table 4.7: Fitted parameter values and AIC evaluated for each of the model versions, as they were fitted to the same dataset of weekly SARI cases and deaths for São Paulo, which consisted of 23 data points. ΔAIC is the difference between the AIC for each model and the minimum overall AIC.

4.6 Conclusion

Modelling the ongoing COVID-19 pandemic represented an unprecedented challenge, as we attempted to make urgent recommendations and predictions based on scarce information about the underlying biology of the virus, its main mechanisms of spread, severity and fatality rates [46].

Known unknowns during the emergence of COVID-19 presented extra challenges and limitations to the models [47], as the lack of individual-level data on mobility patterns to effect of NPIs on individual behaviour and the non-existence of population-based contact diaries data for countries like Brazil. Even the projected contact matrices that existed [122] were estimated as averages per country, whereas we were making simulations at municipality level (which is an important limitation in a country with such extensive sociodemographic heterogeneities as Brazil). Considering these extra layers of uncertainty to model predictions, our job was to make the most out of the information we had in hand to create new methodologies and communicate results and uncertainties in a fair way.

Although previous studies considered heterogeneous contact structures in epidemiological models [104; 129; 138], and many others accounted for the effects of NPIs on transmissibility [49; 65; 67; 4; 5; 11; 116; 71], few to none of them made the connection between both phenomena. Strictly speaking, NPIs can affect the underlying network structure over time due to large-scale social distancing interventions. In this context, we identified a demand for more flexible and dynamic modelling approaches taking into account the non-linear phenomena emerging from social distancing NPIs.

In the case of the COVID-19 pandemic, governments enforced restrictions on movement and contacts, which essentially aimed to remove connections among individuals. To take into account the structural heterogeneities introduced by these social distancing measures on a population's contact network, we proposed an optional improvement to compartmental models that only depends on quantities already computed when modelling NPIs, without the addition of an excessive number of parameters which could lead to model over-fitting.

The effect of our proposed implementation of an NPI-induced percolation effect was clearly shown by the evaluation of the NPI-dependent basic reproduction number, which resulted in a distinctive discrepancy in the transmissibility when comparing model versions with and without percolation. We determined that for high combined NPI adherence, the model without percolation highly overestimates transmission rates. Using the AIC we found a $\Delta\text{AIC} = 191$, between the model versions with and without the percolation effect (Table 4.7), which is orders of magnitude higher than the conventional threshold of $\Delta AIC \leq 2$ to consider both models equally plausible [63]. That is, the model with percolation correction had support from the data analysed. It should be noted that the AIC takes into account biases caused by over-fitting [8]. Therefore, the stronger support of the percolation model cannot be attributed to a spurious effect caused by a few additional parameters.

One alternative solution to the data fitting difficulty would consist of adjusting other model parameters. However, with further exploration of the parameter space, even considering less realistic parameter sets (e.g. tampering with the coverage parameters of NPIs), we could not obtain a good fit to data. For instance, a 30% increase in NPI adherence in the standard model (red curve in Fig. 4.8) did not result in better model fitness, compared to the model version including percolation.

These results highlight the importance of implementing this non-linear response to NPIs in compartmental models to obtain a better representation of the effect of those interventions on a large population.

The proposed improvement to an age-structured SEIR compartmental model described here, although motivated by the need to better represent the dynamics of how COVID-19 spreads, was inspired by network theory and general individual level considerations that could be applied to any population whose underlying

network structure has been affected by similar fragmentation processes.

We implemented the effect of individuals' behavioural changes on a population's macroscopic dynamics without drastically increasing the number of fitted parameters in our compartmental model. Specifically, only two parameters were added to the equations governing the dynamic model and these were fitted to epidemiological data. Hence, the increased model fitness obtained would not imply additional efforts for decision-makers in terms of data collection.

The usefulness of incorporating this fragmentation process into our SARS-CoV-2 epidemiological compartmental model was evident, but it should also be noted that its flexible implementation also permits it to be applied when modelling the transmission dynamics of other communicable diseases under similarly high NPI coverage regimes.

Therefore, our framework may be applied to any compartmental models trying to represent the dynamics of a homogeneously mixed population suffering drastic changes in its connectivity patterns during an epidemic. This result contributes to a more accurate epidemic modelling, potentially implying better control and prevention policy recommendations at a public health level.

The description of the CoMo Consortium framework has been published by Aguas et al. [5] and the theoretical description of the model including percolation is under peer-review and has been made available as a preprint by Franco et al. [69].

5

Further applications and communication of mathematical models for COVID-19

"The single biggest problem in communication is the illusion that it has taken place."

(George Bernard Shaw, 1856-1950)

Since the beginning of the pandemic, health policymakers around the world have struggled with the lack of clear data and definitive scientific evidence to support decision making. The pandemic not only overwhelmed health systems, but also changed how scientific evidence is shared and consumed by scientists, policymakers and the wider public.

In Brazil, the unprecedented national health crisis has also aggravated previously existing health structure and health system inequities in the country. Political challenges also arouse, including uncoordinated actions, variability of public health response among the various states, delayed and insufficient implementation of non-pharmaceutical interventions (NPIs) and a lack of preparedness.

In this context, the *Observatório COVID-19* modelling group engaged in a varied modelling exercises to inform the public, the scientific community and health policy makers. We have worked directly with health secretaries, funerary services and larger policy facing scientific collaborations to inform decision making through mathematical modelling. In the following sections, we will present some of outcomes of these efforts and discuss the communication challenges involved with conveying their meaning to policymakers and the wider public.

5.1 Effects of different COVID-19 mitigation strategies in school settings

Among the various non-pharmaceutical interventions implemented in response to the COVID-19 pandemic in Brazil in 2020, school closure was among the first measures implemented throughout the 27 states of the country aiming to mitigate the disease spread. By the end of 2020, however, planning the best timing for school closure and reopening was a crucial debate, since the short and long-term impacts of prolonged suspension of in-person classes arose as a major concern. In this context, it was also relevant to come up with ways to assess the actual impact of school closure on the disease transmission dynamics, as well as its effectiveness when considered in association with other mitigation strategies, such as contact tracing.

To assess the impact of school reopening and contact tracing strategies upon decreasing COVID-19 cases and deaths, we have established a collaboration with the group lead by infectious diseases expert Prof. Cristiana Toscano (UFG) on a CNPq funded project entitled Modelling SARS-CoV-2 transmission dynamics in Brazil: Real time science to subside evidence based policy making. We have applied the modelling framework described in chapter 4 with an expanded compartmental structure. Mechanisms of contact tracing and quarantining where implemented to access their effect of such strategies in the epidemic dynamics upon school reopening (see the expanded model diagram in figure 5.1.

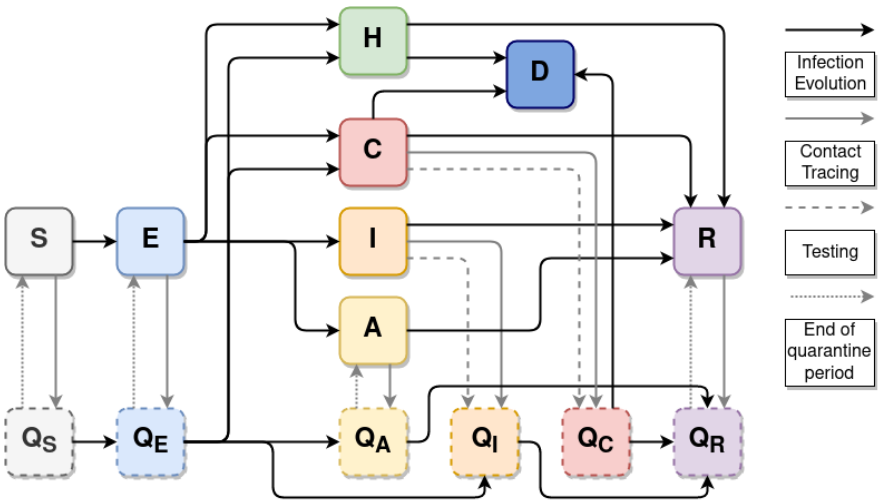


Figure 5.1: Diagram representing the model for school reopening and quarantining dynamics. H comprises all hospitalized compartments (H, ICU, ICUh), and C comprises all critical compartments (Hc, ICUc) who have not received attendance. Solid black arrows describe the infection pathways in the model, as in figure 4.3. Solid grey arrows describe individuals who are quarantined through contact tracing. Dashed grey arrows describe individuals who are isolated/quarantined because of positive testing in an asymptomatic individual. Dotted gray arrows describe the pathway of individuals after the quarantine period. Source: Borges et al. [19].

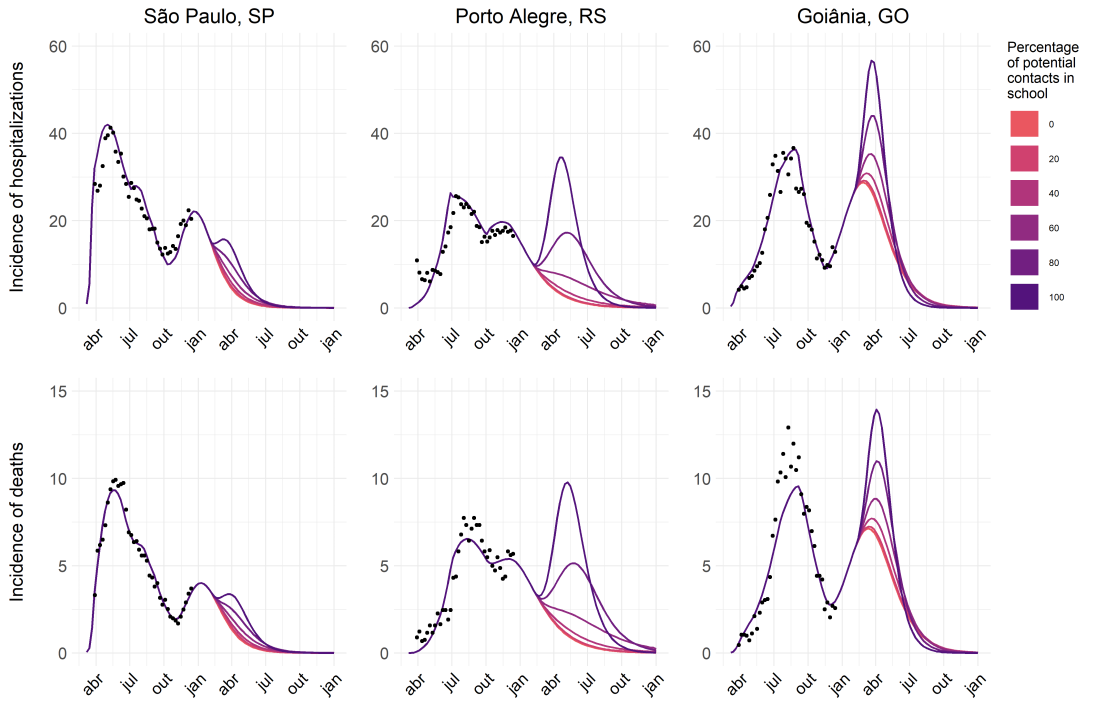


Figure 5.2: Epidemiological dynamics for different scenarios of increase in the percentage of potential contacts in school for three Brazilian capitals: São Paulo, Porto Alegre, and Goiânia. Colours represent the percentage of potential contacts of individuals in the school setting after reopening schools on February 1st, and dots represent the hospitalisation and death data from the SIVEP database. Source: Borges et al. [19].

After fitting epidemiological and demographic data, we have simulated scenarios with increasing school transmission due to school reopening in 3 large urban centres in Brazil: São Paulo, Porto Alegre and Goiânia (figure 5.2). Due to their different epidemiological environments (transmission patterns and timing for implementing mitigation strategies), the impact of school reopening in new cases and deaths varied among these cities. In settings with declining trends in Covid-19 incidence, small impacts were projected after school reopening (as observed in São Paulo), whereas in settings with increasing trends, the magnitude of the projected impact was significant (as projected in Goiânia).

Our model shows that schools reopening results in a non-linear increase of reported COVID-19 cases and deaths (figure 5.3, being highly dependent on infection and disease incidence at the time of reopening. While low rates of within-school transmission resulted in small effects on disease incidence (cases/100,000 pop), intermediate or high rates can severely impact disease trends resulting in escalating rates of new cases even if other interventions remain unchanged. When contact tracing and quarantining are restricted to school and home settings, a large number of daily tests is required to produce significant effects of reducing the total number

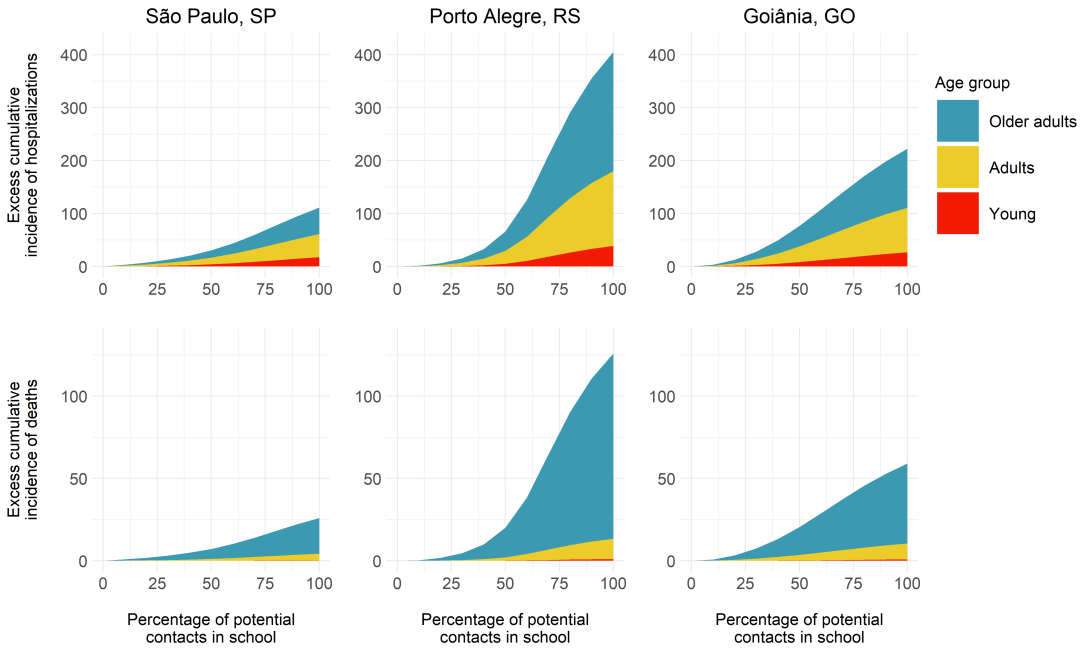


Figure 5.3: Effects of implementing case isolation, contact tracing, and quarantining contacts in schools on the cumulative incidence of Covid-19 cases and deaths, by the daily number of tests available and scenarios of school reopening (percentage of potential contacts in schools). The projected cumulative incidence of cases and deaths presented is for the period from school reopening (February 1st, 2021) to the end of 2021 (December 31, 2021). Source: Borges et al. [19].

of hospitalisations and deaths.

Our results suggest that policymakers should carefully consider the epidemiological context and timing regarding the implementation of school closure and return of in-person school activities. Also, although contact tracing strategies are essential to prevent new infections and outbreaks within school environments, our data suggest that they are alone not sufficient to avoid significant impacts on community transmission in the context of school reopening in settings with high and sustained transmission rates.

A more detailed description of the model and its results can be found as preprint in [19].

5.2 Simulating the burden of new SARS-CoV-2 variants of concern

The SARS-CoV-2 variant of concern (VOC) P.1 (Gamma variant) emerged in the Amazonas State, Brazil, in November 2020 [64; 109], having been detected in 36 countries, with local transmission in at least 5 of those. A range of mutations are

seen in P.1, ten of them in the spike protein. It shares mutations with VOCs previously detected in the United Kingdom (B.1.1.7, Alpha variant) and South Africa (B.1.351, Beta variant). At the beginning of 2021, the consequences of these mutations on SARS-CoV-2 P.1 pathogenicity were still uncertain.

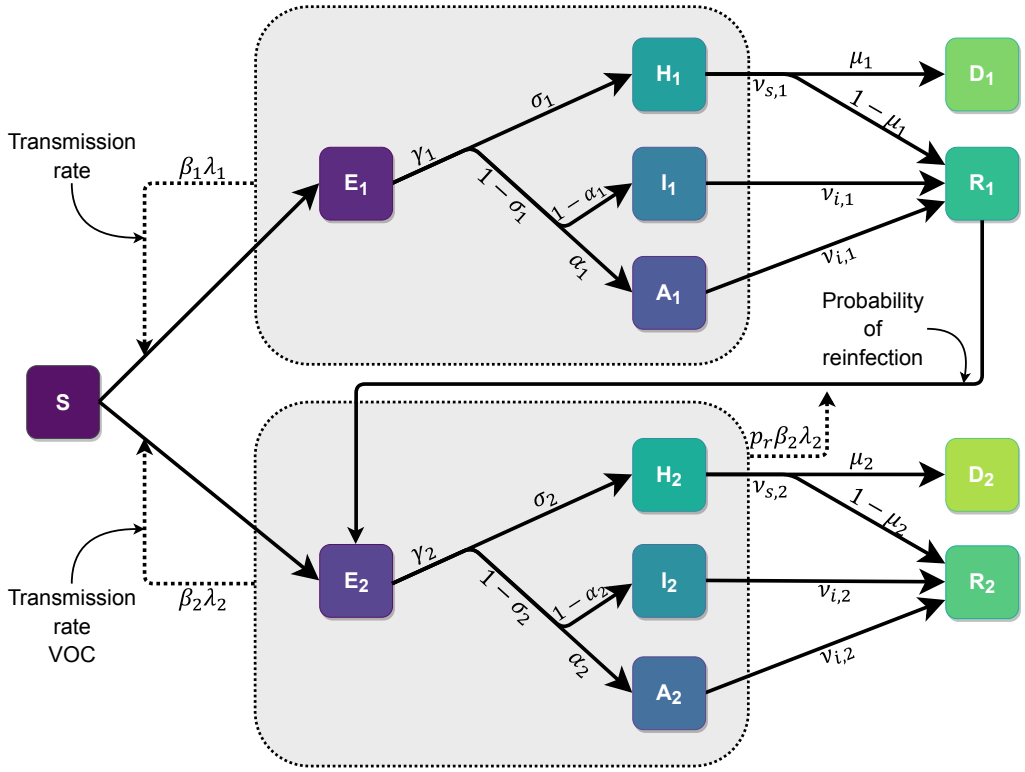


Figure 5.4: Diagram representing the model for SARS-CoV-2 P.1 dynamics. The model compartments and the respective connections between them are summarised in this diagram, and they are named as S: Susceptible, E: Exposed (pre-symptomatic), H: Hospitalized (severe infected individuals), I: Infected (symptomatic individuals, not hospitalized), A: Asymptomatic. D: Deceased, R: Recovered. Compartments are subdivided into 3 age categories, not represented here for simplicity. Compartments with subindex 1 represent the wild-type variant, subindex 2 refers to the VOC P.1. Continuous lines represent flux between each compartment; dashed lines, infection pathways. Small arrows indicate force of reinfection and transmissibility. λ =force of infection. β =relative transmission rate. p_r =relative force of reinfection. γ =average time between being infectious and presenting symptoms. σ =proportion of severe cases that require hospitalization. α =proportion of asymptomatic cases. ν_s =average time between being infectious and recovering for severe cases. ν_i =average time between being infectious and recovering for mild/asymptomatic cases. μ =in hospital mortality ratio. Source: Coutinho et al. [44].

We have analyzed Brazilian national health surveillance data on COVID-19 hospitalizations [22] and the frequency of P.1 among sequences from residents of Manaus city [64; 125] using an extended SEIR compartmental model (figure 5.4 to estimate the *relative transmissibility* in comparison to the previous local variant(s), and *relative force of reinfection* of the P.1 variant, *i.e.*, the ratio between the force of infection by P.1 on previously infected individuals (reinfections) and the force of

infection by P.1 on susceptible ones (new infections). Fitting the surveillance data to our model, we were able to estimate that the P.1 variant is about 2.6 times more transmissible (95% Confidence Interval: 2.4–2.8) than previous circulating variant(s), and 28% of Manaus cases in the period were due to reinfections.

Our estimates ranked P.1 as one of the most transmissible among the SARS-CoV-2 VOCs identified, and potentially as transmissible as the posteriorly detected VOC B.1.617.2 (Delta variant), posing a serious threat and requiring measures to control its global spread.

The full model description have been published by Coutinho et al. [44] and a thorough discussion of its implications on health policy making is presented in Marquitti et al. [100].

5.3 Optimizing COVID-19 vaccines roll-out in a limited supply context

While rising cases and new variants were still an ongoing concern, by the end of 2020, COVID-19 vaccines started to become available around the world. Nevertheless, production and distribution of doses was a major concern, given the high demand scenario. Hence, optimal vaccination strategies were an important topic for investigation.

We used a different SEIR mathematical model, with simplified age classification, compared to the model in chapter 4, but with the addition of a two-dose vaccination schedule, including the between-doses delay modelled through delay differential equations and linear optimisation of vaccination rates (figure 5.5). Simulations for each time window and for different types of vaccines and production rates were run to find the optimal time window between doses, which we define as the one that minimises the number of deaths.

We found that the best strategy depends on an interplay between the vaccine production rate and the relative efficacy of the first dose. In the scenario of low first-dose efficacy, it was always better to apply the second dose as soon as possible, while for high first-dose efficacy, the optimal window depends on the production rate and also on second-dose efficacy provided by each type of vaccine. We also found that the rate of spread of the infection does not affect significantly the thresholds of the optimal window, but is an important factor in the absolute number of total deaths. These conclusions point to the need to carefully take into account both vaccine characteristics and roll-out speed to optimize the outcome of vaccination strategies.

The full model description and detailed results have been published in Souto Ferreira et al. [137].

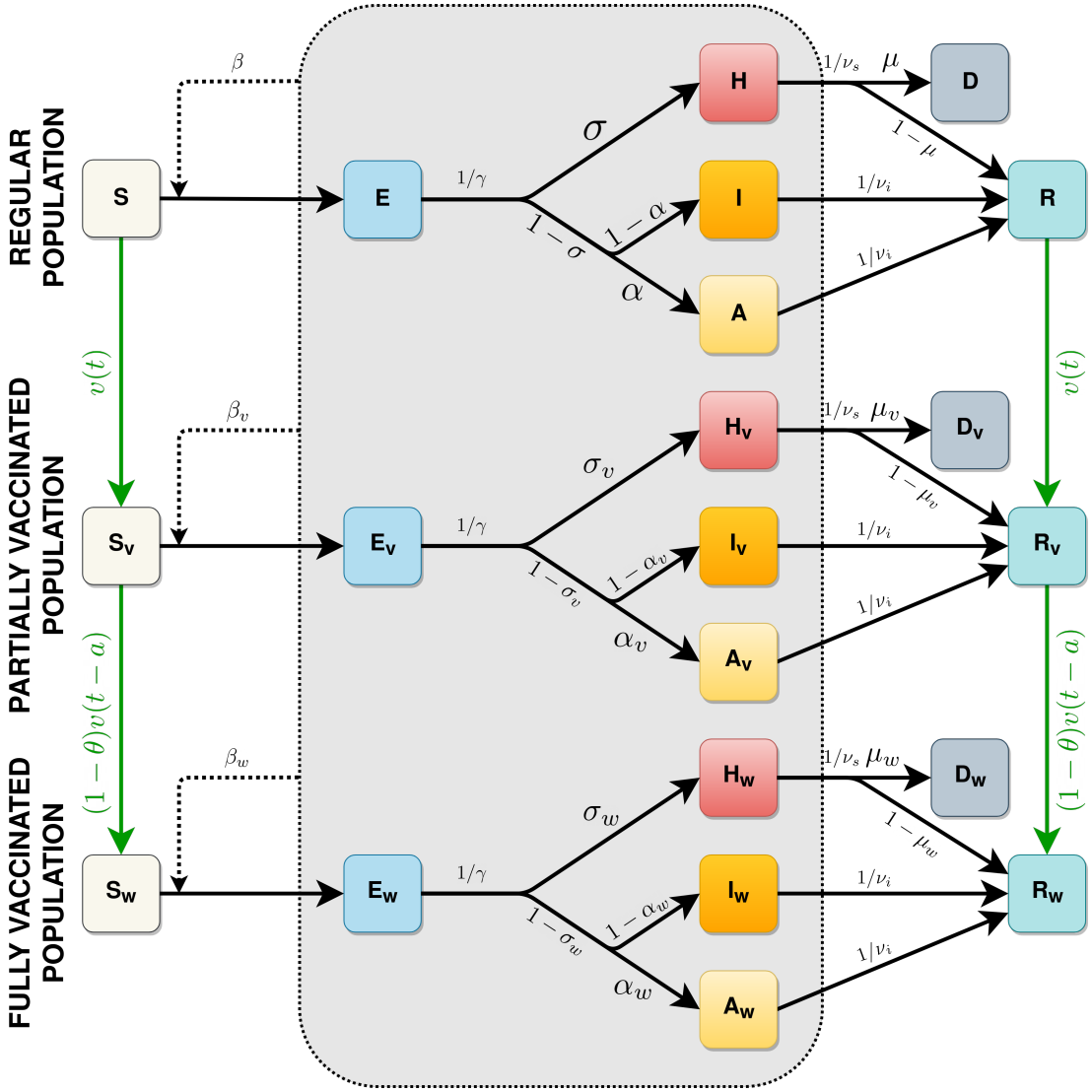


Figure 5.5: Diagram representing the vaccination model structure. Subscripts v and w indicate the first- and second-dose vaccinated classes, respectively. Black arrows indicate transitions between epidemiological stages, green arrows indicate vaccination. All classes pictured inside the grey box are infectious. Because epidemiological progressions happen at time-scales shorter than those related to vaccine effects, infectious classes are not vaccinated in the model. Source: Souto Ferreira et al. [137].

5.4 Communicating model results

The use of mathematical models to compare intervention scenarios, understand aspects of a disease transmission dynamics or optimise the implementation of interventions can be seen as a great way to inform decision making, allowing for evidence-based policies. Policy decisions that do not incorporate current knowledge are invariably weakened [154], so why not always use mathematical models as policy making tools?

Conveying relevant scientific results into policy recommendations is essential. Nevertheless, it can be challenging either due to timing or underlying uncertainties. Rigorous analysis might be produced in a slower pace compared to the demanded by policymakers, specially during an emergency situation like the COVID-19 pandemic. Additionally, in the case of mathematical modelling, communicating the lack of evidence that results either in wide statistical confidence intervals (e.g. when predicting case figures or parameter values) or analysis with only qualitative results (e.g. scenario comparison) is a challenge by itself [47].

With the mathematical models presented in this chapter, we tried to address burning issues regarding to the evolution and spread of SARS-CoV-2 pandemic in the most rigorous, yet timely, way possible. The scientific news press frequently provided a good media through which scientifically accurate synthesis were produced [39; 99; 60; 59], as well as some dedicated newspaper columns [70]. Notwithstanding, we have always encountered challenges to effectively communicate the impacts of our findings to policy makers.

The issue might be two-fold: in one hand, academics need to be prepared to translate their rigorous original papers into well synthesised policy briefs; in the other hand, policy makers need to become more open to evidence-based decision making. This might required training, actions towards integrating research and governmental institutions, and a cultural transformation in both sides, but it would be a step towards a society better prepared for the health emergenciess of the future. As the medical historian Mark Honigsbaum wrote on his book *The Pandemic Century*: “The only thing that is certain is that there will be new plagues and new pandemics. It is not a question of if, but when.” [80].

Part III

Modelling towards elimination of an endemic disease

6 The malaria endemic in Brazilian Amazon

Malaria is an acute febrile-illness disease that produced an estimated 229 million cases and 409,000 deaths worldwide in 2018 [158]. It is caused by five different species of Plasmodium parasites (Plasmodium falciparum, Plasmodium malariae, Plasmodium vivax, Plasmodium ovale and Plasmodium knowlesi) [35], of which at least four are present in the Americas [114]. The malaria parasites are transmitted to people through the bites of infected female Anopheles mosquitoes, with P. falciparum and P. vivax species posing the greatest threat worldwide.

Following The World Health Organization World malaria report 2020 [158], in 2019 nearly half of the world's population was at risk of malaria. Most malaria cases and deaths occurred in sub-Saharan Africa. However, South-East Asia, Eastern Mediterranean, Western Pacific, and the Americas were also at risk. P. vivax is the predominant parasite in the American continent, representing 75% of malaria cases, though P. falciparum and P. malariae are also consistently present.

In Brazil, 90% of malaria is caused by P. vivax and most of the other cases are caused by P. falciparum [115; 155]. Currently, more than 99% of the Brazilian malaria cases are concentrated within the Amazon region [135], where climatic conditions and lack of strategic allocation of resources perpetuates the challenge to move from an ongoing endemic state to an elimination pathway.

6.1 Malaria epidemiological cycle

The life cycle of malaria parasites involves humans and the female *Anopheles* mosquitoes as hosts. The mosquitoes lay their eggs in a variety of bodies of water, having their larvae hatching rate and longevity dependent on climate variables, such as rainfall, temperature and humidity.

The within-host cycle can be slightly different among *Plasmodium* species. Namely, *P. vivax*, the most common species in Brazil, has an additional parasite dormant

stage (hypnozoite) that can persist in a patients liver and causes relapses if untreated. Also, *P. vivax* tends to cause less severe disease and deaths, compared to *P. falciparum*, which is predominant in Sub-Saharan Africa. The diagram in figure

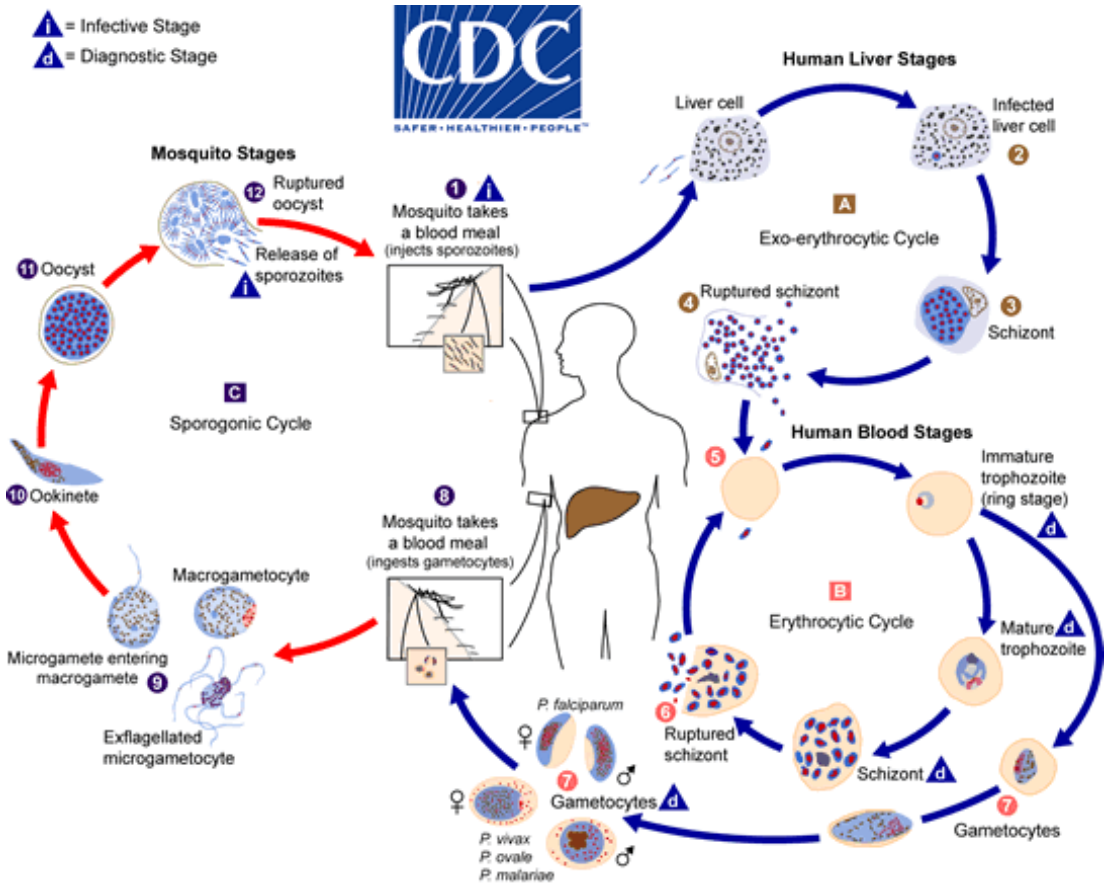


Figure 6.1: The malaria parasite life cycle involves two hosts. During a blood meal, a malaria-infected female Anopheles mosquito inoculates sporozoites into the human host (1) . Sporozoites infect liver cells (2) and mature into schizonts (3), which rupture and release merozoites (4). (Of note, in P. vivax and P. ovale a dormant stage [hypnozoites] can persist in the liver (if untreated) and cause relapses by invading the bloodstream weeks, or even years later.) After this initial replication in the liver (exo-erythrocytic schizogony (A)), the parasites undergo asexual multiplication in the erythrocytes (erythrocytic schizogony (B)). Merozoites infect red blood cells (5). The ring stage trophozoites mature into schizonts, which rupture releasing merozoites (5). Some parasites differentiate into sexual erythrocytic stages (gametocytes) (7). Blood stage parasites are responsible for the clinical manifestations of the disease. The gametocytes, male (microgametocytes) and female (macrogametocytes), are ingested by an Anopheles mosquito during a blood meal (8). The parasites' multiplication in the mosquito is known as the sporogonic cycle (C). While in the mosquito's stomach, the microgametes penetrate the macrogametes generating zygotes (9). The zygotes in turn become motile and elongated (ookinetes) (1) which invade the midgut wall of the mosquito where they develop into oocysts (11). The oocysts grow, rupture, and release sporozoites (12), which make their way to the mosquito's salivary glands. Inoculation of the sporozoites (1) into a new human host perpetuates the malaria life cycle. (Figure and caption reproduced from CDC [35])

7.1, reproduced from the Center for Disease Control and Prevention (CDC) website [35], explains the parasite dual-host life cycle in detail.

6.2 Current treatment protocol

The World Health Organization recommends a single dose of primaquine as a treatment for *P. falciparum*. For *P. vivax*, radical control is recommended with 7 (or 14) days of chloroquine + primaquine [155].

In Brazil, the treatment for confirmed non-complicated *P. vivax* infections consists in a 7 day therapy combining cloroquine and primaquine. In the case of confirmed *P. falciparum*, a 6 days combination therapy is recommended using quinine, doxicicline and primaquine. For mixed infections (*P. falciparum* and *P. vivax*), mefloquine and primaquine are recommended, durinf a 7 days therapy. The recommendation for *P. malariae* is a 3 day cloroquine therapy [30].

6.3 Malaria in Brazil

The history of malaria in Brazil began in the XVI century, with the introduction of *Plasmodium falciparum* by African slaves, which were brought to Brazil by European colonizers [76]. Throughout the colonial period, slaves, Amerindians and colonizers became malarious. Yet the biggest drivers for the spread of malaria throughout the country were the rubber industry and railroad construction, responsible for an increasing flux of migrants to the Amazon region in the XIX century.

With the abolition of slavery in 1888 and consequent interruption of agricultural activities in São Paulo (SP) and Rio de Janeiro (RJ), many plains were left susceptible as mosquito breeding sites, introducing highly malarious areas in the Southeast region. By 1891, the increasing number of malaria cases emerged as an important national issue and the first anti-malaria commissions were created.

At 1903 the mosquito *Anopheles cruzi* was pinpointed as the disease vector by the Adolpho Lutz, from the Bacteriological Institute of São Paulo. Though, people would still take some time before acknowledging that malaria was not an "infection of the home", therefore effective vector control would only start in the 1940s with the creation of the Northeastern Malaria Service and the National Malaria Service (SNM).

Carlos Chagas, who worked with Oswaldo Cruz, was assigned to deal with highly malarious regions in the 1900s and applied "offensive" prevention methods such as killing mosquitoes and larvae breeding sites home by home, regularly administering quinine to workers and advising contractors to fire people who would not take the medicine or were chronically infected.

Many attempts to establish a nationwide malaria control program occurred throughout the XX century with more or less success rates. In the early years, malaria was spread all over the country but by the end of the XX century, it was mainly restricted to the Amazon region. One of the biggest efforts to control the epidemic was adopted by the Malaria Eradication Campaign (CEM), guided by World Health Organization (WHO) recommendations: DDT spraying and distribution of antimalarial were apllied widely. Though it was effective at first, the wide distribution of chloroquinine (CQ) rapidly culminated in *P. falciparum* resistance (*P. vivax* would not show resistance before 2007).

Other drugs were sought as potential substitutions to CQ. Sulphadoxine pyrimethamine started on trials on the 1960s but by 1972 resistance had already been reported and by 1989 90% of parasites were sulfadoxine-pyrimethamine-resistant. At this point, other options were quinine-tetracycline followed by quinine + doxycycline and mefloquine + primaquine as secondary drugs.

With the introduction of the Program for the Intensification of Malaria Control (PIACM) in 1999, a more politically engaged program, *P. falciparum* malaria cases drastically decayed. Advances from other parts like the National Health Foundation (FUNASA) and Program for Malaria Prevention and Control (PNCM) also continued. By 2002 rapid detection methods were implemented and the diagnostic could be obtained within 24 h. By 2005 the Malaria Epidemiological Surveillance Information System was created.

More recently, the Amazon still accounts for more than 99% of the Brazilian malaria cases. By 2010, the Global Fund funded a Ministry of Health project to strengthen public health capacity in the Amazon region, which apparently led to some decrease in cases between 2010-2013 [155].

Table 6.1 shows a timeline of important events that affected the spread of malaria in Brazil, leading to it current endemic state. Right now, the challenge is to continue the efforts towards elimination, perhaps focusing in more effective ways to deliver treatment and prophylaxis.

6.4 Conclusion

Strategic decisions informed by mathematical modelling and data analysis have been proved to be a valuable tool to overcome such challenging predicaments [133; 139; 5; 100; 44]. There have been many successful endeavours to inform health policy making using mathematical models in Africa [133], which carries 94% of the current malaria burden [142; 135], and Asia, where antimalarial drug resistance is a recurring issue [139; 7].
In spite of that, there haven't been many efforts directed towards using mathematical analysis to inform health policy makers in the Americas, where

765,000 cases and 340 deaths were reported in 2018 [114]. These relevant numbers show the importance and urgency of the development of effective modelling tools, built with the aim to communicate with and support policy making, not only in the Brazilian Amazon, but throughout all malaria affected regions in the Americas.

Elimination of this lethal disease is a realistic goal for the region, but only if multiple interventions are combined to address the various *Plasmodium* species that infect humans. The most recent modelling activity commissioned by the Global Fund to Fight Aids, Tuberculosis and Malaria focuses on the Guyana Shield region, which encompasses Suriname, Guyana, French Guiana, Venezuela, and the Brazilian states of Amapá, Pará, and Roraima.

In the following chapters, we propose the use of mathematical models to simulate the malaria dynamics both from a theoretical and an applied point of view. In chapter 7 we propose two independent model structures for *P. falciparum* and *P. vivax* parametrising them consistently with the malaria dynamics in Brazilian Amazon. In chapter 8, we develop a theoretical framework to allow interspecific interactions to be taken into account in multi-species models, giving examples of its applications in modelling malaria elimination strategies in Asia-Pacific and Guyana Shield regions.

Table 6.1: Timeline for malaria in Brazil. Summarised from from Griffing et al. [76].

1500	Colonisation.
1560	**Possible introduction of *P. falciparum* by African slaves**.
1587	1^{st} **report: Tupinamba Indians**.
1610	Possible introduction of *P. vivax* from an European strain.
1815	End of colonial period.
1870	**Rubber industry railroad construction / great flux of immigrants to Amazon**.
1888	Abolition of slavery / SP and RJ with highly malarious regions.
1891	First anti-malaria commissions.
1900	At this point, half of the population (6 million) was malarious.
1903	*An. cruzi* identified as a vector by Adolpho Lutz.
1905	Carlos Chagas surveyed Itatinga-SP and started "offensive" prevention methods: gave quinine to workers, killed mosquitoes and larvae.
1907	**First report of quinine resistance**.
1910	House screens and bed nets started being used.
1912	Oswaldo Cruz creates a plan to sanitizing Amazon region but finds many challenges due to the scattered population.
1920	National Department of Public Health (DNSP) was created to deal with Chagas disease, hookworn and malaria.
1928	Introduction of the African vector *An. gambiae* and large outbreak of malaria in the Northeastern region.
1939	**Creation of the Northeastern Malaria Service: fumigation of post on outgoing roads, homes, boats and planes were sprayed with insecticide**.
1940	Cases begin to diminish / More than 50% were outside Amazon.
1941	The National Malaria Service (SNM) is created and starts fighting "Bromeliad malaria".
1942	Foundation of the Special Public Health Service in Amazon (SESP), created to provide sanitation and provide training for health staff.
1945	**Introduction of DDT spraying as profilaxy and chloroquine (CQ) tablets as treatment**.
1957	**In response to the WHO establishment of the Global Eradication of Malaria, SNM changed to the Malaria Eradication Campaign (CEM), which decided to put CQ in table salt in areas where DDT was difficult to apply (which migh have encouraged early parasite resistance)**.

1960 CEM sprays SP houses with DDT every six months, tested population every year and treated patients / Sulphadoxine pyrimethamine starts being used in trials.

1965 **The government adopts the CEM model recommended by the WHO: DDT spraying and antimalarial distribution**.

1968 68% of the population lived in a malaria-free zone, with 31% living in the consolidation phase / Mild resistance to pyrimethamine alone is reported.

1970 CEM was suspended due to the fewer cases and concerns over DDT / Malaria had been reduced to 1% of its incidence in 1950.

1970 The Superintendent of Public Campaigns of Brasil (SUCAM) integrated the malaria eradication program.

1972 **First report for sulfadoxine-pyrimethamine resistance**.

1974 All Brazilian malaria regions had been through the attack phase.

1975 **DDT resistance reported in Brazil and Colombia**.

1976 Malaria under control in S, SE, NE ans Central-W, but transmission was increasing due to the flux to Amazon.

1980 Due to Amazon forests deforestation, - encouraged by Agrarian reform, agriculture, construction of roads and hydroelectric plants - the population in Amazon Region rose and 96% of all cases occurred there.

1980 **SUCAM focused efforts on Amazonian malaria control with new techniques: outdoor ultra-low-volume nebulisation, mass treatment, impregnated curtains ans new insecticides / Cases are still increasing due to migration between regions / CQ resistance is reported throughout Brazil**.

1989 The Amazon Basin Malaria Control Project (PCMAM) is created / 90% of parasites were Sulfadoxine-pyrimethamine resistant / quinine-tetracycline starts being used, followed by quinine + doxycycline and mefloquine + primaquine as a secondary drug.

1996 With depleted funding of PCMAM, malaria resurges and malaria control plans are now administered by the National Health Foundation (FUNASA).

1999 Introduction of the Program for the Intensification of Malaria Control (PIACM).

2001 By 2001, cases were almost halved. *P. falciparum* cases were reduced by 35% and *P. vivax* by 41%.

2002 FUNASA implements rapid detection methods where a case could be diagnosed within 24h.

2005 Malaria control is undertaken by the Program for Malaria Prevention and Control (PNCM), which created the Malaria Epidemiological Surveillance Information System.

2007 ***P. vivax* start showing resistance to CQ in Manaus**.

2010 Global Fund funds a Ministry of Health Project to strengthen public health capacity in the Amazon.

7

Single-species models for malaria in Brazilian Amazon

Here we illustrate how mechanistic models can be used to study the effect of the available interventions on the malaria dynamics, based on historic information, clinical literature and data. We present two separate compartmental models describing *Plasmodium vivax* and *Plasmodium falciparum* dynamics, with relevant interventions being implemented, namely, early detection and treatment (EDAT) and vector control (VC). We parametrise both models to represent the dynamics in Brazilian Amazon, based on available incidence data and the timeline of implementation of interventions.

7.1 Single-species malaria models

7.1.1 Model for *P. falciparum*

For the baseline model structure, we have adopted an extended SEIRS model structure with demography, including sub-populations of individuals that are susceptible (S); infectious and clinically ill (I_C); infectious and asymptomatic (I_A); infectious, asymptomatic and undetectable (by rapid tests or blood smear microscopy) (I_U); and recovered and (partially) immune (R).

Natural recovery occurs as symptomatic individuals become asymptomatic (at rate ν_C), followed by undetectable (at rate ν_A) and then recovered (at rate ν_U). We assume that immunity can be lost at a rate ω and individuals can be re-infected even while infectious (super-infection) or recovered, with lower re-infection probability in the latter case.

Finally, the total population P is assumed constant, as individuals die and emigrate at the rate μ_{OUT} and are substituted at the same rate $\mu + \mu_C + \mu_A + \mu_U = \mu_{OUT}$ due

to new births or immigration.

As transmission depends on mosquito population dynamics, the *baseline transmission rate*, $\tilde{\beta}$, is evaluated as

$$\tilde{\beta} = Seas(t)\frac{b\epsilon_h b_h \epsilon_m}{P\left(b_h\epsilon_h + \delta_m\right)\left(\frac{\gamma_m}{\gamma_m+\delta_m}\right)} \tag{7.1}$$

where b is the per mosquito biting rate, ϵ_h is the human per bite probability of infection, b_h is the per human biting rate, ϵ_m is the mosquito per bite probability of infection, δ_m is the mosquito death rate and γ_m is the mosquito transition rate from latent to infectious. Note that the factor $1/\left(b_h\epsilon_h + \delta_m\right)$ represents the duration of a mosquito infectious period and $\frac{\gamma_m}{\gamma_m+\delta_m}$ accounts for the probability for each mosquito to survive through its latent stage, before becoming infectious. Finally, $Seas(t)$ is an optional seasonality factor with amplitude α and phase ϕ written as

$$Seas(t) = \left(1 + \alpha\cos(2\pi t - \phi)\right). \tag{7.2}$$

Note that, with the above definition of $\tilde{\beta}$, we are assuming the mosquito population dynamics has reached a asymptotic stationary state, with oscillations exclusively caused by environmental seasonality.

Treatment for *P. falciparum*

To implement treatment as an intervention in the model, we assume that all symptomatic cases are eventually tended, though treatment seeking might be delayed (which is reasonable for the Brazilian Amazon region, where treatment is free and fairly easily accessible). The number of untreated people decays exponentially with time at the rate ω_T, which also defines the rate at which people seek treatment. After receiving treatment, people can comply or not. The adherent fraction is considered to be protected from new infections for a limited period.

This mechanism is taken into account by adding two extra compartments in the model: sub-population of individuals under treatment (T) and recovered but still under the effect of antimalarial drugs (hence completely immune) (R_D).

For both species of *Plasmodium*, we have considered that the business-as-usual treatment protocol, as described in section 6.2, was applied.

Figure 7.1 depicts the model compartments and flows among them, including the baseline dynamics and treatment.

Vector control

The vector control interventions most commonly adopted against malaria in the Brazilian Amazon region are indoor residual spraying (IRS) and insecticide treated

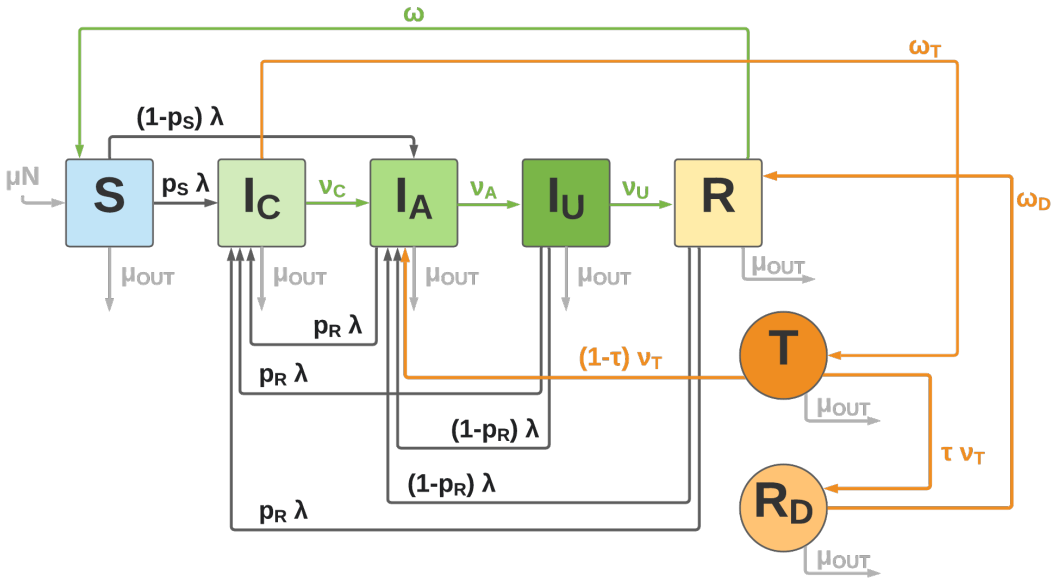


Figure 7.1: *P. falciparum* malaria model diagram. The variables in the compartments represent individuals that are S: susceptible, I_C: infectious (symptomatic), I_A: asymptomatic infectious, I_U: asymptomatic infectious and undetectable, R: recovered and immune, T: under treatment, and R_D: recovered after treatment and still under the effect of drugs. Black lines represent infection, green lines represent natural recovery, orange lines represent flows related to treatment interventions and grey lines represent demographic flows (births, migration and non disease-related deaths).

nets (ITN). These intervention is taken into account as a reduction of the effective contact rate, as previously defined in equation 7.1, lead by a decrease in the total mosquito population.

Considering IRS and ITN are applied, β can be written as $\beta = \tilde{\beta} \prod_i [1 - \epsilon_i c_i]$, or

$$\beta = Seas(t) \frac{b\epsilon_h b_h \epsilon_m}{P\left(b_h \epsilon_h + \delta_m\right)\left(\frac{\gamma_m}{\gamma_m + \delta_m}\right)} [1 - \epsilon_{IRS} c_{IRS}][1 - \epsilon_{ITN} c_{ITN}]\,, \tag{7.3}$$

where ϵ_i and c_i represent, respectively, effectiveness and coverage for intervention i ($i = \{IRS,\ ITN\}$).

The following equations represent the full *P. falciparum* model, including treatment and vector control intervention dynamics.

$$\frac{dS}{dt} = \mu P - \mu_{OUT} S + \omega R - \lambda S \tag{7.4}$$

$$\frac{dI_C}{dt} = \mu_C P - \mu_{OUT} I_C + p_S \lambda S + p_R \lambda R + p_R \lambda I_U + p_R \lambda I_A - \nu_C I_C - \omega_T I_C \tag{7.5}$$

$$\frac{dI_A}{dt} = \mu_A P - \mu_{OUT} I_A + (1 - p_S)\lambda S + (1 - p_R)\lambda R + (1 - p_R)\lambda I_U - p_R \lambda I_A + \nu_C I_C - \nu_A I_A + (1 - \tau)\nu_T T \tag{7.6}$$

$$\frac{dI_U}{dt} = \mu_U P - \mu_{OUT} I_U - \lambda I_U - \nu_U I_U + \nu_A I_A \tag{7.7}$$

$$\frac{dR}{dt} = -\mu_{OUT} R - \omega R - \lambda R + \nu_U I_U + \omega_D R_D \tag{7.8}$$

$$\frac{dT}{dt} = -\mu_{OUT} T - \nu_T T + \omega_T I_C \tag{7.9}$$

$$\frac{dR_D}{dt} = -\mu_{OUT} R_D + \tau \nu_T T - \omega_D R_D \tag{7.10}$$

where the total population $P = S + I_C + I_A + I_U + R + T + R_D$ is constant and the system's force of infection is written

$$\lambda = \frac{\beta}{P}(I_C + \rho_A I_A + \rho_U I_U) \tag{7.11}$$

with ρ_A and ρ_U being the relative infectivity i=of asymptomatic and undetectable individuals, respectively. The transmission rate, β, is defined as in equation 7.3.

Note that treatment related terms are represented by the orange terms in the system. Compliance and delayed treatment are taken into account.

7.1.2 Model for *P. vivax*

A very similar model structure can be adopted to model *P. vivax* dynamics, with the addition of a hypnozoite dormant state compartment (H). This means that an infected asymptomatic individual can either recover (with probability $1 - h$) or go into a dormant liver stage (with probability h), which is asymptomatic but can be both re-infected (rate λ) and suffer relapses (rate ω_R), if the liver parasites reach bloodstream. Figure 7.2 depicts P. vivax model structure and the flows between compartments.

The full set of equation for this model is presented below, where terms related to the latent hypnozoite liver state, which arises from an unsuccessful natural recovery, are included in blue:

$$\frac{dS}{dt} = \mu P - \mu_{OUT} S + \omega R - \lambda S \tag{7.12}$$

$$\frac{dI_C}{dt} = \mu_C P - \mu_{OUT} I_C + p_S \lambda S + p_R \lambda R + p_R \lambda I_U + p_R \lambda I_A - \nu_C I_C - \omega_T I_C + p_H(\omega_R + \lambda) H \tag{7.13}$$

$$\frac{dI_A}{dt} = \mu_A P - \mu_{OUT} I_A + (1 - p_S)\lambda S + (1 - p_R)\lambda R + (1 - p_R)\lambda I_U - p_R \lambda I_A + \nu_C I_C - \nu_A I_A + (1 - p_H)(\omega_R + \lambda) H \tag{7.14}$$

$$\frac{dI_U}{dt} = \mu_U P - \mu_{OUT} I_U - \lambda I_U - \nu_U I_U + \nu_A I_A \tag{7.15}$$

$$\frac{dR}{dt} = -\mu_{OUT} R - \omega R - \lambda R + (1 - h)\nu_U I_U + \omega_D R_D + (1 - \tau)(1 - h)\nu_T T \tag{7.16}$$

$$\frac{dT}{dt} = -\mu_{OUT} T - \nu_T T + \omega_T I_C \tag{7.17}$$

$$\frac{dR_D}{dt} = -\mu_{OUT} R_D + \tau \nu_T T - \omega_D R_D \tag{7.18}$$

$$\frac{dH}{dt} = \mu_H P - \mu_{OUT} H + h \nu_U I_U - (\omega_R + \lambda) H + (1 - \tau) h \nu_T T \tag{7.19}$$

where again the total population $P = S + I_C + I_A + I_U + R + T + R_D + H$ is constant and the system's force of infection is

$$\lambda = \beta \left(I_C + \rho_A I_A + \rho_U I_U \right) / P \,, \tag{7.20}$$

with the equation for beta is given by 7.3.

7.2 Data sources and parametrisation

In Chapter 2 we have introduced the Brazilian data bases that can be used for malaria case data acquisition and some of the work behind data processing and aggregation [68]. Here we used historical time-series, together with parameters values informed by clinical literature, vector control and treatment seeking data available in World Health Organisation reports [155; 158]. Values and sources for parameter values are included in table 7.1.

7.2.1 Single-species model results

Vector control for *P. falciparum*

To qualitatively illustrate the resulting model dynamics, simulations were run using the parameters specified in table 7.1 for the *P. falciparum* malaria model (equations

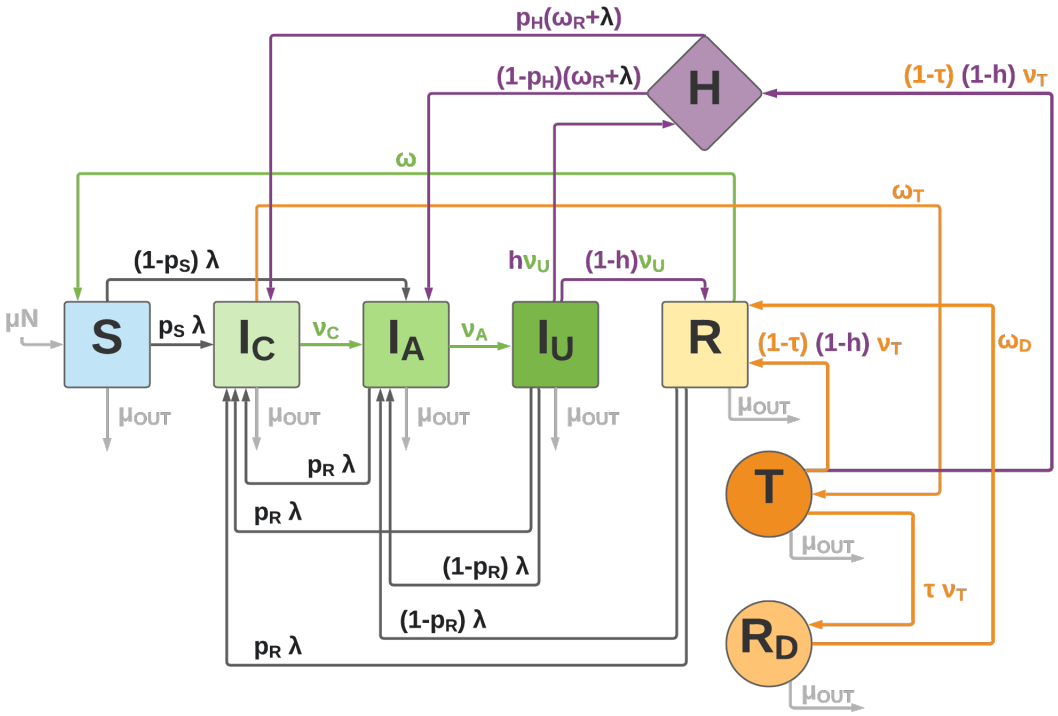


Figure 7.2: *P. vivax* malaria model diagram. The variables in the compartments represent individuals that are S: susceptible, I_C: infectious (symptomatic), I_A: asymptomatic infectious, I_U: asymptomatic infectious and undetectable, R: recovered and immune, T: under treatment, R_D: recovered after treatment and still under the effect of drugs, and H: hypnozoite latent. Black lines represent infection, green lines represent natural recovery, orange lines represent flows related to treatment interventions, grey lines represent demographic flows (births, migration and non disease-related deaths) and purple lines represent relapses and flows affected by the existence of hypnozoites.

7.4-7.10) with and without vector control interventions. No seasonality was included, for simplicity, since it does not affect the dynamics and here we are just going to focus on the qualitative behaviour of the system under this different intervention scenarios.

As expected for a SEIRS-like model with demographic dynamics, an endemic state is reached in a hypothetical scenario where no interventions are applied (figure 7.3).

The effect of the introducing a continuous and highly effective vector control intervention can be seen in figure 7.4, where it is clear that, being an intervention that directly affects the system's transmission rate, it can be very successful if applied widely.

The qualitative behaviour of the P. vivax malaria model (equations 7.12-7.19) is equivalent, since it is still based on a SEIRS-like structure (H can be seen as an extra recovered compartment), and the vector control interventions affect the force of infection in the exact same way.

Table 7.1: Parameter values and sources for *P. falciparum* and *P. vivax* malaria models.

Symbol	Definition	Value	Range	Units	Source
μ	net birth and death rate	1/50		$year^{-1}$	[156]
μ_{OUT}	rate of death + emigration	1/50		$year^{-1}$	*
ω	rate of immunity loss	1/2		$year^{-1}$	†
b	per mosquito biting rate	365/3		$year^{-1}$	†
b_h	per human biting rate	$11/(1+\alpha)$		$year^{-1}$	†
ϵ_h	human per bite probability of infection	23/100	0-1	-	†
ϵ_m	mosquito per bite probability of infection	50/100	0-1	-	†
δ_m	mosquito death rate	365/14		$year^{-1}$	[15; 86]
γ_m	mosquito transition rate from latent to infectious	365/10		$year^{-1}$	[102]
α	relative amplitude of seasonality	0	0-1	-	-
ϕ	phase angle of seasonality	π	0-1	rad	-
ϵ_{VC}	vector control effectiveness ($VC = IRS$ or ITN)	0.5	0-1	-	†
c_{VC}	coverage of vector control	0.7	0-1	-	[155]
ρ_A	relative infectivity of I_A (compared to I_C)	0.55	0-1	-	[136]
ρ_A	relative infectivity of I_U (compared to I_C)	0.17	0-1	-	[136]
μ_C	rate of importation of I_C cases	1/1000		$year^{-1}$	†
μ_A	rate of importation of I_A cases	1/1000		$year^{-1}$	†
μ_U	rate of importation of I_U cases	1/1000		$year^{-1}$	†
μ_H	rate of importation of I_H cases	1/1000		$year^{-1}$	†
p_S	proportion of all non-immune new I_C	90/100	0-1	-	[151]
p_R	proportion of all immune new I_C	20/100	0-1	-	[150]
ν_C	rate of relief of clinical symptoms in the absence of any treatment	365/3		$year^{-1}$	[93]
ν_A	rate of transition from I_A to I_U	365/60		$year^{-1}$	[151]
ν_U	rate of transition from I_U to R state	365/100		$year^{-1}$	[51]
ν_T	rate of transition from T to R_D state	P.f.: 365/1 P.v.: 365/7		$year^{-1}$	[155]
ϵ_T	coverage of EDAT	55	0-1	-	[135]
c_T	compliance to treatment	86	0-1	-	[135]
ω_T	rate of treatment seeking	365/10		$year^{-1}$	†
τ	adherence to treatment	86/100		-	[155]
ω_D	rate of loss of protection by drug	365/10		$year^{-1}$	†
h	proportion of non-treated cases becoming latent (H)	0.7	0-1		†
ω_R	relapse rate	12/8		$year^{-1}$	†
p_H	index of clinical recurrences	365/10		$year^{-1}$	†

* Calculated within the model.
† Assumed.

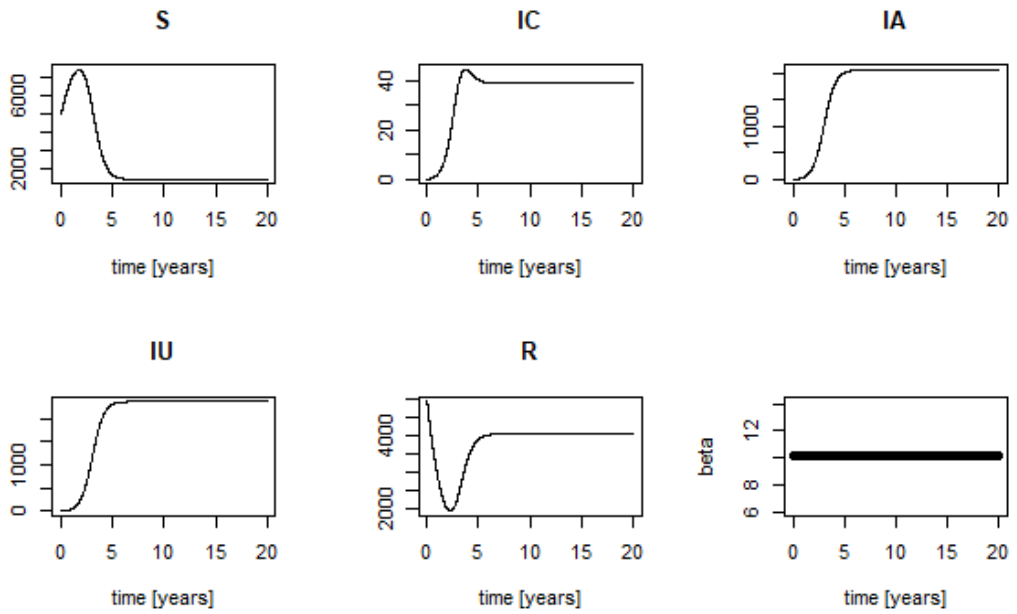


Figure 7.3: Baseline *P. falciparum* model dynamics showing that, if no interventions are made, the disease reaches a stable endemic state. In this case, β assumes its baseline constant value at all times.

Treatment for both species

Here we show the effects of different intervention scenarios on both P. falciparum and *P. vivax* model outputs. The *business as usual* scenario, as in figure 7.5 (a), is a parametric set that represents the invention history until present, and the maintenance of current strategies in the future. In scenarios (b) and (c), vector control is maintained at the same levels as in the *business as usual* scenario, but treatment compliance and coverage linearly decreases to 0% after 2017.

The importance of keeping of maintaining the current vector control and early detection and treatment (EDAT) interventions in imminent.

Treatment coverage and compliance

Here we present the outputs for the full model, which intrinsically correlates the *P. falciparum* and *P. vivax* models presented on subsections 7.1.1 and 7.1.2, respectively. It is assumed that the baseline parameters are shared between both systems while varying two key parameters in the treatment dynamics: coverage (related to ω_T) and compliance (related to τ).

The two-species dynamics outputs, as seen in figure 7.5, can be reproduced through the interactive application available at cfranco.shinyapps.io/malariabr/, where the model parameters can be adjusted to explore different scenarios.

Starting from values corresponding to the present situation in Amazonian malaria

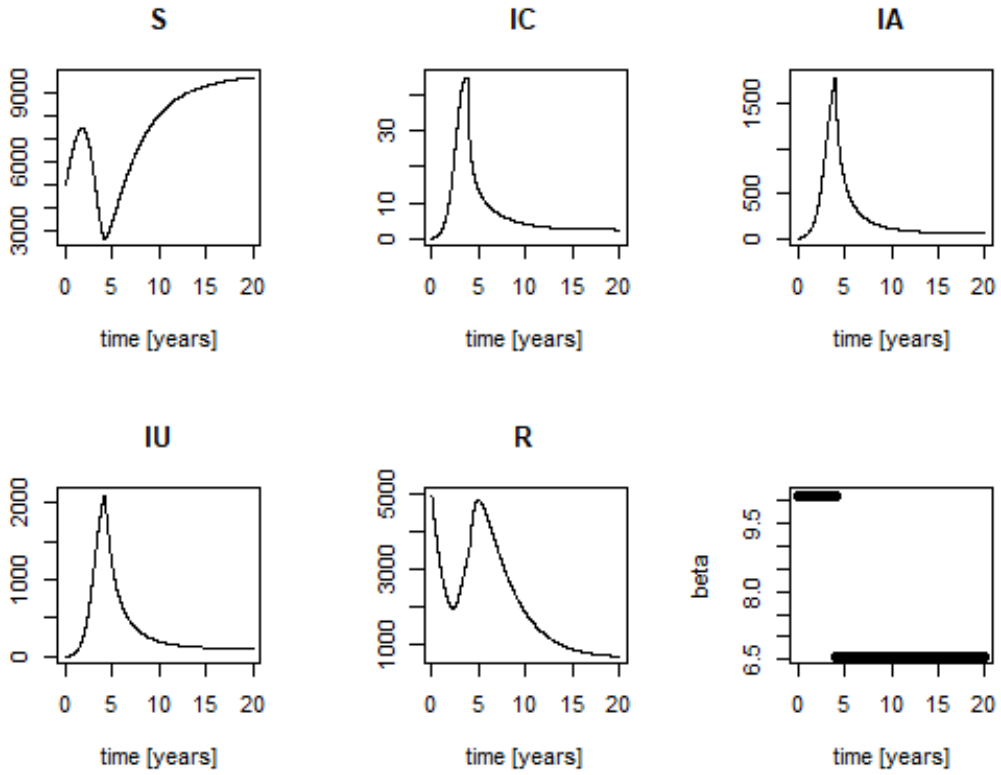
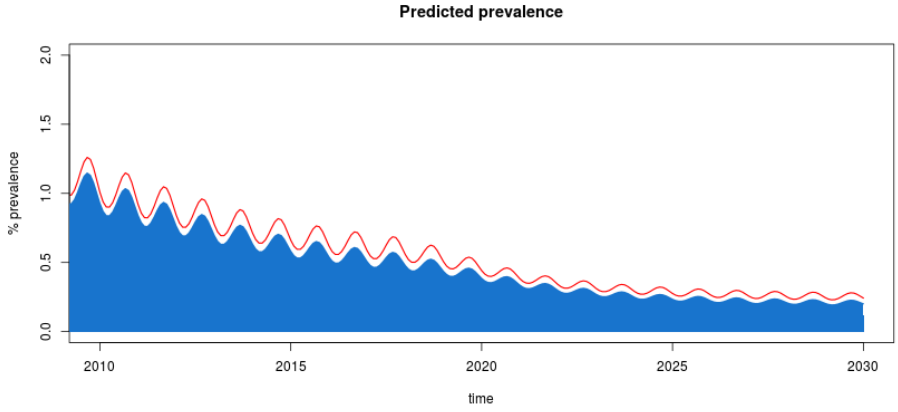


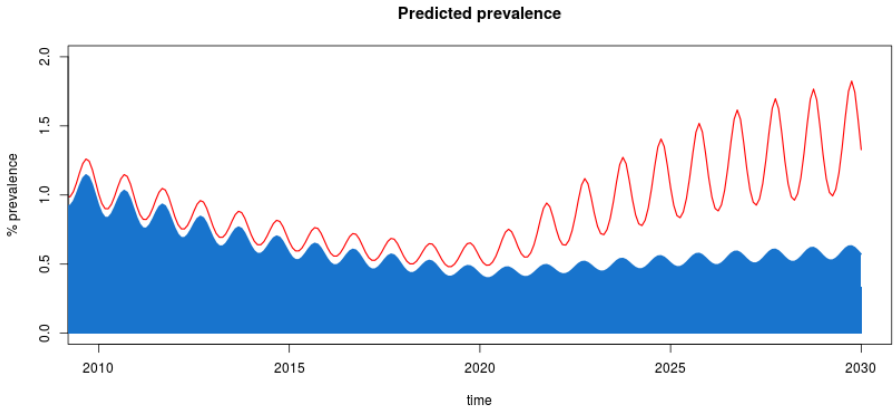
Figure 7.4: Model dynamics for the scenario where a highly effective vector control intervention is applied (after 4 years) and continuously undertaken in the same system, a drastic decrease in malaria cases is observed. Here β changes in time according to the coverage and effectiveness of the intervention.

dynamics, we simulated the effect of interventions leading to the increase of either coverage or compliance. The system dynamics was ran for a wide range of final values for both of theses parameters (Figure 7.7), looking to find out which possible intervention would most effectively lead to a decrease in the total number of cases. From the relation between the percentual decrease in prevalence and compliance/coverage (Figure 7.6), we conclude that actions towards the increase of EDAT coverage might be a slightly more effective as a control measure (Figure 7.6).

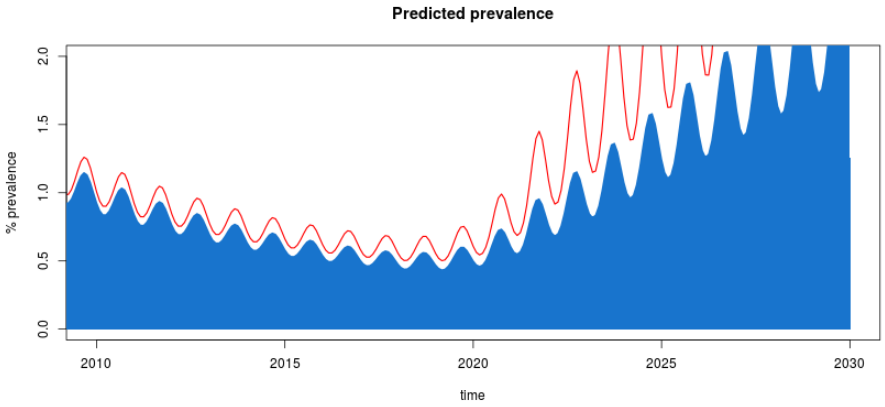
However, the system is apparently already in an optimal regime, concerning the mentioned parameters, and such improvements might not be ultimately required. Even though, the continuous maintenance of an effective treatment system is essential to keep controlled levels of the disease, as we can see by the dark lines of simulations in Figure 7.7.



(a) Business as usual



(b) Decreasing compliance to treatment



(c) Decreasing coverage of treatment

Figure 7.5: *P. vivax* malaria (blue) and all-species malaria (red) true prevalence outputs when an intervention starts in 2018 and linearly varies coverage and compliance until 2021 (as in Figure 7.7). (a) shows the scenario when both interventions are applied successfully. Complementary, in (b) we see the effect of decreasing compliance to treatment to 0% and the same in (c), for decreasing coverage to 0%. Other parameters correspond to the values in table 7.1.

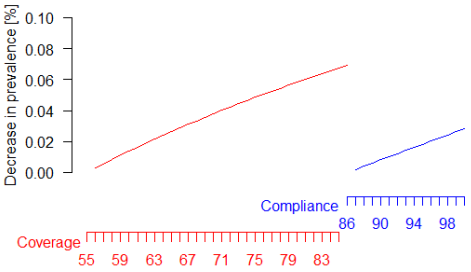
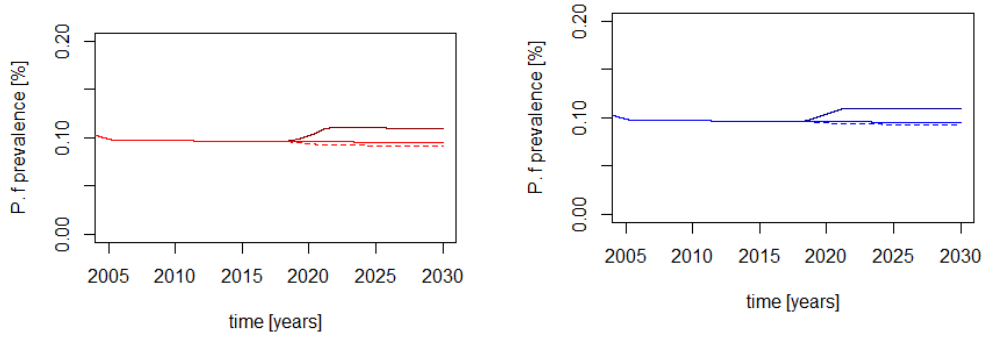


Figure 7.6: Preliminary sensitivity analysis for the percentual coverage of EDAT and compliance to treatment. We can see that a slightly higher percentual decrease in prevalence can be obtained with broader coverage of EDAT. Parameters were the same as in Figure 7.6.



(a) Coverage (ω_T) scenarios

(b) Compliance (τ) scenarios

Figure 7.7: True prevalence outputs when an intervention starts in 2018 and linearly varies a treatment parameter until 2021. (a) shows the effect of increasing to 85% (red dashed line), maintaining at 55% (red solid line) and decreasing to 0% (dark red solid line) the **coverage (ω_T)** of EDAT (related to the parameter ω_T) in the *P. falciparum* model (without seasonality). Analogously, (b) shows the effect of increasing to 100% (blue dashed line), maintaining at 86% (blue solid line) and decreasing to 0% (blue red solid line) the **compliance (τ)** to treatment. Other parameters correspond to the values in table 7.1. Analogous plot for *P. vivax* are qualitatively equivalent.

7.3 Conclusion and perspectives

Single-species models for malaria can be useful to qualitatively reproduce the disease dynamics at population level, with similar structures to the ones described in this chapter having been used to inform health policy making in different settings [139; 172; 132]. Nevertheless, at the time, the lack of more detailed data regarding case importations, testing and intervention coverage/adherence in the Brazilian Amazon limited the impact of our models. Further communication with stakeholders is fundamental, in order to acquire the relevant data to produce

quantitative and useful outputs, aiming to support strategic decision making within the national malaria control program.

Additionally, the lack of explicit coupling between the two species dynamics in the models discussed in this chapter ignores the existence of co-infected individuals, which can transmit and respond to treatment for more than one species at a time. Hence the need for multi-species models that incorporate such interactions is evident.

In chapter 8, we describe how multi-species models can be built from the coupling between single-species systems. The TRansmission of Related Infectionswith Light Linkage (TRILL) approximation is intruduced as a framework to make multi-species modelling feasible. We also present examples of applications of the TRILL framework in recent modelling exercises aiming to inform malaria elimination strategic planning in Asia-Pacific [133] and, more recently, in the Guyana Shield region, lead by S. Silal *et al.*

8 TRILL approximation for multi-species systems

Models for multiple strains or species of pathogens are challenging due to the number of equations required to accurately account for the inter-species interactions, which increase exponentially with the number of species, making numerical and structural identifiability increasingly improbable. We propose an approximation which allows the dimension of the resulting model to increase linearly with the number of species. For a two-species SIRS model, we show through mathematical analysis that this approximation is appropriate for species with low levels of interaction, which we term TRansmission of Related Infections with Light Linkage (TRILL). We apply numerical analysis to confirm the analytical result and explore the less analytically tractable example of a multi-species malaria model.

8.1 Multi-species epidemiological modelling

Realistic multi-species models become increasingly complicated in an exponential fashion as we increase the number of interacting species being modelled. Considering an underlying closed population dynamics where all possible combinations of interactions between N species are taken into account, we can arrive at $(s^N - 1)$ equations when writing the system dynamics in terms of its s^N discrete state variables, with s being the number of discrete states (e.g., $s = 3$ for a SIR or SIRS system).

Pathogens with multiple variants or species might interact in complex ways, including through different levels of cross-immunity or response to treatments. These mechanisms are not completely understood, representing an additional challenge when trying to model [97; 92]. Multiple-strain population models have been previously proposed to investigate immunity duration in a system with different *P. falciparum* [77; 78] or influenza [92; 73] strains. Nevertheless, multi-species models are rare, since different species might have very different life

cycles and and interact in even more complex ways. A few two-species malaria models can be found in the literature, but they usually include overly simplifying assumptions regarding the multiple species life-cycle [98; 123], they are not made at population-level [52; 119] or do not cosider inter-species super-infection [119].

Aiming to be able to model multi-species epidemiological dynamics with less extensive, but not over-simplistic, systems of equations, we propose an approximation which allows the dimension of such systems to increase linearly with the number of species, rather than exponentially.

In the following sections, we first introduce a one species SIRS dynamic system as a baseline model. We then replicate its dynamics for two species (with similar epidemiological cycles, e. g. *Plasmodium falciparum* and *Plasmodium vivax* if we were modelling malaria) including all possible interactions terms between them, leading to a set of 9 equations that represent the dynamics of the two-species state-specific variables. The full model is subsequently re-written in terms of ecological variables, reducing the number of equations to 5. Assuming a weak interaction regime (or *TRansmission of Related Infections with Light Linkage (TRILL))*, we prove that the system can be further reduced to only 4 equations.

Finally, we propose that such framework can be generalised for N-species systems under the same *low linkage* regime so, by redefining variables at ecological level [97] and imposing the TRILL approximation, we can reduce it to an equivalent set of $sN - N$ equations. Therefore, being able to replace the exponential association between number of species and equations by a linear one.

8.2 Baseline model

The baseline model describing an SIRS structure for a single species has $(s-1)$ equations where s is the number of discrete states (3 in this example):

$$X' = -\beta XY + \omega(1-X) \tag{8.1}$$
$$Y' = \beta XY - (\nu+\omega)Y \tag{8.2}$$

Here we can see X as susceptible, Y as infectious and Z as recovered individuals. The parameters represent transmissibility (β), recovery (ν) and loss of immunity (ω) rates. Note that the Z ($Z' = \nu Y - \omega Z$) dynamic equation is redundant under the assumption of aclosed population, i.e., $X + Y + Z = 1$.

8.2.1 Full model

An extended two-species model considering the same underlying natural history, where individuals can be in any combination of the s states for each species, would be described by (s^2-1) equations. In the case of the example system presented in

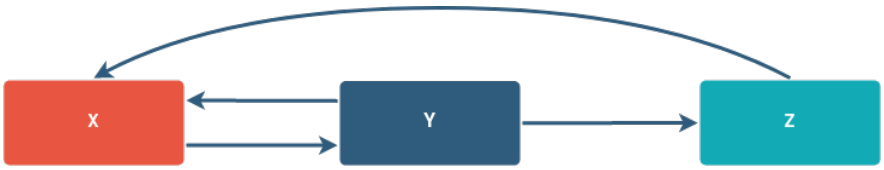


Figure 8.1: Diagram representing the baseline model compartments and flows. X represents susceptible, Y is the infectious and Z is the recovered sub-population.

the previous section, we would have the following set of 9 equations representing the two-species dynamics:

$$\begin{aligned}
V_{\{X_i,X_j\}}{}' &= -(\Lambda_i+\Lambda_j)\,V_{\{X_i,X_j\}} + \omega_j(V_{\{X_i,Y_j\}} + V_{\{X_i,Z_j\}}) \\
&\quad +\omega_i(V_{\{Y_i,X_j\}} + V_{\{Z_i,X_j\}}) && (8.3)\\
V_{\{Y_i,X_j\}}{}' &= \Lambda_i V_{\{X_i,X_j\}} - (1-\alpha_{ij})\Lambda_j V_{\{Y_i,X_j\}} + \omega_j(V_{\{Y_i,Y_j\}} + V_{\{Y_i,Z_j\}}) \\
&\quad -(\omega_i+\nu_i)V_{\{Y_i,X_j\}} && (8.4)\\
V_{\{Z_i,X_j\}}{}' &= \nu_i V_{\{Y_i,X_j\}} + \omega_j(V_{\{Z_i,Y_j\}} + V_{\{Z_i,Z_j\}}) - \omega_i V_{\{Z_i,X_j\}} \\
&\quad -(1-\alpha_{ij})\Lambda_j V_{\{Z_i,X_j\}} && (8.5)\\
V_{\{Z_i,Y_j\}}{}' &= \nu_i V_{\{Y_i,Y_j\}} + (1-\alpha_{ij})\Lambda_j V_{\{Z_i,X_j\}} - (\nu_j+\omega_i+\omega_j)V_{\{Z_i,Y_j\}} && (8.6)\\
V_{\{Y_i,Y_j\}}{}' &= (1-\alpha_{ji})\Lambda_i V_{\{X_i,Y_j\}} + (1-\alpha_{ij})\Lambda_j V_{\{Y_i,X_j\}} \\
&\quad -(\nu_i+\nu_j+\omega_i+\omega_j)V_{\{Y_i,Y_j\}} && (8.7)\\
V_{\{Z_j,Z_j\}}{}' &= \nu_i V_{\{Y_i,Z_j\}} + \nu_j V_{\{Z_i,Y_j\}} - (\omega_i+\omega_j)V_{\{Z_i,Z_j\}} && (8.8)
\end{aligned}$$

with $(i,j) \in \{(1,2),(2,1)\}$ and where

$$\Lambda_i = \beta_i(V_{\{Y_i,X_j\}} + V_{\{Y_i,Y_j\}} + V_{\{Y_i,Z_j\}}) \tag{8.9}$$

represents the system's force of infection. All parameters are defined for each of the species the same way as in the above section, except for α_{ij}, which can be defined as a cross-immunity index.

The full model set of equations in this example can be reduced to only 5 ordinary differential equations, when written in terms of its ecological variables, X_i, Y_i, Z_i, defined as:

$$\begin{aligned}
X_i &= V_{\{X_i,X_j\}} + V_{\{X_i,Y_j\}} + V_{\{X_i,Z_j\}} && (8.10)\\
Y_i &= V_{\{Y_i,X_j\}} + V_{\{Y_i,Y_j\}} + V_{\{Y_i,Z_j\}} && (8.11)\\
Z_i &= V_{\{Z_i,X_j\}} + V_{\{Z_i,Y_j\}} + V_{\{Z_i,Z_j\}} && (8.12)
\end{aligned}$$

with the constant total population constraint

$$X_i + Y_i + Z_i = 1 \tag{8.13}$$

which simplifies the system by making the Z_1 and Z_2 dynamic equations redundant.

With the above definitions, we can re-write equations (8.3)-(8.8) as:

$$\begin{aligned} V_{\{X_i,X_j\}}{}' &= -(\beta_i Y_i + \beta_j Y_j)V_{\{X_i,X_j\}} + \omega_i(X_i - V_{\{X_i,X_j\}}) \\ &\quad +\omega_j(X_j - V_{\{X_i,X_j\}}) \qquad (8.14) \\ X_i' &= -\beta_i Y_i \left[V_{\{X_i,X_j\}} + (1-\alpha_{ji})(X_i - V_{\{X_i,X_j\}})\right] \\ &\quad +\omega_i(1 - X_i) \qquad (8.15) \\ Y_i' &= \beta_i Y_i \left[V_{\{X_i,X_j\}} + (1-\alpha_{ji})(X_i - V_{\{X_i,X_j\}})\right] \\ &\quad -(\nu_i + \omega_i)Y_i \qquad (8.16) \end{aligned}$$

again with $(i,j) \in \{(1,2),(2,1)\}$.

Note this is merely an artefact of the chosen structure. The structure was chosen to demonstrate the concept, but numerical analysis can be used in the case of less tractable structures, such as the multi-species malaria models exemplified in section 8.5.

8.3 TRansmission of Related Infections with Light Linkage (TRILL) approximation

We propose an approximation in order to reduce the dimension of the full system. The approximation assumes that at each time point, the proportion of the population in the state of uninfected and unimmune to either species, $V_{\{X_i,X_j\}}$, is the product of the proportions in the two composite states of X_1 and X_2, the proportions of the population in the state of uninfected and unimmune to each species.

First, let $W_{\{X_i,X_j\}} = X_i X_j$

$$\begin{aligned} W_{\{X_i,X_j\}}{}' &= X_i{}' X_j + X_i X_j{}' \\ &= -\beta_i Y_i \left[V_{\{X_i,X_j\}} + (1-\alpha_{ji})(X_i - V_{\{X_i,X_j\}})\right] X_j \\ &\quad -\beta_j Y_j \left[V_{\{X_i,X_j\}} + (1-\alpha_{ij})(X_j - V_{\{X_i,X_j\}})\right] X_i \\ &\quad +\omega_i\left(X_j - X_i X_j\right) + \omega_j\left(X_i - X_i X_j\right) \\ &= -\beta_i Y_i \left[W_{\{X_i,X_j\}} - \alpha_{ji}(W_{\{X_i,X_j\}} - X_j V_{\{X_i,X_j\}})\right] \\ &\quad -\beta_j Y_j \left[W_{\{X_i,X_j\}} - \alpha_{ji}(W_{\{X_i,X_j\}} - X_i V_{\{X_i,X_j\}})\right] \\ &\quad +\omega_i\left(X_j - W_{\{X_i,X_j\}}\right) \\ &\quad +\omega_j\left(X_i - W_{\{X_i,X_j\}}\right). \qquad (8.17) \end{aligned}$$

Then, rearranging equation (8.14), we can re-write $V'_{\{X_i,X_j\}}$ as

$$\begin{aligned} V'_{\{X_i,X_j\}} &= -\beta_i Y_i V_{\{X_i,X_j\}} \\ &\quad -\beta_j Y_j V_{\{X_i,X_j\}} \\ &\quad +\omega_i (X_j - V_{\{X_i,X_j\}}) \\ &\quad +\omega_j (X_i - V_{\{X_i,X_j\}}). \end{aligned} \tag{8.18}$$

By comparing the terms of $W'_{\{X_i,X_j\}}$ and $V'_{\{X_i,X_j\}}$, we can clearly see that $W_{\{X_i,X_j\}} \to V_{\{X_i,X_j\}}$ as $\alpha_{ij} \to 0$ and/or $\beta_i \to 0$.

8.3.1 Error and equilibrium analysis

In order to analytically evaluate the difference between the full and approximate models we define the error function

$$E_{\{X_i,X_j\}} = V_{\{X_i,X_j\}} - W_{\{X_i,X_j\}} \tag{8.19}$$

with first derivative given by

$$\begin{aligned} E' &= -\left[\beta_i Y_i (1-\alpha_{ji}) + \beta_j Y_j (1-\alpha_{ij}) + \omega_i + \omega_j\right] E - \\ &\quad \left[\alpha_{ji}\beta_i Y_i (1-X_j) V_{\{X_i,X_j\}} + \alpha_{ij}\beta_j Y_j (1-X_i) V_{\{X_i,X_j\}}\right]. \end{aligned} \tag{8.20}$$

The second derivative of the error is negative, therefore the error will always tend to it's equilibrium value. The error at equilibrium, E^*, is given by:

$$E^* = -\frac{\alpha_{ji}\beta_i Y_i^* \left(1-X_j^*\right) V^*_{\{X_i,X_j\}} + \alpha_{ij}\beta_j Y_j^* (1-X_i^*) V^*_{\{X_i,X_j\}}}{\beta_i Y_i^* (1-\alpha_{ji}) + \beta_j Y_j^* (1-\alpha_{ij}) + \omega_i + \omega_j} \tag{8.21}$$

The terms in the numerator are cubic in the equilibrium values of the the system variables, while the terms in the denominator are linear in the equilibrium values of the the system variables. Since all the system variables are normalised to less than unity, this would imply that the equilibrium value of the error term would be small.

There are four equilibria of the system. The disease-free equilibrium, the single species equilibrium and the coexistence equilibrium. In all but the coexistence equilibrium, $E^* = 0$. The size of the error for the coexistence equlibrium is corellated most strongly with with α_{ji} as it appears in the numerator and its complement appears in the denominator. The rate of loss of immunity, ω_i, appears in the denominator and thus the size of the error is also correlated with the duration of immunity, $\frac{1}{\omega_i}$.

We can therefore conclude that the approximation is appropriate for species with low levels of cross-immunity and short duration of immunity. We term this weak

form of inter-species interaction as *TRansmission of Related Infections with Light Linkage (TRILL).*

The approximate model can therefore be written in its simplest form as 4 ODEs in 4 variables.

8.4 Approximate model equations

Here we define the lower-case variables x, y, z and $v_{\{x_i,y_j\}}$ as the *TRILL* approximations for the previously defined X, Y, Z and $V_{\{X_i,Y_j\}}$

$$v_{\{x_i,x_j\}} = x_i x_j \tag{8.22}$$

$$v_{\{x_i,x_j\}}{}' = x_i{}' x_j + x_i x_j{}' \tag{8.23}$$

$$x_i{}' = -\beta_i y_i \left[x_j + (1-\alpha_{ji})(1-x_j)\right] x_i + \omega_i (1-x_i) \tag{8.24}$$

$$y_i{}' = \beta_i y_i \left[x_j + (1-\alpha_{ji})(1-x_j)\right] x_i - (\omega_i + \nu_i) y_i, \tag{8.25}$$

$(i,j) \in \{(1,2),(2,1)\}$.

Note that, due to the way we defined X, Y and Z for the specific system in Section 8.2, we only end up with $V_{\{X_i,Y_j\}}$ terms left. For more complicated systems, other mixed terms could appear. In that case, we assume the same reasoning holds and we can make approximations such as $V_{\{U_i,W_j\}} \approx U_i W_j$.

8.5 Multi-species malaria models

The theoretical framework presented here has been applied by Silal et al. [133] to model *P.falciparum* - *P. vivax* transmission and costing in the Asia-Pacific (METCAP). Two independent single species models (similar to the ones described in chapter 7) are coupled as weak interspecific interactions are taken into account. The interactions implemented in the METCAP model were dual treatment, triggering and radical cure, masking, competition between species and the protective effect of G6PDd against the disease.

The model is also spatially explicit, as intervention packages are modelled for 22 different countries in the region, modelled as interconnected patches. Figure 8.2 shows the minimum intervention package to allow elimination at national level, as obtained by the model simulations. The simulations took into account *P. vivax*, *P. falciparum* and mixed incidence, making clear the importance of considering interspecific interactions when modelling towards elimination of the disease as a whole.

A similar modelling effort is in progress towards malaria elimination in the Guyana Shield region, which encompasses Guyana, Suriname, French Guiana, as well as

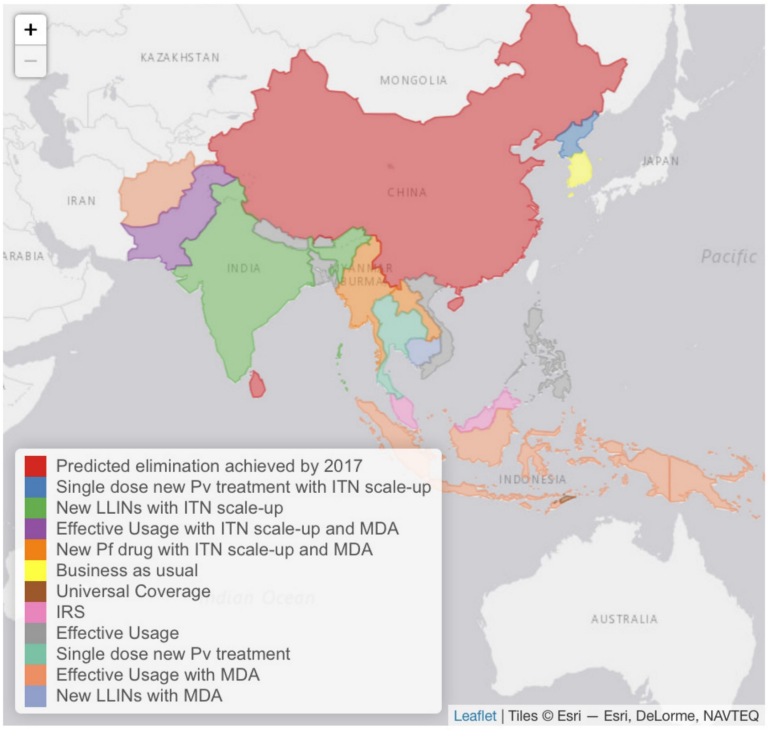


Figure 8.2: Predicted minimum intervention package to achieve malaria elimination by 2030. A minimum package refers to minimum effort where, given the nested nature of the scenarios, Business As Usual < Universal Coverage < IRS < Effective Usage < Single Dose Radical Cure < New LLINs < New Pf Drug. All scenarios are considered 'less effort' without the addition of Mass Drug Administration. (Figure and caption reproduced from Silal et al. [133])

some states in the Brazilian Amazon (Amapá, Pará and Roraima) with varied levels of endemicity among countries.

Due to mining, logging and other activities, significant migration lays out a very interconnected environment. In this context, the the achievement and maintenance of malaria elimination depends on coordinated technical and financial efforts. Hence, the underlying models need to reflect this interconectedness together with multi-species interactions. This is an ongoing project within or modelling group.

8.6 Discussion

Multiple species are generally considered separately [119] or are intrinsically linked with very complex models [92]. In reality species do interact but the complexity of full interactions in models deters modelling approaches.

The 2 species example SIRS system that was initially being modelled with a set of 9 equations, in terms of explicit state variables, was substituted by an approximately equivalent (under the TRILL regimen) system of 4 equations, written now in terms

of ecological variables. Extending this reasoning for any system with s baseline state variables and N species, we obtain that the initial system with s^N equations can be reduced to a set of $sN - N$ equations, simply by defining the variables in a different way under the TRILL regimen. Note that the new *ecological variables* are potentially more relevant in the epidemiological context.

Our proposed modelling approach could be seen as an *ecological neutrality approximation*, where the dynamics of the ecological variables (X_i, Y_i, Z_i) is made sure to be independent of the strain dynamics within the *neutral* regimen, correspondent to the parametric region allowing coexistence only for low levels of cross-immunity and short duration of immunity. Such contribution provided the theoretical support for multi-species modelling exercises, such as the one used by Silal et al. [133] to model two-species malaria in Asia-Pacific.

The TRILL approximation framework can be seen as a powerful tool that makes it feasible to deal with many previously mathematically, numerically and computationally intractable modelling challenges. This could lead to important implications for evidence-based horizontal public-health strategy development. For example, we will be able to consider more *Plasmodium* species simultaneously when planning strategies towards malaria eradication.

Future work includes generalising our analytical proof for systems with more than two species and with more complex single species dynamics.

9

Final remarks

"We all know that art is not truth. Art is a lie that makes us realise truth, at least the truth that is given us to understand. The artist must know the manner whereby to convince others of the truthfulness of his lies."

(Pablo Picasso, 1923)

The mathematical modelling of infectious diseases is an extremely interdisciplinary field, where we need to amalgamate different areas of expertise, cultures and scientific backgrounds. Maintenance of solid inter-disciplinary collaboration networks is essential for fact-checking, critically discussing results and hence producing any policy-relevant conclusions.

Based on biological and epidemiological data, compartmental models can be tailored to investigate different research questions regarding a disease spread dynamics within a population. With robust model fitting to data and well informed parametrisation, both qualitative and quantitative predictions regarding the behaviour of epidemics or endemics are possible.

In our COVID-19 modelling exercise, we wanted to produce a robust model for non-pharmaceutical intervention scenarios comparison that took into account household-level heterogeneities. To bridge the gap between compartmental and household-level models for COVID-19, we have used concepts from network theory inside a differential equation model, resulting in a mean-field approach to the population behaviour under social distancing interventions. It was both a challenging and rewarding experience to be able to contribute with new methodologies and inform policy making through our COVID-19 models, which were built in an extremely interdisciplinary, timely and international collaborative environment.

To tackle the issue of qualitatively investigating the effect of different COVID-19 vaccination strategies, we have adopted a simpler model structure (less

compartments and implemented interventions), but used delayed differential equations to account for the time lag between doses. Hence, we were able to evaluate optimal vaccine delivery strategies.

A different model was used for the same disease, when investigating its multi-variant dynamics. In this case, we were interested in quantitatively estimating the value of specific parameters, since the research question revolved around whether a new variant of concern was more transmissible than the wild virus.

In the context of endemic malaria, we presented single-species models that not only can be useful, but have been used to inform health policymakers. Nevertheless, as multiple species can coexist in some parts of the world, it is valuable to evaluate their combined burden while taken into account their interactions. Hence, a novel multi-species modelling framework was proposed: the TRansmission of Related Infectionswith Light Linkage (TRILL) approximation is introduced as a framework to enable the coupling of multiple species dynamics in a simplified way.

Even though all these models were built from a SEIRS-like baseline structure, the varied mechanisms, strains or interventions implemented unravelled different insights, which can be further extended and applied for modelling alternative diseases.

Going back to the "All models are wrong, but some are useful" aphorism (George Box), the enquiry we want to keep in mind is not whether our model is the right or true one (which will never be the case), but whether it is illuminating and useful for a particular purpose. As Georg Rash once wrote, " no models are [true] — not even the Newtonian laws. When you construct a model you leave out all the details which you, with the knowledge at your disposal, consider inessential. Models should not be true, but it is important that they are applicable, and whether they are applicable for any given purpose must of course be investigated. This also means that a model is never accepted finally, only on trial."

Ultimately, the work of a mathematical modeller is everlasting, as there will never be one true model, but rather an infinitude of model variations appropriate to answer another infinitude of research questions.

Bibliography

[1] CoMo Consortium: Covid-19 App, 2020. URL https://comomodel.net.

[2] Observatório COVID-19 BR: Comparação entre cenários, 2020. URL https://covid19br.github.io/index.html.

[3] Observatório COVID-19 BR: Casos graves de COVID-19 e SRAG em São Paulo, 2021. URL https://covid19br.github.io/municipios?aba=aba1&uf=SP&mun=Sao_Paulo&q=dia.

[4] D. Adam. Special report: The simulations driving the world's response to COVID-19. *Nature*, 580(7802):316–319, 2020.

[5] R. Aguas, L. White, N. Hupert, R. Shretta, W. Pan-Ngum, O. Celhay, A. Moldokmatova, F. Arifi, A. Mirzazadeh, H. Sharifi, K. Adib, M. N. Sahak, C. Franco, and R. Coutinho. Modelling the COVID-19 pandemic in context: an international participatory approach. *BMJ Global Health*, 5(12), 2020. doi:10.1136/bmjgh-2020-003126. URL https://gh.bmj.com/content/5/12/e003126.

[6] L. J. Allen, F. Brauer, P. Van den Driessche, and J. Wu. *Mathematical epidemiology*, volume 1945. Springer, 2008.

[7] A.M. Dondorp, S. Yeung, L. White, et al. Artemisinin resistance: current status and scenarios for containment. *Nature Reviews Microbiology*, 8:272–280, 2010.

[8] D. R. Anderson. *Information Theory and Entropy*. Springer, New York, 2008.

[9] E. J. Anderson, N. G. Rouphael, A. T. Widge, L. A. Jackson, P. C. Roberts, M. Makhene, J. D. Chappell, M. R. Denison, L. J. Stevens, A. J. Pruijssers, A. B. McDermott, B. Flach, B. C. Lin, N. A. Doria-Rose, S. O'Dell, S. D. Schmidt,

K. S. Corbett, P. A. Swanson, M. Padilla, K. M. Neuzil, H. Bennett, B. Leav, M. Makowski, J. Albert, K. Cross, V. V. Edara, K. Floyd, M. S. Suthar, D. R. Martinez, R. Baric, W. Buchanan, C. J. Luke, V. K. Phadke, C. A. Rostad, J. E. Ledgerwood, B. S. Graham, and J. H. Beigel. Safety and Immunogenicity of SARS-CoV-2 mRNA-1273 Vaccine in Older Adults. *New England Journal of Medicine*, 383(25):2427–2438, 2020. doi:10.1056/NEJMoa2028436. URL https://doi.org/10.1056/NEJMoa2028436. PMID: 32991794.

[10] R. M. Anderson. Discussion: The Kermack-McKendrick epidemic threshold theorem. *Bulletin of Mathematical Biology*, 53(1):3–32, 1991. ISSN 0092-8240. doi:https://doi.org/10.1016/S0092-8240(05)80039-4. URL https://www.sciencedirect.com/science/article/pii/S0092824005800394.

[11] A. Anirudh. Mathematical modeling and the transmission dynamics in predicting the Covid-19 - What next in combating the pandemic. *Infectious Disease Modelling*, 5:366–374, 2020.

[12] J. Arino, F. Brauer, P. van den Driessche, J. Watmough, and J. Wu. A model for influenza with vaccination and antiviral treatment. *Journal of Theoretical Biology*, 253(1):118–130, July 2008. doi:10.1016/j.jtbi.2008.02.026. URL https://doi.org/10.1016/j.jtbi.2008.02.026.

[13] S. Bansal, B. T. Grenfell, and L. A. Meyers. When individual behaviour matters: homogeneous and network models in epidemiology. *Journal of the Royal Society Interface*, 4(16):4879—-891, October 2007. ISSN 1742-5662. doi:10.1098/rsif.2007.1100. URL https://royalsocietypublishing.org/doi/10.1098/rsif.2007.1100.

[14] L. S. Bastos, T. Economou, M. F. C. Gomes, D. A. M. Villela, F. C. Coelho, O. G. Cruz, O. Stoner, T. Bailey, and C. T. Codeço. A modelling approach for correcting reporting delays in disease surveillance data. *Statistics in Medicine*, 38(22):4363–4377, 2019. doi:https://doi.org/10.1002/sim.8303. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.8303.

[15] L. M. Beck-Johnson, W. A. Nelson, K. P. Paaijmans, A. F. Read, M. B. Thomas, and O. N. Bjørnstad. The effect of temperature on anopheles mosquito population dynamics and the potential for malaria transmission. *PLoS ONE*, 8, 2013. doi:10.1371/journal.pone.0079276.

[16] D. Bernoulli. Essai d'une nouvelle analyse de la mortalite causee par la petite verole. *Mém. Math. Phys. Acad. R. Sci. (1760)*, 1(45), 1766.

[17] D. Bernoulli and S. Blower. An attempt at a new analysis of the mortality caused by smallpox and of the advantages of inoculation to prevent it. *Reviews in Medical Virology*, 14(5):275–288, 2004. doi:https://doi.org/10.1002/rmv.443. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/rmv.443.

[18] Q. Bi, Y. Wu, S. Mei, C. Ye, X. Zou, Z. Zhang, X. Liu, L. Wei, S. A. Truelove, T. Zhang, W. Gao, C. Cheng, X. Tang, X. Wu, Y. Wu, B. Sun, S. Huang, Y. Sun, J. Zhang, T. Ma, J. Lessler, and T. Feng. Epidemiology and transmission of covid-19 in 391 cases and 1286 of their close contacts in shenzhen, china: a retrospective cohort study. *The Lancet Infectious Diseases*, 20:911–919, 2020. doi:10.1016/S1473-3099(20)30287-5. URL https://doi.org/10.1016/S1473-3099(20)30287-5.

[19] M. E. Borges, L. S. Ferreira, S. Poloni, Â. M. Bagattini, C. Franco, M. Q. M. da Rosa, L. M. Simon, S. A. Camey, R. d. S. Kuchenbecker, J. A. F. D. Filho, R. A. Kraenkel, R. M. Coutinho, and C. M. Toscano. Modeling the impact of school reopening and contact tracing strategies on covid-19 dynamics in different epidemiologic settings in brazil. *medRxiv*, 2021. doi:10.1101/2021.10.22.21264706. URL https://www.medrxiv.org/content/early/2021/10/25/2021.10.22.21264706.

[20] BRASIL. Lei nº 12.527, de 18 de novembro de 2011. URL http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm.

[21] BRASIL. Constituição da república federativa do brasil (1988), 1988. URL http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm.

[22] BRASIL. Datasus. SRAG 2020 - Banco de Dados de Síndrome Respiratória Aguda Grave - incluindo dados da COVID-19, 2020. URL https://opendatasus.saude.gov.br/dataset/bd-srag-2020.

[23] BRASIL. Governo Federal. Acesso à informação. URL https://www.gov.br/acessoainformacao/pt-br. Accessed: 18/11/2021.

[24] BRASIL. Governo Federal. LocalizaSUS - Dados e informações sobre a Campanha Nacional de Vacinação contra a Covid-19 – 2021, 2021. URL https://localizasus.saude.gov.br/. Accessed: 22/11/2022.

[25] BRASIL. Ministério da Saúde. SIVEP Gripe - Sistema de informação de vigilância Epidemilógica da Gripe, Dicionário de Dados - Ficha de Registro Individual – Casos de Síndrome Respiratória Aguda Grave Hospitalizados, 2020. URL https://opendatasus.saude.gov.br/dataset/ae90fa8f-3e94-467e-a33f-94adbb66edf8/resource/8f571374-c555-4ec0-8e44-00b1e8b11c25/download/dicionario-de-dados-srag-hospitalizado-27.07.2020-final.pdf.

[26] BRASIL. Ministério da Saúde. SRAG 2021 - Banco de Dados de Síndrome Respiratória Aguda Grave - incluindo dados da COVID-19, 2021. URL https://opendatasus.saude.gov.br/dataset/bd-srag-2021.

[27] BRASIL. Ministério da Saúde. DataSUS. Suporte ao sistema de registro de notificações e-sus notifica, 2021. URL https://datasus.saude.gov.br/notifica/.

[28] BRASIL. Ministério da Saúde. Gabinete do Ministro. Portaria nº 1.792, de 17 de julho de 2020, 2020. URL https://www.in.gov.br/en/web/dou/-/portaria-n-1.792-de-17-de-julho-de-2020-267730859#:~:text=%C3%89%20obrigat%C3%B3ria%20a%20notifica%C3%A7%C3%A3o%20ao,outros%2C%20em%20todo%20territ%C3%B3rio%20nacional.

[29] BRASIL. Ministério da Saúde. Rede Interagencial de Informações para Saúde. Indicadores de morbidade: D.4 Índice parasitário anual (ipa) de malária linh, 2021. URL http://tabnet.datasus.gov.br/cgi/deftohtm.exe?idb2012/d04.def.

[30] BRASIL. Ministério da Saúde. Secretaria de Vigilância em Saúde. Diretoria Técnica de Gestão. *Ações de Controle da Malária*. 2005.

[31] BRASIL. Ministério da Saúde. Secretaria Executiva. Plano de dados abertos (PDA) - Ministério da Saúde 2020-2022, 2020. URL https://www.gov.br/saude/pt-br/acesso-a-informacao/dados-abertos/.

[32] F. Brauer. The Kermack–McKendrick epidemic model revisited. *Mathematical Biosciences*, 198(2):119–131, 2005. ISSN 0025-5564. doi:https://doi.org/10.1016/j.mbs.2005.07.006. URL https://www.sciencedirect.com/science/article/pii/S0025556405001331.

[33] N. Britton. *Essential mathematical biology*. Springer, 2003. ISBN 1-85233-536-X.

[34] K. P. Burnham and D. R. Anderson. *Model Selection and multi-model inference: A practical information-theoretic approach*. Springer New York, 2013.

[35] Centers for Disease Control and Prevention. Malaria biology. URL https://www.cdc.gov/malaria/about/biology/index.html.

[36] Centers for Disease Control and Prevention. Definition of Epidemiology. In: Principles of Epidemiology in Public Health Practice, Third Edition: An Introduction to Applied Epidemiology and Biostatistics, 2012. URL https://www.cdc.gov/csels/dsepd/ss1978/lesson1/section1.html.

[37] Centers for Disease Control and Prevention COVID-19 Response Team. Severe Outcomes Among Patients with Coronavirus Disease 2019 (COVID-19) - United States, February 12 – March 16, 2020. *MMWR Morb Mortal Wkly Rep*, 69:343––6, 2020.

[38] M. Cevik, M. Tate, O. Lloyd, A. E. Maraolo, J. Schafers, and A. Ho. SARS-CoV-2, SARS-CoV-1 and MERS-CoV viral load dynamics, duration of

viral shedding and infectiousness – a living systematic review and meta-analysis. July 2020. doi:10.1101/2020.07.25.20162107. URL https://doi.org/10.1101/2020.07.25.20162107.

[39] L. R. Chaves. Uma ferramenta para acompanhar a epidemia em São Paulo, 24/04/2020. URL https://revistapesquisa.fapesp.br/uma-ferramenta-para-acompanhar-a-epidemia-em-sao-paulo/.

[40] Y. Chen, G. Paul, R. Cohen, S. Havlin, S. P. Borgatti, F. Liljeros, and H. Eugene Stanley. Percolation theory and fragmentation measures in social networks. *Physica A: Statistical Mechanics and its Applications*, 378(1):11–19, 2007. ISSN 0378-4371. doi:https://doi.org/10.1016/j.physa.2006.11.074. URL https://www.sciencedirect.com/science/article/pii/S0378437106012611.

[41] D. K. Chu, E. A. Akl, S. Duda, K. Solo, S. Yaacoub, H. J. Schünemann, D. K. Chu, E. A. Akl, A. El-harakeh, A. Bognanni, T. Lotfi, M. Loeb, A. Hajizadeh, A. Bak, A. Izcovich, C. A. Cuello-Garcia, C. Chen, D. J. Harris, E. Borowiack, F. Chamseddine, F. Schünemann, G. P. Morgano, G. E. U. M. Schünemann, G. Chen, H. Zhao, I. Neumann, J. Chan, J. Khabsa, L. Hneiny, L. Harrison, M. Smith, N. Rizk, P. G. Rossi, P. AbiHanna, R. El-khoury, R. Stalteri, T. Baldeh, T. Piggott, Y. Zhang, Z. Saad, A. Khamis, M. Reinap, S. Duda, K. Solo, S. Yaacoub, and H. J. Schünemann. Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. *The Lancet*, 395 (10242):1973–1987, June 2020. doi:10.1016/s0140-6736(20)31142-9. URL https://doi.org/10.1016/s0140-6736(20)31142-9.

[42] G. E. Coelho, P. L. Leal, M. d. P. Cerroni, A. C. R. Simplicio, and J. B. Siqueira, Jr. Sensitivity of the dengue surveillance system in brazil for detecting hospitalized cases. *PLOS Neglected Tropical Diseases*, 10(5):1–12, 05 2016. doi:10.1371/journal.pntd.0004705. URL https://doi.org/10.1371/journal.pntd.0004705.

[43] V. M. Corman, O. Landt, and M. Kaiser et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro surveillance*, 25(3), Sep 2020. doi:10.2807/1560-7917.ES.2020.25.3.2000045. URL https://doi.org/10.2807/1560-7917.ES.2020.25.3.2000045.

[44] R. M. Coutinho, F. M. D. Marquitti, L. S. Ferreira, M. E. Borges, R. L. P. da Silva, O. Canton, T. P. Portella, S. Poloni, C. Franco, M. M. Plucinski, F. C. Lessa, A. A. M. da Silva, R. A. Kraenkel, M. A. de Sousa Mascena Veras, and P. I. Prado. Model-based estimation of transmissibility and reinfection of SARS-CoV-2 P.1 variant. *Communications Medicine*, 1(1):1–8, Nov. 2021. ISSN 2730-664X. doi:10.1038/s43856-021-00048-6. URL https://www.nature.com/articles/s43856-021-00048-6.

[45] B. J. Cowling and G. M. Leung. Epidemiological research priorities for public health control of the ongoing global novel coronavirus (2019-nCoV) outbreak. *Euro surveillance*, 25(6), Feb 2020. doi:doi.org/10.2807/1560-7917.ES.2020.25.6.2000110. URL https://doi.org/10.2807/1560-7917.ES.2020.25.6.2000110.

[46] D. Cyranoski. Profile of a killer: the complex biology powering the coronavirus pandemic. *Nature*, 581(7806):22–26, May 2020. doi:10.1038/d41586-020-01315-73.

[47] G. Davey Smith, M. Blastland, and M. Munafò. Covid-19's known unknowns. *BMJ*, 371, 2020. doi:10.1136/bmj.m3979. URL https://www.bmj.com/content/371/bmj.m3979.

[48] N. G. Davies, P. Klepac, Y. Liu, K. P. A. M. Jit, C. C.-. working group, and R. M. Eggo. Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nature Medicine*, 26:1205–1211, 2020. doi:10.1038/s41591-020-0962-9.

[49] N. G. Davies, A. J. Kucharski, R. M. Eggo, A. Gimma, W. J. Edmunds, T. Jombart, K. O'Reilly, A. Endo, J. Hellewell, E. S. Nightingale, et al. Effects of non-pharmaceutical interventions on covid-19 cases, deaths, and demand for hospital services in the UK: a modelling study. *The Lancet Public Health*, 5(7): e375–e385, 2020.

[50] M. Day. Covid-19: four fifths of cases are asymptomatic, china figures indicate. *BMJ*, 369, 2020. doi:10.1136/bmj.m1375. URL https://www.bmj.com/content/369/bmj.m1375.

[51] E. DE and Y. MD. The duration of untreated or inadequately treated Plasmodium falciparum infections in the human host. *Journal of the National Malaria Society*, 10:327–36, 1951.

[52] R. D. DEMASSE and A. DUCROT. An age-structured within-host model for multistrain malaria infections. *SIAM Journal on Applied Mathematics*, 73(1): 572–593, 2013. ISSN 00361399. URL http://www.jstor.org/stable/23479888.

[53] Disease Outbreak News - World Health Organization. SARS-CoV-2 mink-associated variant strain – Denmark, 2020. URL https://www.who.int/emergencies/disease-outbreak-news/item/2020-DON301.

[54] Disease Outbreak News - World Health Organization. SARS-CoV-2 Variants, 2020. URL https://www.who.int/emergencies/disease-outbreak-news/item/2020-DON305.

[55] Disease Outbreak News - World Health Organization. SARS-CoV-2 Variant – United Kingdom of Great Britain and Northern Ireland, 2020. URL https://www.who.int/emergencies/disease-outbreak-news/item/2020-DON304.

[56] M. Ehrhardt, J. Gašper, and S. Kilianová. SIR-based mathematical modeling of infectious diseases with vaccination and waning immunity. *Journal of Computational Science*, 37:101027, Oct. 2019. doi:10.1016/j.jocs.2019.101027. URL https://doi.org/10.1016/j.jocs.2019.101027.

[57] E. H. Elbasha and A. P. Galvani. Vaccination against multiple HPV types. *Mathematical Biosciences*, 197(1):88–117, Sept. 2005. doi:10.1016/j.mbs.2005.05.004. URL https://doi.org/10.1016/j.mbs.2005.05.004.

[58] T. V. Elzhov, K. M. Mullen, A.-N. Spiess, and B. Bolker. *minpack.lm: R Interface to the Levenberg-Marquardt Nonlinear Least-Squares Algorithm Found in MINPACK, Plus Support for Bounds*, 2016. URL https://CRAN.R-project.org/package=minpack.lm. R package version 1.2-1.

[59] H. Escobar. Flexibilização do isolamento tende a endurecer pandemia, 19/06/2020. URL https://jornal.usp.br/atualidades/flexibilizacao-do-isolamento-tende-a-endurecer-pandemia/.

[60] H. Escobar. Pesquisadores alertam para risco de suspender isolamento em São Paulo, 26/05/2020. URL https://jornal.usp.br/ciencias/pesquisadores-alertam-para-risco-de-suspender-isolamento-em-sao-paulo/.

[61] J. W. Essam. Percolation theory. *Reports on Progress in Physics*, 43(7):833–912, July 1980. doi:10.1088/0034-4885/43/7/001. URL https://doi.org/10.1088/0034-4885/43/7/001.

[62] F. Zhou, T. Yu, R. Du, *et al.* Clinical course and risk factors for mortality of adult inpatients with covid-19 in wuhan, china: a retrospective cohort study. *Lancet*, 3:1054–1062, 2020. doi:10.1016/S0140-6736(20)30566-3. URL https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(20)30566-3/fulltext.

[63] F. J. Fabozzi, S. M. Focardi, S. T. Rachev, and B. G. Arshanapalli. *Appendix E: Model Selection Criterion: AIC and BIC*, pages 399–403. John Wiley & Sons, Ltd, 2014. ISBN 9781118856406. doi:https://doi.org/10.1002/9781118856406.app5. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118856406.app5.

[64] N. R. Faria, T. A. Mellan, C. Whittaker, I. M. Claro, D. da S. Candido, S. Mishra, M. A. E. Crispim, F. C. S. Sales, I. Hawryluk, J. T. McCrone, R. J. G. Hulswit, L. A. M. Franco, M. S. Ramundo, J. G. de Jesus, P. S. Andrade, T. M. Coletti, G. M. Ferreira, C. A. M. Silva, E. R. Manuli, R. H. M. Pereira, P. S. Peixoto, M. U. G. Kraemer, N. Gaburo, C. da C. Camilo, H. Hoeltgebaum, W. M. Souza, E. C. Rocha, L. M. de Souza, M. C. de Pinho, L. J. T. Araujo, F. S. V. Malta, A. B.

de Lima, J. do P. Silva, D. A. G. Zauli, A. C. de S. Ferreira, R. P. Schnekenberg, D. J. Laydon, P. G. T. Walker, H. M. Schlüter, A. L. P. dos Santos, M. S. Vidal, V. S. D. Caro, R. M. F. Filho, H. M. dos Santos, R. S. Aguiar, J. L. Proença-Modena, B. Nelson, J. A. Hay, M. Monod, X. Miscouridou, H. Coupland, R. Sonabend, M. Vollmer, A. Gandy, C. A. Prete, V. H. Nascimento, M. A. Suchard, T. A. Bowden, S. L. K. Pond, C.-H. Wu, O. Ratmann, N. M. Ferguson, C. Dye, N. J. Loman, P. Lemey, A. Rambaut, N. A. Fraiji, M. do P. S. S. Carvalho, O. G. Pybus, S. Flaxman, S. Bhatt, and E. C. Sabino. Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. *Science*, 372(6544):815–821, 2021. doi:10.1126/science.abh2644. URL https://doi.org/10.1126/science.abh2644.

[65] N. Ferguson, D. Laydon, G. Nedjati Gilani, N. Imai, K. Ainslie, M. Baguelin, S. Bhatia, A. Boonyasiri, Z. Cucunuba Perez, G. Cuomo-Dannenburg, et al. Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce covid19 mortality and healthcare demand. Technical report, Imperial College London, 2020.

[66] Ferguson, N., Keeling, M., John Edmunds, W. et al. Planning for smallpox outbreaks. *Nature*, 425:681–685, 2003. doi:10.1038/nature02007.

[67] S. Flaxman, S. Mishra, A. Gandy, H. J. T. Unwin, T. A. Mellan, H. Coupland, C. Whittaker, H. Zhu, T. Berah, J. W. Eaton, et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in europe. *Nature*, 584(7820): 257–261, 2020.

[68] C. Franco. Malaria cases in Brazil, Synapse ID: syn26483904, 2021.

[69] C. Franco, L. S. Ferreira, V. Sudbrack, M. E. Borges, S. Poloni, P. I. Prado, L. J. White, R. Águas, R. A. Kraenkel, and R. M. Coutinho. Percolation across households in mechanistic models of non-pharmaceutical interventions in sars-cov-2 disease dynamics. *Epidemics*, 39:100551, 2022. ISSN 1755-4365. doi:https://doi.org/10.1016/j.epidem.2022.100551. URL https://www.sciencedirect.com/science/article/pii/S1755436522000111.

[70] C. Franco, R. Coutinho, and R. Kraenkel. Com o avanço da COVID-19, o Brasil deve adotar já medidas drásticas de confinamento? SIM, 21/03/2020.

[71] L. Fumanelli, M. Ajelli, S. Merler, N. M. Ferguson, and S. Cauchemez. Model-based comprehensive analysis of school closure policies for mitigating influenza epidemics and pandemics. *PLOS Computational Biology*, 12(1):1–15, 01 2016. doi:10.1371/journal.pcbi.1004681. URL https://doi.org/10.1371/journal.pcbi.1004681.

[72] G. Giordano, F. Blanchini, and R. Bruno et al. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy.

Nature Medicine, 26:855–860, Apr 2020. doi:10.1038/s41591-020-0883-7. URL https://doi.org/10.1038/s41591-020-0883-7.

[73] J. R. Gog and B. T. Grenfell. Dynamics and selection of many-strain pathogens. *Proceedings of the National Academy of Sciences*, 99(26): 17209–17214, 2002. ISSN 0027-8424. doi:10.1073/pnas.252512799. URL https://www.pnas.org/content/99/26/17209.

[74] Google. Covid-19 mobility reports. Technical report, Google, 2020. URL https://www.google.com/covid19/mobility/.

[75] D. Greenhalgh. Vaccination campaigns for common childhood diseases. *Mathematical Biosciences*, 100(2):201–240, July 1990. doi:10.1016/0025-5564(90)90040-6. URL https://doi.org/10.1016/0025-5564(90)90040-6.

[76] S. M. Griffing, V. Tauil, P. L. NAD Udhayakumar, and L. Silva-Flannery. A historical perspective on malaria control in brazil. *Mem Inst Oswaldo Cruz*, 110(6):701–718, 2015.

[77] S. Gupta, K. Trenholme, R. M. Anderson, and K. P. Day. Antigenic diversity and the transmission dynamics of plasmodium falciparum. *Science*, 263(5149):961–963, 1994. doi:10.1126/science.8310293. URL https://www.science.org/doi/abs/10.1126/science.8310293.

[78] Gupta, S., Maiden, M., Feavers, I., et al. The maintenance of strain structure in populations of recombining infectious agents. *Nat Med*, (2):437––442, 1996. doi:10.1038/nm0496-437.

[79] W. Hamer. The milroy lectures on epidemic disease in england—the evidence of variability and persistence of type. *Lancet*, 1:733–739, 1906.

[80] M. Honigsbaum. *The pandemic century: one hundred years of panic, hysteria and hubris*. Hurst, 2019.

[81] Instituto Brasileiro de Geografia e Estatística. *Metodologia do censo demográfico 2010*. IBGE, Rio de Janeiro, 2016. ISBN 9788524043628. URL https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=296501.

[82] Instituto Brasileiro de Geografia e Estatística. Tábuas completas de mortalidade, 2019. URL https://www.ibge.gov.br/estatisticas/sociais/populacao/9126-tabuas-completas-de-mortalidade.html?=&t=resultados.

[83] Instituto Brasileiro de Geografia e Estatística. Pesquisas estatísticas do registro civil - tabela 2679 - nascidos vivos, por ano de nascimento, idade da

mãe na ocasião do parto, sexo e lugar do registro, 2021. URL https://sidra.ibge.gov.br/tabela/2679.

[84] International Organization for Standardization (ISO). *INTERNATIONAL STANDARD ISO 8601 - Data elements and interchange formats — Information interchange — Representation of dates and times*. 3 edition, 2004.

[85] J. Hasell, E. Mathieu, D. Beltekian *et al.* A cross-country database of covid-19 testing. *Nature Scientific Data*, 7(345), 2020. doi:doi.org/10.1038/s41597-020-00688-8.

[86] C. JD, Smith, B. PF, T. W, L. EOK, and M. JHET. Survival and infection probabilities of anthropophagic anophelines from an area of high prevalence of plasmodium falciparum in humans. *Bulletin of Entomological Research*, 87: 445–53, 2009. doi:10.1017/S0007485300041304.

[87] K. Kameda, M. Malheiros, Barbeitas, R. Caetano, I. LöwyAna, C. D. de Oliveira, M. C. D. V. Corrêa, and M. Cassier. A testagem para COVID-19 no Brasil: esforços fragmentados e desafios para ampliar a capacidade diagnóstica no Sistema Único de Saúde. *Cadernos de Saúde Pública*, 37, 2021. doi:10.1590/0102-311X00277420.

[88] M. J. Keeling and P. Rohani. *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press, 2008.

[89] W. O. Kermack, A. G. McKendrick, and G. T. Walker. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 115 (772):700–721, 1927. doi:10.1098/rspa.1927.0118. URL https://royalsocietypublishing.org/doi/abs/10.1098/rspa.1927.0118.

[90] M. Khalili, M. Karamouzian, N. Nasiri, S. Javadi, A. Mirzazadeh, and H. Sharifi. Epidemiological characteristics of covid-19: a systematic review and meta-analysis. *Epidemiology and Infection*, 148:e130, 2020. doi:10.1017/S0950268820001430.

[91] A. N. Kraay, M. E. Gallagher, Y. Ge, P. Han, J. M. Baker, K. Koelle, A. Handel, and B. A. Lopman. Modeling the use of SARS-CoV-2 vaccination to safely relax non-pharmaceutical interventions. Mar. 2021. doi:10.1101/2021.03.12.21253481. URL https://doi.org/10.1101/2021.03.12.21253481.

[92] A. J. Kucharski, V. Andreasen, and J. R. Gog. Capturing the dynamics of pathogens with many strains. *Journal of Mathematical Biology*, 72:1–24, 2016. doi:10.1007/s00285-015-0873-4.

[93] L. W. Church LW, T. P. Le, J. P. Bryan, D. M. Gordon, R. Edelman, L. Fries, *et al.* Clinical manifestations of Plasmodium falciparum malaria experimentally induced by mosquito challenge. *Journal of Infectious Diseases*, 175, 1997.

[94] J. M. Last. *A Dictionary of epidemiology*. Oxford University Press, 1983.

[95] J. A. Lewnard, V. X. Liu, M. L. Jackson, M. A. Schmidt, B. L. Jewell, J. P. Flores, C. Jentz, G. R. Northrup, A. Mahmud, A. L. Reingold, M. Petersen, N. P. Jewell, S. Young, and J. Bellows. Incidence, clinical outcomes, and transmission dynamics of severe coronavirus disease 2019 in california and washington: prospective cohort study. *BMJ*, 369, 2020. doi:10.1136/bmj.m1923. URL https://www.bmj.com/content/369/bmj.m1923.

[96] N. M. Linton, T. Kobayashi, Y. Yang, K. Hayashi, A. R. Akhmetzhanov, S.-m. Jung, B. Yuan, R. Kinoshita, and H. Nishiura. Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: A statistical analysis of publicly available case data. *Journal of Clinical Medicine*, 9(2), 2020. ISSN 2077-0383. doi:10.3390/jcm9020538. URL https://www.mdpi.com/2077-0383/9/2/538.

[97] M. Lipsitch, C. Colijn, T. Cohen, W. P. Hanage, and C. Fraser. No coexistence for free: Neutral null models for multistrain pathogens. *Epidemics*, 1(1):2–13, Mar. 2009. ISSN 17554365. doi:10.1016/j.epidem.2008.07.001. URL https://linkinghub.elsevier.com/retrieve/pii/S1755436508000029.

[98] B. Mahato, B. K. Mishra, A. Jayswal, and B. K. Mishra. Mathematical model of malaria for co-infection of plasmodium vivax and plasmodium falciparum in india. *British Microbiology Research Journal*, 5(3):285–299, november 2014. ISSN 2231-0886. doi:10.9734/BMRJ/2015/10267. URL http://www.sciencedomain.org/abstract/6851.

[99] F. Mariz. Observatório COVID-19 BR ajuda a compreender evolução da pandemia no Brasil, 23/03/2020. URL https://jornal.usp.br/ciencias/ciencias-biologicas/observatorio-covid-19-br-ajuda-publico-a-compreender-evolucao-da-pandemia-no-brasil/.

[100] F. M. D. Marquitti, R. M. Coutinho, L. S. Ferreira, M. E. Borges, T. P. Portella, R. L. P. da Silva, O. Canton, S. Poloni, C. Franco, V. Coelho, L. Barberia, M. de Bolle, A. C. Boing, M. R. Donalisio, A. F. Boing, A. A. M. da Silva, P. I. Prado, M. A. de Sousa Mascena Veras, and R. A. Kraenkel.

[101] L. Matrajt, T. Britton, M. E. Halloran, and I. M. Longini. One versus two doses: What is the best use of vaccine in an influenza pandemic? *Epidemics*, 13: 17–27, Dec. 2015. doi:10.1016/j.epidem.2015.06.001. URL https://doi.org/10.1016/j.epidem.2015.06.001.

[102] K. Matuschewski. Getting infectious: formation and maturation of plasmodium sporozoites in the anopheles vector. *Cell Microbiology*, 8:1547–56, 2006. doi:10.1111/j.1462-5822.2006.00778.x.

[103] S. F. McGough, M. A. Johansson, M. Lipsitch, and N. A. Menzies. Nowcasting by bayesian smoothing: A flexible, generalizable model for real-time epidemic tracking. *PLOS Computational Biology*, 16(4):1–20, 04 2020. doi:10.1371/journal.pcbi.1007735. URL https://doi.org/10.1371/journal.pcbi.1007735.

[104] W. min Liu, H. W. Hethcote, and S. A. Levin. Dynamical behavior of epidemiological models with nonlinear incidence rates. *Journal of Mathematical Biology*, 25(4):359–380, Sept. 1987. doi:10.1007/bf00277162. URL https://doi.org/10.1007/bf00277162.

[105] Ministério da Saúde. Portaria nº 264, de 17 de fevereiro de 2020, 2020. URL https://bvsms.saude.gov.br/bvs/saudelegis/gm/2020/prt0264_19_02_2020.html.

[106] K. Mizumoto, K. Kagaya, A. Zarebski, and G. Chowell. Estimating the asymptomatic proportion of coronavirus disease 2019 (covid-19) cases on board the diamond princess cruise ship, yokohama, japan, 2020. *Eurosurveillance*, 25(10):2000180, 2020. doi:https://doi.org/10.2807/1560-7917.ES.2020.25.10.2000180. URL https://www.eurosurveillance.org/content/10.2807/1560-7917.ES.2020.25.10.2000180.

[107] J. Mossong, N. Hens, M. Jit, P. Beutels, K. Auranen, R. Mikolajczyk, M. Massari, S. Salmaso, G. S. Tomba, J. Wallinga, J. Heijne, M. Sadkowska-Todys, M. Rosinska, and W. J. Edmunds. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLOS Medicine*, 5 (3):1–1, 03 2008. doi:10.1371/journal.pmed.0050074. URL https://doi.org/10.1371/journal.pmed.0050074.

[108] N. Ferguson, D. Laydon and G. Nedjati Gilani et al. Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand, 16 March 2020.

[109] F. Naveca, V. Nascimento, V. Souza, A. Corado, F. Nascimento, G. Silva, Á. Costa, D. Duarte, K. Pessoa, M. Mejía, M. Brandão, M. Jesus, L. Gonçalves, C. da Costa, V. Sampaio, D. Barros, M. Silva, T. Mattos, G. Pontes, L. Abdalla, J. Santos, I. Arantes, F. Dezordi, M. Siqueira, G. Wallau, P. Resende, E. Delatorre, T. Gräff, and G. Bello. COVID-19 in Amazonas, Brazil, was driven by the persistence of endemic lineages and P.1 emergence. *Nature Medicine*, pages 1–9, 2021. doi:https://doi.org/10.1038/s41591-021-01378-7. URL https://www.nature.com/articles/s41591-021-01378-7.

[110] News - World Health Organization. WHO issues its first emergency use validation for a COVID-19 vaccine and emphasizes need for equitable global access, 2020. URL https://www.who.int/news/item/31-12-2020-who-issues-its-first-emergency-use-validation-for-a-covid-19-vaccine-and-emphasiz

[111] News - World Health Organization. COVAX Announces new agreement, plans for first deliveries, 2020. URL https://www.who.int/news/item/22-01-2021-covax-announces-new-agreement-plans-for-first-deliveries.

[112] N. B. Noll, I. Aksamentov, V. Druelle, A. Badenhorst, B. Ronzani, G. Jefferies, J. Albert, and R. A. Neher. Covid-19 scenarios: an interactive tool to explore the spread and associated morbidity and mortality of sars-cov-2. *medRxiv*, 2020. doi:10.1101/2020.05.05.20091363. URL https://www.medrxiv.org/content/early/2020/05/12/2020.05.05.20091363.

[113] Observatório COVID-19 BR. Repositório do site https://covid19br.github.io/. https://github.com/covid19br/covid19br.github.io, 2021.

[114] Pan American Health Organization. Malaria, . URL https://www.paho.org/en/topics/malaria.

[115] Pan American Health Organization. Interactive malaria statistics, . URL https://www3.paho.org/hq/index.php?option=com_content&view=article&id=2632:2010-interactive-malaria-statistics&Itemid=2130&lang=en.

[116] J. Panovska-Griffiths, C. Kerr, W. Waites, and R. Stuart. *Mathematical modeling as a tool for policy decision making: Applications to the COVID-19 pandemic*. Elsevier, 2021.

[117] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani. Epidemic processes in complex networks. *Rev. Mod. Phys.*, 87:925–979, Aug 2015. doi:10.1103/RevModPhys.87.925. URL https://link.aps.org/doi/10.1103/RevModPhys.87.925.

[118] C. M. Petrilli, S. A. Jones, J. Yang, H. Rajagopalan, L. O'Donnell, Y. Chernyak, K. A. Tobin, R. J. Cerfolio, F. Francois, and L. I. Horwitz. Factors associated with hospitalization and critical illness among 4,103 patients with covid-19 disease in new york city. *medRxiv*, 2020. doi:10.1101/2020.04.08.20057794. URL https://www.medrxiv.org/content/early/2020/04/11/2020.04.08.20057794.

[119] Pizzitutti, F., Pan, W., Barbieri, A. et al. A validated agent-based model to study the spatial and temporal heterogeneities of malaria incidence in the rainforest environment.

[120] F. P. Polack, S. J. Thomas, N. Kitchin, J. Absalon, A. Gurtman, S. Lockhart, J. L. Perez, G. Pérez Marc, E. D. Moreira, C. Zerbini, R. Bailey, K. A. Swanson,

S. Roychoudhury, K. Koury, P. Li, W. V. Kalina, D. Cooper, R. W. Frenck, L. L. Hammitt, Türeci, H. Nell, A. Schaefer, S. Ünal, D. B. Tresnan, S. Mather, P. R. Dormitzer, U. Şahin, K. U. Jansen, and W. C. Gruber. Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. *New England Journal of Medicine*, 383 (27):2603–2615, 2020. doi:10.1056/NEJMoa2034577. URL https://doi.org/10.1056/NEJMoa2034577. PMID: 33301246.

[121] T. P. Portella, S. R. Mortara, R. Lopes, A. Sánchez-Tapia, M. R. Donalísio, M. C. Castro, V. R. Venturieri, C. G. Estevam, A. F. Ribeiro, R. M. Coutinho, M. A. d. Sousa Mascena Veras, P. Inácio Prado, and R. André Kraenkel. Temporal and geographical variation of covid-19 in-hospital fatality rate in brazil. *medRxiv*, 2021. doi:10.1101/2021.02.19.21251949. URL https://www.medrxiv.org/content/early/2021/02/23/2021.02.19.21251949.

[122] K. Prem, K. van Zandvoort, P. Klepac, R. M. Eggo, N. G. Davies, A. R. Cook, and M. J. and. Projecting contact matrices in 177 geographical regions: An update and comparison with empirical data for the COVID-19 era. *PLOS Computational Biology*, 17(7):e1009098, July 2021. doi:10.1371/journal.pcbi.1009098. URL https://doi.org/10.1371/journal.pcbi.1009098.

[123] O. Prosper and M. Martcheva. Impact of enhanced malaria control on the competition between plasmodium falciparum and plasmodium vivax in india. *Mathematical Biosciences*, 242(1):33–50, 2013. ISSN 0025-5564. doi:https://doi.org/10.1016/j.mbs.2012.11.015. URL https://www.sciencedirect.com/science/article/pii/S0025556412002350.

[124] H. Rahmandad and J. Sterman. Heterogeneity and network structure in the dynamics of diffusion: Comparing agent-based and differential equation models. *Management Science*, 54(5):998–1014, May 2008. doi:10.1287/mnsc.1070.0787. URL https://doi.org/10.1287/mnsc.1070.0787.

[125] Rede Genômica Fiocruz. Plots of lineages presence by state, 2021. http://www.genomahcov.fiocruz.br/presenca-das-linhagens-por-estado, Accessed: 28/02/2021.

[126] S. Richardson, J. S. Hirsch, M. Narasimhan, J. M. Crawford, T. McGinn, K. W. Davidson, , and the Northwell COVID-19 Research Consortium. Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. *JAMA*, 323(20): 2052–2059, 05 2020. ISSN 0098-7484. doi:10.1001/jama.2020.6775. URL https://doi.org/10.1001/jama.2020.6775.

[127] RIO DE JANEIRO. Secretaria de Estado de Saúde. Sistema de Vigilância Epidemiológica da Gripe – SIVEP-Gripe, Casos de Síndrome Respiratória

Aguda Grave (SRAG), Nota Técnica, 2021. URL http://sistemas.saude.rj.gov.br/tabnetbd/sivep_gripe/SIVEP_Gripe.pdf.

[128] R. Ross. The prevention of malaria. 1906.

[129] M. Roy and M. Pascual. On representing network heterogeneities in the incidence rate of simple epidemic models. *Ecological Complexity*, 3(1):80–90, Mar. 2006. doi:10.1016/j.ecocom.2005.09.001. URL https://doi.org/10.1016/j.ecocom.2005.09.001.

[130] J. Sadoff, G. Gray, A. Vandebosch, V. Cárdenas, G. Shukarev, B. Grinsztejn, P. A. Goepfert, C. Truyers, H. Fennema, B. Spiessens, K. Offergeld, G. Scheper, K. L. Taylor, M. L. Robb, J. Treanor, D. H. Barouch, J. Stoddard, M. F. Ryser, M. A. Marovich, K. M. Neuzil, L. Corey, N. Cauwenberghs, T. Tanner, K. Hardt, J. Ruiz-Guiñazú, M. Le Gars, H. Schuitemaker, J. Van Hoof, F. Struyf, and M. Douoguih. Safety and efficacy of single-dose ad26.cov2.s vaccine against covid-19. *New England Journal of Medicine*, 384(23):2187–2201, 2021. doi:10.1056/NEJMoa2101544. URL https://doi.org/10.1056/NEJMoa2101544. PMID: 33882225.

[131] H. Salje, C. T. Kiem, N. Lefrancq, N. Courtejoie, P. Bosetti, J. Paireau, A. Andronico, N. Hozé, J. Richet, C.-L. Dubost, Y. L. Strat, J. Lessler, D. Levy-Bruhl, A. Fontanet, L. Opatowski, P.-Y. Boelle, and S. Cauchemez. Estimating the burden of SARS-CoV-2 in france. *Science*, 369(6500):208–211, May 2020. doi:10.1126/science.abc3517. URL https://doi.org/10.1126/science.abc3517.

[132] Silal, S P, Little, F, Barnes, K I et al. Towards malaria elimination in Mpumalanga, South Africa: a population-level mathematical modelling approach. *Malaria Journal*, 13(297), 2014. doi:10.1186/1475-2875-13-297.

[133] Silal, S P, Shretta, R, Celhay, O J, et al. Malaria elimination transmission and costing in the asia-pacific: a multi-species dynamic transmission model [version 2; peer review: 1 approved, 1 approved with reservations, 2 not approved]. *Wellcome Open Res 2019*, 4. doi:10.12688/wellcomeopenres.14771.2.

[134] A. Siqueira, P. Marchesini, R. M. Torres, S. Rodovalho, and T. Chaves. *Malária na atenção básica*. Nescon/UFMG, 2018.

[135] A. M. e. a. Siqueira. Plasmodium vivax landscape in brazil: Scenario and challenges. *The American Journal of Tropical Medicine and Hygiene*, 95(6 Suppl):87–96, 2016. doi:10.4269/ajtmh.16-0204.

[136] Slater HC, Ross A, Ouedraogo AL, White LJ, Nguon C, Walker PG, et al. Assessing the impact of next-generation rapid diagnostic tests on

Plasmodium falciparum malaria elimination strategies. *Nature*, 528:S94–101, 2015. doi:10.1038/nature16040.

[137] L. Souto Ferreira, O. Canton, R. L. P. da Silva, S. Poloni, V. Sudbrack, M. E. Borges, C. Franco, F. M. D. Marquitti, J. C. de Moraes, M. A. d. S. M. Veras, R. A. Kraenkel, and R. M. Coutinho. Assessing the best time interval between doses in a two-dose vaccination regimen to reduce the number of deaths in an ongoing epidemic of sars-cov-2. *PLOS Computational Biology*, 18(3):1–15, 03 2022. doi:10.1371/journal.pcbi.1009978. URL https://doi.org/10.1371/journal.pcbi.1009978.

[138] P. D. Stroud, S. J. Sydoriak, J. M. Riese, J. P. Smith, S. M. Mniszewski, and P. R. Romero. Semi-empirical power-law scaling of new infection rate to model epidemic dynamics with inhomogeneous mixing. *Mathematical Biosciences*, 203(2):301–318, Oct. 2006. doi:10.1016/j.mbs.2006.01.007. URL https://doi.org/10.1016/j.mbs.2006.01.007.

[139] S.T.T. Tun, L. von Seidlein, T. Pongvongsa, T. *et al.* Towards malaria elimination in savannakhet, lao pdr: mathematical modelling driven strategy design. *Malaria Journal*, 16(483), 2017. URL https://doi.org/10.1186/s12936-017-2130-3.

[140] W. W. Sun, F. Ling, J. R. Pan, J. Cai, Z. P. Miao, S. L. Liu, W. Cheng, and E. F. Chen. Epidemiological characteristics of COVID-19 family clustering in Zhejiang Province. *Chinese journal of preventive medicine*, 54(6):625–629, 2020. ISSN 02539624. doi:10.3760/cma.j.cn112150-20200227-00199.

[141] SÃO PAULO. Secretaria Municipal de Mobilidade e Transportes. Passageiros transportados - 2020, 2020. URL https://www.prefeitura.sp.gov.br/cidade/secretarias/transportes/institucional/sptrans/acesso_a_informacao/agenda/index.php?p=292723.

[142] T. Vos *et al.* Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet*, 396:1204 – 1222, 2020.

[143] The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team. The Epidemiological Characteristics of an Outbreak of 2019 Novel Coronavirus Diseases (COVID-19) — China. Technical report, 2020.

[144] C. M. Toscano, A. F. R. Lima, L. L. S. Silva, P. F. Razia, L. F. A. Pavão, D. A. Polli, R. F. Moraes, and M. A. Cavalcanti. Medidas de distanciamento social e evolução da covid-19 no brasil, 2020. URL https://medidas-covidbr-iptsp.shinyapps.io/painel/.

[145] R. Verity, L. C. Okell, I. Dorigatti, P. Winskill, C. Whittaker, N. Imai, G. Cuomo-Dannenburg, H. Thompson, P. G. Walker, H. Fu, A. Dighe, J. T. Griffin, M. Baguelin, S. Bhatia, A. Boonyasiri, A. Cori, Z. Cucunubá, R. FitzJohn, K. Gaythorpe, W. Green, A. Hamlet, W. Hinsley, D. Laydon, G. Nedjati-Gilani, S. Riley, S. van Elsland, E. Volz, H. Wang, Y. Wang, X. Xi, C. A. Donnelly, A. C. Ghani, and N. M. Ferguson. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *The Lancet Infectious Diseases*, 20(6):669–677, 2020. ISSN 14744457. doi:10.1016/S1473-3099(20)30243-7.

[146] E. Volz and L. A. Meyers. Susceptible-infected-recovered epidemics in dynamic contact networks. *Proceedings of the Royal Society B: Biological Sciences*, 274(1628):2925–2934, 2007. doi:10.1098/rspb.2007.1159. URL https://royalsocietypublishing.org/doi/abs/10.1098/rspb.2007.1159.

[147] M. Voysey and A. Marites et al. Single-dose administration and the influence of the timing of the booster dose on immunogenicity and efficacy of ChAdOx1 nCoV-19 (AZD1222) vaccine: a pooled analysis of four randomised trials. *The Lancet*, 397(10277):881–891, Mar. 2021. doi:10.1016/s0140-6736(21)00432-3. URL https://doi.org/10.1016/s0140-6736(21)00432-3.

[148] H. Wang, X. Li, T. Li, S. Zhang, L. Wang, X. Wu, and J. Liu. The genetic sequence, origin, and diagnosis of SARS-CoV-2. *Eur J Clin Microbiol Infect Dis*, 39(9):1629–1635, Sep 2020. doi:10.1007/s10096-020-03899-4. URL https://link.springer.com/article/10.1007%2Fs10096-020-03899-4.

[149] Z. Wang, J. S. Ji, Y. Liu, R. Liu, Y. Zha, X. Chang, L. Zhang, Q. Liu, Y. Zhang, J. Zeng, T. Dong, X. Xu, L. Zhou, J. He, Y. Deng, B. Zhong, and X. Wu. Survival analysis of hospital length of stay of novel coronavirus (covid-19) pneumonia patients in sichuan, china. *medRxiv*, 2020. doi:10.1101/2020.04.07.20057299. URL https://www.medrxiv.org/content/early/2020/04/10/2020.04.07.20057299.

[150] C. WE and J. GM. A retrospective examination of secondary sporozoite- and trophozoite-induced infections with Plasmodium falciparum: development of parasitologic and clinical immunity following secondary infection. *American Journal of Tropical Medicine and Hygiene*, 61:20–35, 1999.

[151] C. WE and J. GM. A retrospective examination of sporozoite- and trophozoite-induced infections with Plasmodium falciparum: development of parasitologic and clinical immunity during primary infection. *American Journal of Tropical Medicine and Hygiene*, 61:4–19, 1999.

[152] W. E. Wei, Z. Li, C. J. Chiew, S. E. Yong, M. P. Toh, and V. J. Lee. Presymptomatic transmission of SARS-CoV-2 — singapore, january 23–march 16, 2020. *MMWR. Morbidity and Mortality Weekly Report*, 69(14):411–415, Apr.

2020. doi:10.15585/mmwr.mm6914e1. URL https://doi.org/10.15585/mmwr.mm6914e1.

[153] Y. Wei, L. Wei, Y. Liu, L. Huang, S. Shen, R. Zhang, J. Chen, Y. Zhao, H. Shen, and F. Chen. A systematic review and meta-analysis reveals long and dispersive incubation period of covid-19. *medRxiv*, 2020.

[154] C. Whitty. What makes an academic paper useful for health policy? *BMC Medicine*, 13(301), 2015. doi:10.1186/s12916-015-0544-8.

[155] WHO Global Health Observatory (GHO) data. Brazil malaria profile.

[156] World Health Organization. Life expectancy data by who region. URL http://apps.who.int/gho/data/view.main.SDG2016LEXv?lang=en.

[157] World Health Organization. Communicable diseases and health analysis (cha), health information and analysis (cha/ha), epidemiological calendar 2016: A basic element for the use of the time variable in health surveillance, 2016. URL https://www.paho.org/hq/dmdocuments/2016/2016-cha-epidemiological-calendar.pdf.

[158] World Health Organization. World malaria report 2020: 20 years of global progress and challenges, 2020.

[159] World Health Organization. Global partnership to make available 120 million affordable, quality COVID-19 rapid tests for low- and middle-income countries, 2020. URL https://www.who.int/news/item/28-09-2020-global-partnership-to-make-available-120-million-affordable-quality-covid-19-rapid-

[160] World Health Organization. Timeline: WHO's COVID-19 response, 2020. URL https://www.who.int/emergencies/diseases/novel-coronavirus-2019/interactive-timeline.

[161] World Health Organization. Disease Outbreak News - COVID-19- China, 2020-01-05. URL https://www.who.int/emergencies/disease-outbreak-news/item/2020-DON229.

[162] World Health Organization. Coronavirus disease 2019 (COVID-19): situation report, 51. Technical documents, 2020-03-11.

[163] World Health Organization. Interim recommendations for use of the Moderna mRNA-1273 vaccine against COVID-19: interim guidance, 25 January 2021. Technical documents, 2021.

[164] World Health Organization. Interim recommendations for use of the Pfizer–BioNTech COVID-19 vaccine, BNT162b2, under emergency use listing: interim guidance, 8 January 2021. Technical documents, 2021.

[165] World Health Organization. Interim recommendations for use of the ChAdOx1-S [recombinant] vaccine against COVID-19 (AstraZeneca COVID-19 vaccine AZD1222 Vaxzevria™, SII COVISHIELD™): interim guidance, first issued: 10 February 2021, updated: 21 April 2021, last updated: 30 July 2021. Technical documents, 2021.

[166] World Health Organization. Interim recommendations for the use of the Janssen Ad26.COV2.S (COVID-19) vaccine: interim guidance, 17 March 2021. Technical documents, 2021.

[167] World Health Organization. WHO SAGE Roadmap For Prioritizing Uses Of COVID-19 Vaccines In The Context Of Limited Supply, 16 July 2021. Technical documents, 2021.

[168] World Health Organization. COVID-19 weekly epidemiological update - 12 January 2021. Technical documents, 2021-01-12.

[169] World Health Organization. COVID-19 weekly epidemiological update, edition 54, 24 August 2021. Technical documents, 2021-08-24.

[170] World Health Organization. Naming the coronavirus disease (COVID-19) and the virus that causes it, 2021-11-30. URL https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it.

[171] S. Zhao, Z. Xu, and Y. Lu. A mathematical model of hepatitis b virus transmission and its application for vaccination strategy in china. *International Journal of Epidemiology*, 29(4):744–752, Aug. 2000. doi:10.1093/ije/29.4.744. URL https://doi.org/10.1093/ije/29.4.744.

[172] R. Águas, L. White, R. Snow, and M. Gomes. Prospects for malaria eradication in sub-saharan africa. *PloS One*, 3, 2008.

[173] R. Águas, A. Mahdi, R. Shretta, P. Horby, M. Landray, and L. W. . CoMo Consortium. Potential health and economic impacts of dexamethasone treatment for patients with covid-19. *Nature Communications*, 12(915), 2021. doi:10.1038/s41467-021-21134-2. URL https://www.nature.com/articles/s43856-021-00048-6.