UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia "Galileo Galilei"

Master Degree in Physics of Data

Final Dissertation

Dynamics of interdependent epidemics: a physics approach

Thesis supervisor
Prof. Sandro Azaele
Thesis external supervisor
Prof.ssa Chiara Poletto
Thesis co-supervisor
Prof. Amos Maritan

Candidate
Marika Sartore

Academic Year 2021/2022

Contents

List of Figures

Abstract

Every year respiratory viruses such as influenza, respiratory syncytial virus, and human parainfluenza virus cause seasonal epidemics with peaks in hospitalizations and deaths. The SARS-CoV-2 pandemic has altered this seasonal pattern due to the severe non pharmaceutical interventions implemented against it [1][2]. However, with the relaxation of the interventions against SARS-CoV-2, the circulation of the other viruses is rising again, thus the understanding of the dynamics of interdependent epidemics remains critical. The co-circulation of multiple viruses is the result of factors concurring at different scales: from the microscopic scale of within-host infection mechanisms to the scale of the human encounters and mobility [3]. Extensive virological data becoming increasingly available are providing evidence of a complex network of virus-virus interactions [4]. Still, the extent of these interactions is not clear, and their role in the epidemic dynamics is far from being understood. For convenience purposes the epidemics caused by each virus are mainly studied separately. However a proper accounting for virus-virus interactions becomes essential to understand the interdependent epidemics and anticipate their future course. This requires a new approximate theory beyond current approaches [5][6] to tackle the coupled system of viruses' dynamical equations and enable scalable numerical simulations. This work introduces a new multi-pathogen dynamical model accounting for the competitive interaction between pathogens. The aim of this project is to explore the phase space of possible dynamical regimes. Understanding gained by numerical simulations will be backed up by theoretical considerations. Model trajectories will be compared with patterns observed in the real data.

Introduction

Each season there are multiple respiratory viruses circulating in the population and representing a burden for the public health. Among them, influenza, respiratory syncytial virus, human parainfluenza virus and others with a similar behavior. They share the same biological targets, they attack the same age groups and the symptoms of the infections coincide. For these reasons they are generally studied together, under the name of influenza like illness (ILI). Recently, multiple PCR tests have been developed and progressively employed in large scale studies. Through them it is possible to test individuals for many viruses with just one swab. This allows for the tracking of all the viruses currently circulating in a population.

Many studies are starting to monitor the composition of ILI cases revealing the viral agents causing the infections. This enables reconstructing viruses' co-circulation at the population level and provides evidence of virus-virus interactions [3]. Virus interactions can be inferred also from within-host observations, *e.g.* by analyzing virus co-occurrences, or their absence [4]. Most viruses are found to interfere with each other and are subject to cross-immunity: after an infection with a virus the hosts gain a temporary protection against infections with other viruses. The extension of such protection depends on the level of similarity between the first pathogen and a potential secondary one. The more similar they are, the stronger the protection while the more they differ, the weaker it is. The viral co-circulation dynamics resulting from these heterogeneous interactions forms a complex system.

Being able to describe virus co-circulation at the population level through dynamical models is necessary to understand what happens in reality and to have the possibility of exploring possible epidemiological and intervention scenarios. Compartmental models, such as SI, SIS and SIR, are a versatile technique used in epidemiology to study the spreading of infectious diseases at the population level. Majority of compartmental models focus on the spread of a single pathogen. Still, a few compartmental models take into account multiple pathogens and their interaction mechanisms. They are built upon two main frameworks introduced in [5] and in [6]. They adopt two different views, the first one the host's and the second the pathogen's one. The host-view model allows for both temporary and long-term cross-immunity, thus it is flexible, but it can not be upscaled to many viruses since the number of variables grows exponentially with the pathogens considered. On the contrary, the pathogen-view model is scalable to many viruses, but it is not flexible, it can account only for long-term cross-immunity thus it is more suitable for systems with highly similar viruses, *e.g.* strains of the same virus.

The thesis aims at filling the gap in the literature of multi-pathogen modeling by introducing a framework able to combine flexibility and scalability. We develop a multi-pathogen compartmental model that adopts the pathogen-view but is able to describe short term interaction. Interactions among viruses are included with the introduction of convalescent compartments that replicate the cross-immunity mechanism.

We use the model to investigate the consequences of multi-pathogen interaction on the multi-pathogen co-circulation dynamics. For illustrative purposes we take the perspective of influenza, by far the most studied virus. We reproduce flu incidence dynamics found in surveillance data. We try to answer three practical questions within our framework to gain fundamental understanding of the impact of other respiratory infections on influenza incidence:

- What is the effect of the presence of other respiratory viruses on influenza dynamics?
- What is the effect of the presence of other respiratory viruses on influenza parameters estimation?
- Is there a relation between the interactions strength matrix and the matrix of the correlations between the incidences?

The thesis is structured as follows. In chapter 1 we provide an introduction to epidemiology and to compartmental models, both mathematically and practically. In chapter 2 we make an overview of the variety of respiratory viruses and their interactions. Then we present the state-of-the-art of the compartmental models describing such scenarios. The model developed during this project is presented in chapter 3, where it is also compared to the host-view model. We retrieve the dynamics of a virus similar to influenza with our model and we use it as a reference during the analysis described in chapter 4. Eventually, in chapter 5 there are the conclusions of this work.

Chapter 1

Models in epidemiology

In the field of epidemiology, mathematical models have a fundamental role. It is necessary to develop the right model to describe the scenario under consideration. Modeling should be a trade-off between reality abstraction, in order to maintain a simple mathematical description, and accuracy to replicate the main aspects of the system. Once the model is properly defined, it becomes a predictive tool and allows for understanding the fundamental epidemiological processes.

This thesis work is about creating a framework capable of describing systems populated by many respiratory viruses. Since the next chapter goes into the development process, with the explanation of the biological context and the state-of-the-art of the actual models, this first chapter is dedicated to an introduction to mathematical models in epidemiology. In section 1.1 we briefly present the field of epidemiology and its history. Then, in section 1.2 we provide an overview on compartmental models from the mathematical point of view and in section 1.3 some of their practical applications are shown.

1.1 History of epidemiology

The word epidemiology comes from the union of the Greek words *epi* (= upon), *demos* (= people), *logos* (= study of), literally the study of what occurs among the population. Indeed, according to 'A Dictionary of Epidemiology' [7], its definition is:

'*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*'.

In other words, epidemiology studies the occurrence of diseases (or other health-related characteristics) in a specific population and tries to control them. A disease can be either infectious or noninfectious. The first one can be passed between individuals, whereas the second develops over an individual's lifespan. For this reason, the epidemiology of noninfectious diseases investigates the risk factors associated with the chance of developing the disease. In opposition, the primary cause of catching an infectious disease is the presence of infectious cases in the local population. This work focuses on the latter, where mathematical models have great predictive power at the population scale and over relatively short time scales.

The origins of mathematical epidemiology are far in time. There are studies about the spreading

Figure 1.1: Spot map made by Snow to track cholera cases and water pumps localization [9].

of infectious diseases from the past centuries, some examples are the ones by Bernoulli and Snow. Bernoulli in 1766 analyzed mortality induced by smallpox and highlighted the importance of prevention against the disease [8]. In 1854 Snow investigated the localization of cholera cases in London, during a cholera epidemic. He marked on a 'spot map', reported in Figure 1.1, all the cases because he believed in a correlation between households with cholera and water pumps localization [9]. Through his studies he was able to demonstrate that the source of disease in the area of interest was a specific pump. But the foundation of modern epidemiology can be attributed to Ross in 1911 [10]. In his works, he studied a system of differential equations to model the spreading dynamics of malaria. He was able to show that malaria can persist only if the number of mosquitos (the source of the infection) is above a certain threshold. As a consequence, it is not necessary to kill all the mosquitos to eradicate malaria, but only a fraction of them. Such work can be considered the first mathematical model of malaria transmission.

Similar epidemic models were later developed by Kermack and McKendrick in 1927 [11]. They summarized the spreading mechanism of a general infectious disease as follows: one (or more) infected person enters a population where people are more or less susceptible to that disease. The disease spreads from infected to susceptible by contact. Each infected individual has to go through the course of the infection and can get out of it either by recovery or death. The proportion of people involved, the timing and the severity are all characteristics depending on the single disease. They demonstrated that the epidemic does not need to terminate the susceptible population to be extinct. For each specific set of infectivity, recovery and death rates, there exists a critical susceptible population density below which the epidemic does not even start. The more this threshold is exceeded the smaller the population density will be at the end of the epidemic. In this case, the epidemic continues to increase as long as the density of the unaffected population is greater than the threshold density. When this point is reached the epidemic begins to wane and eventually dies out.

Still nowadays the progress of an infectious disease can be defined qualitatively in terms of pathogen load within the host, which in turn is determined by the growth rate of the pathogen and the interaction between the pathogen and the host's immune response. The scheme in Figure 1.2 shows the temporal evolution of an infection along with the corresponding response of the host immune system. Initially the host is susceptible to the infection: the pathogen is not present in his organism, and he has a low level of nonspecific immunity. At time 0 the host is exposed to an infectious individual and becomes infected with a microparasite from him. From this moment the internal abundance of the parasite grows over time. During this early phase the individual may exhibit no evident signs of infection and the abundance can be too low to cause a further transmission. Once the abundance is sufficiently large within the host, there is the possibility of transmitting the infection to other susceptible hosts. In this phase the host is infectious. At last, when the host's immune system has defeated the parasite attack, the host is no longer infectious and is recovered. In this simplified scenario, at this point the immune response developed to oppose the virus renders the host immune to further infections for a lifelong duration.

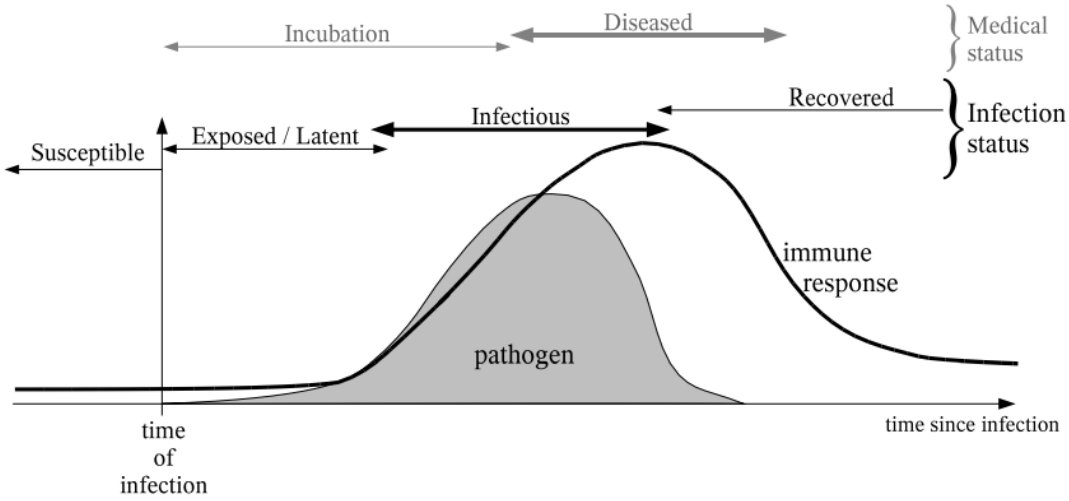

Figure 1.2: Figure from [5]. Temporal evolution of an infection from the point of view of the pathogen (gray curve) and the host's immune response (black curve). All the S-E-I-R phases are shown.

1.2 Compartmental models

Starting from Kermack and McKendrick work [11], epidemiological models are often formulated in terms of mutually exclusive compartments: hosts are grouped by their health status with respect to a pathogen. This modeling technique is called *compartmental model*. In this section we provide an overview of some of the most common compartmental models. We illustrate their versatility while maintaining a simple and clear mathematical description. All the mathematical details can be found in the book by Keeling and Rohani [5].

In compartmental models transitions between compartments, *i.e.* health status in Figure 1.2, occur according to rates. For example, the transition from susceptible to infected is mediated by the event of contagion and occurs at a rate that is proportional to the factors that affect such an event, *e.g.* the fraction of infected present in the population. Rate translates into probability of an individual moving from a compartment to the next one per unit time. As for all the models,

this is an abstraction of reality, done for mathematical convenience. Indeed, systems modeled by compartmental models can be easily represented by a set of ordinary differential equations.

The population is assumed to follow the *homogeneous mixing* hypothesis: all individuals are considered to behave equally and they have the same number of contacts $\langle k \rangle$ through which transmission can occur. This simplifies the mathematics of the model but it can result in an unrealistic scenario when dealing with a large population.

SI Model

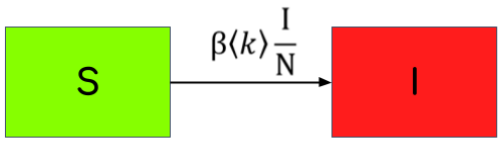

Figure 1.3: Scheme of SI compartmental model.

The most simple dynamics that a compartmental model can describe is the SI (Figure 1.3). The SI dynamics represent the contagion process where susceptible become infected with a rate proportional to:

- β, the transmissibility parameter which regulates the speed of the spreading;
- $\langle k \rangle$, the mean number of contacts that each individual has;
- $\frac{I}{N}$, the fraction of infected people in the population.

After catching the disease, infected individuals stay infectious forever. Such a dynamic leads to the spreading of the disease over the entire population N: in the end $I = N$. Speaking in terms of densities: $\frac{S}{N} + \frac{I}{N} = s + i = 1$. This could suit chronic infections or pathogens that always have a fatal outcome. An example from animals' diseases is the highly pathogenic avian influenza subtype (H5N1) [12].

SI model is described by the simple set of ordinary differential Equations 1.1:

$$\begin{aligned} \frac{ds}{dt} &= -\beta \langle k \rangle si \\ \frac{di}{dt} &= \beta \langle k \rangle si \end{aligned} \tag{1.1}$$

. By solving this ODE system, the solution in terms of density of infected is:

$$i(t) = \frac{i_0 e^{\beta t}}{1 - i_0 + i_0 e^{\beta t}}$$

with i_0 being the initial fraction of infected over the entire population. The resulting behavior of $i(t)$ is a Sigmoid function: during the early phase the growth is exponential and ruled by β, then the slope decreases because there are less susceptible to be infected and then it saturates to 1, all hosts end up getting infected.

SIS Model

Figure 1.4: Scheme of SIS compartmental model.

On the opposite side with respect to the SI model, where there is no possibility of recovering from the infection, there is the SIS model. In the SIS, after being infectious for a period, hosts can heal and return to be susceptible to that disease. The infection does not confer immunity to the hosts. Thus the transitions described are two as shown in Figure 1.4:

- the contagion, which is mediated by the infected people through the rate $\beta\langle k\rangle\frac{I}{N}$;
- the healing from the infection, that happens spontaneously at the recovery rate $\mu = \tau^{-1}$ which is the inverse of the infectious period.

The ODE system describing the SIS dynamics is expressed by Equation 1.2:

$$\begin{aligned}\frac{ds}{dt} &= -\beta\langle k\rangle si + \mu i \\ \frac{di}{dt} &= \beta\langle k\rangle si - \mu i\end{aligned} \tag{1.2}$$

which gives as a solution for the density of infected again a Sigmoid-like behavior:

$$i(t) = i_0\frac{(\beta-\mu)e^{(\beta-\mu)t}}{\beta-\mu+\beta i_0 e^{(\beta-\mu)t}}.$$

Since there is the possibility of recovering, $i(t)$ does not saturate anymore at 1 like in SI case, but at a value that can be retrieved by searching for the stationary states $\frac{di}{dt} = 0$. The results are the no infected scenario ($i(t) = 0$) which is a stable equilibrium if $\beta < \beta_c = \frac{\mu}{\langle k\rangle}$, but becomes unstable if $\beta > \beta_c$, and the dynamical equilibrium at $i(t) = \frac{\beta\langle k\rangle-\mu}{\beta\langle k\rangle}$, that is feasible only in the case of $\beta > \beta_c$. That situation is called *endemic equilibrium* and represents the case in which the number of new recovered equals the number of new infections.

During the early phase of the spreading $s \approx 1, i \ll 1$, thus $i(t)$ can be approximated with $i(t) \sim i_0 e^{(\beta\langle k\rangle-\mu)t}$. Consequentially the initial transient can determine two different results for the spreading process:

- the case $\beta\langle k\rangle < \mu$ leads to a decreasing dynamics that ends up at zero infected;
- the case $\beta\langle k\rangle > \mu$ leads to an increasing dynamics.

Again, β_c is a threshold value, it is called *epidemic threshold* since it establishes whether an epidemic grows exponentially or goes extinct. Thus it is the minimum value of the infection

probability for which the disease survives. The pathogen's survival condition can be rewritten as:

$$\frac{\beta\langle k\rangle}{\mu} = \beta\langle k\rangle\tau = R_0 > 1. \tag{1.3}$$

This defines a fundamental quantity in epidemiology, the *basic reproductive ratio* $R_0 = \beta\langle k\rangle\tau$. By definition R_0 is the number of cases generated by a case in a fully susceptible population. In other words, the epidemic threshold condition becomes: 'assuming that the entire population is initially susceptible ($s(0) = 1$), a pathogen can invade only if $R_0 > 1$'.

SIS dynamics is mainly used to model a disease in which individuals have a very short temporary immunity period after which they become susceptible again to the infection. For example, this assumption is reasonable for sexually transmitted infections like chlamydia [13] and gonorrhoea [14], where there is hardly any immunity and repeated infections are common.

SIR Model

Figure 1.5: Scheme of SIR compartmental model.

The majority of the infectious diseases leave the hosts with an immunity to a secondary infection. In terms of compartmental models this is described with a SIR dynamics. The transitions involved are the same just illustrated for the SIS model, but in this case another compartment is present, *i.e* the recovered (Figure 1.5), where hosts have gained lifelong immunity to the disease from which they healed. Such dynamics well represent the one of measles [15] and chickenpox [16] that can infect only once in a lifetime. Otherwise, if the temporal window observed is limited, then the majority of the infectious diseases can be approximated with a SIR dynamics, even those with limited immunity duration. For instance, for influenza, COVID-19 and many other respiratory infections immunity lasts several months or even years. Thus, if we aim at describing an epidemic during the course of a winter season, the SIR model provides a convenient approximation.

The SIR dynamics is the one described by Equation 1.4:

$$\begin{aligned} \frac{ds}{dt} &= -\beta\langle k\rangle si \\ \frac{di}{dt} &= \beta\langle k\rangle si - \mu i \\ \frac{dr}{dt} &= \mu i \end{aligned} \tag{1.4}$$

where $r = \frac{R}{N}$ is the density of recovered individuals. Despite its extreme simplicity, this model can not be solved explicitly to obtain analytical expressions for $s(t)$ and $i(t)$. Thus, here it is reported the idea of what happens at the beginning and at long-term. Since i is affected by the same mechanisms introduced for the SIS model, at the initial stages the two dynamics

are similar and there is the same threshold behavior at $\beta_c = \frac{\mu}{\langle k \rangle}$, as the SIS case. Only if the pathogen has $R_0 > 1$ it can invade the population, otherwise it dies out. The asymptotic state can be retrieved by following all the analytical steps explained in [5] under the assumptions of $t \to \infty$ and $r_0 = 0$:

$$1 - r(\infty) = s_0 e^{-r(\infty)R_0}. \tag{1.5}$$

The epidemic terminates without reaching the whole population. There is a portion of the people that remains susceptible, and the epidemic declines because infected individuals do not find enough susceptible to infect. It is important to notice that the final value $r(\infty)$ is independent of the specific values of β and μ, it is a function of R_0 only.

If a portion of the population ν is initially immune to the pathogen, either by vaccination or by previous infection, it can be included in the system in this way:

$$\frac{di}{dt} = \beta\langle k \rangle (1-\nu) si - \mu i \overset{s(0)\sim 1}{\approx} \beta\langle k \rangle (1-\nu) i - \mu i.$$

The concept of basic reproductive ratio can be extended to the *reproductive ratio*: $R = \beta\langle k \rangle \tau s$, the number of cases generated by a case in a partially susceptible population. Merging together the definition of R_0 and the fraction of immunized, R can be rewritten as $R = R_0(1-\nu)$. If

$$R_0(1-\nu) = 1 \to \nu = 1 - \frac{1}{R_0}$$

the epidemic does not start. This condition is called *herd immunity threshold*. It shows that vaccination can be used to reduce the proportion of susceptible below $\frac{1}{R_0}$ and hence eradicate the disease. In general, R_0 can be estimated from surveillance data and interventions can modify its value and thus the progress of the epidemic.

SEIR Model

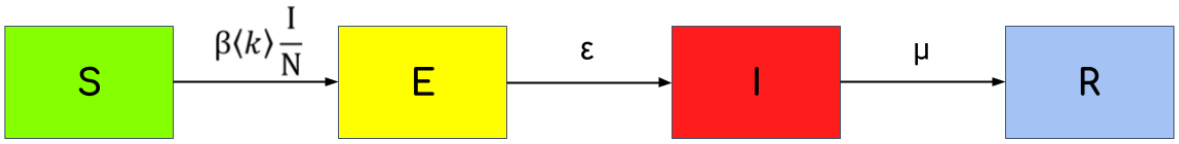

Figure 1.6: Scheme of SEIR compartmental model.

As seen in section 1.1, when a host contracts an infection, there is a period between the time zero of the contagion and the moment when it becomes infectious. That is the exposed (or incubation) period, during which the pathogen abundance grows inside the host organism to reach the threshold abundance level for the host to be infectious. In many cases, like the ones just presented, this passage can be neglected in order to simplify the description. But in the case of a longer incubation period, this can be explicitly modeled by adding an exposed compartment. An example where the SEIR model is adopted because it well suits the disease characteristics is the Ebola case, which has an average incubation period of 9 days [17].

The scheme of the SEIR dynamics is shown in Figure 1.6 and its representation from the

mathematical point of view is in Equation 1.6:

$$\begin{aligned} \frac{ds}{dt} &= -\beta\langle k\rangle si \\ \frac{de}{dt} &= \beta\langle k\rangle si - \epsilon e \\ \frac{di}{dt} &= \epsilon e - \mu i \\ \frac{dr}{dt} &= \mu i \end{aligned} \tag{1.6}$$

where $e = \frac{E}{N}$ is the density of exposed individuals and ϵ the inverse of the latent period. The resulting dynamics behaves similarly to a SIR, with the same final equilibrium state found in Equation 1.5. However the SEIR model has a slower growth rate, since individuals need to pass through the E compartment before they can contribute to the transmission. The exposed stage adds a delay to the dynamics.

Generalizations

Compartmental models are extremely versatile, they can be easily adapted to fit different epidemic situations, following the knowledge about the spreading mechanism of each pathogen. To the basic frameworks just exposed, more state-variables and transitions can be added. Some examples can be:

- Waning of immunity: many viruses have an immunity duration that is temporary. Once it is waned the hosts return to be susceptible. The most appropriate compartmental model to describe this dynamics is the SIRS. For example, influenza viruses fit in this situation [18].
- Birth and death rates: if the population can not be considered constant over the observational period, birth and death events must be included in the dynamics. Births add susceptible and deaths can remove individuals from each compartment.
- Risk structure: usually pathogens do not affect all the hosts with the same intensity. To have a more accurate description of the population, risk structures can be included in the model. In childhood diseases, such as measles, different age cohorts have different risk of contracting the infection. This can be accounted for by adding to the compartmental model compartments corresponding to the different age groups and by assuming they have a different risk of getting infected[19].

In addition to compartmental models, there exist other epidemiological modeling frameworks, such as contact networks, metapopulation models and agent based models. They have an increasing level of complexity: contact networks include into the spreading dynamics of a disease, the specific structure of contacts among individuals; metapopulation models include spatial heterogeneities and mobility; agent based models are a complex combination of multiple agents interacting among them and with the environment, according to specific rules. With these models it is possible to reach very accurate results, but still compartmental models are widely used: they are more flexible and less computational intensive, thus they are more suitable to be integrated in computational intensive studies, such as the ones involving data fitting.

1.3 Use of compartmental models

In the previous section, compartmental models have been introduced from the mathematical point of view. Their versatility has been claimed and in this section some of their practical applications are finally presented.

Compartmental models have two main roles: understanding and predicting. They can be used to understand real world disease spreading. Building them from an individual-level knowledge of epidemiological factors, compartmental models give back the population-level epidemic dynamics. Such dynamics can be affected by many factors and by the means of models it is possible to identify the effect of each factor on the resulting dynamics, increasing our understanding of the infection under study or a specific epidemic. On the other hand, compartmental models have a great predictive power. They are useful to simulate the transmission dynamics to anticipate the future course of an epidemic and to compare possible scenarios in response, *e.g.* intervention measures.

Here are some examples of public health issues that compartmental models can help to address.

Biological understanding. Models can be used to test different biological hypotheses to see which better describe available data. There are multiple influenza subtypes circulating among the population simultaneously every season (this concept is further explained in section 2.1). In the study [20] the authors develop a compartmental model that comprehends all these viruses and their possible interactions. They fit with the model flu incidence data collected in Hong Kong in the period 1998-2018. From that, they are able to estimate the strength of the cross-immunity between each virus-pair, the timing and frequency of changes in population immunity in response to antigenic mutations in influenza viruses, and key epidemiological parameters.

Outbreak analysis and forecasting. By analyzing outbreaks it is possible to obtain important information on the virus causing it to quantify the intensity of the epidemic and possibly make forecasts about the future spreading. This has a fundamental role especially when new pathogens emerge, whether they are new strains or completely new viruses. The most recent example is the COVID-19 pandemic. A new virus from the human coronavirus family emerged and attacked the population whose immune system was completely unprepared. During the early phases of the spreading, there were no clues about the virus' characteristics and analysis on available data have been fundamental to quantify the importance of such a phenomenon. In the work [21], they provide estimations of R_0, per-day mortality and recovery rates about the early phase of the COVID-19 outbreak in Hubei, China, through a compartmental model having as compartment Susceptible-Infectious-Recovered-Dead (SIDR). By calibrating the parameters of the SIRD model to the reported data, they also attempt to forecast the evolution of the outbreak. For example, Figure 1.7 is taken from the original paper and shows the predictions on the cumulative number of infected by COVID-19 (red solid line), from their SIRD model calibrated on the available cases data at that time (red dots).

Vaccination study. The goal of influenza vaccination is to maximize health benefits through efficient use of limited resources. According to World Health Organization recommendations, influenza vaccination programmes have targeted individuals older than 65 years and those at risk. In [22] the authors develop a method that exploits a SIRS model to evaluate how changing target populations in the seasonal vaccination programme would affect infection rate and mortality. In Figure 1.8 they show the results of their study: they compare the consequences on morbidity and mortality of the actual vaccination strategy (in red, 70% vaccine coverage among people 65+) to an hypothetical extension to the age group of 5-16 years (in cyan, 70%

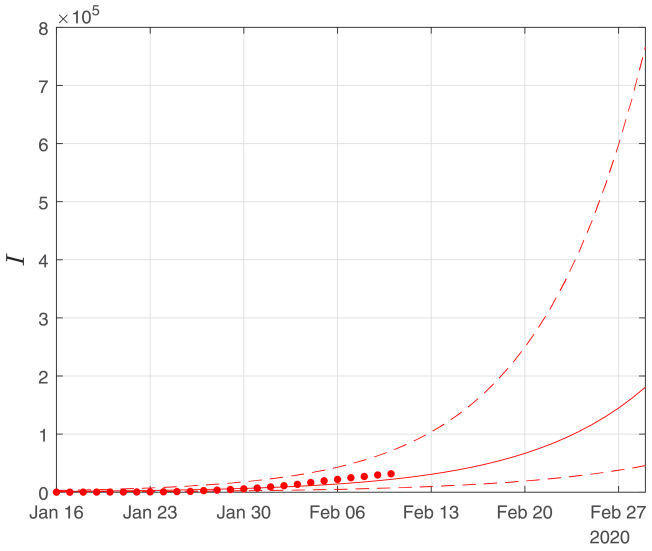

Figure 1.7: Figure from [21]. Simulations until the 29th of February of the cumulative number of infected by COVID-19 in Hubei as obtained using the SIRD model. Dots are the confirmed number of cases, the solid line is the model prediction and the dashed lines are the lower and upper bounds.

vaccine coverage among people 65+ and 30% among children aged 5-16).

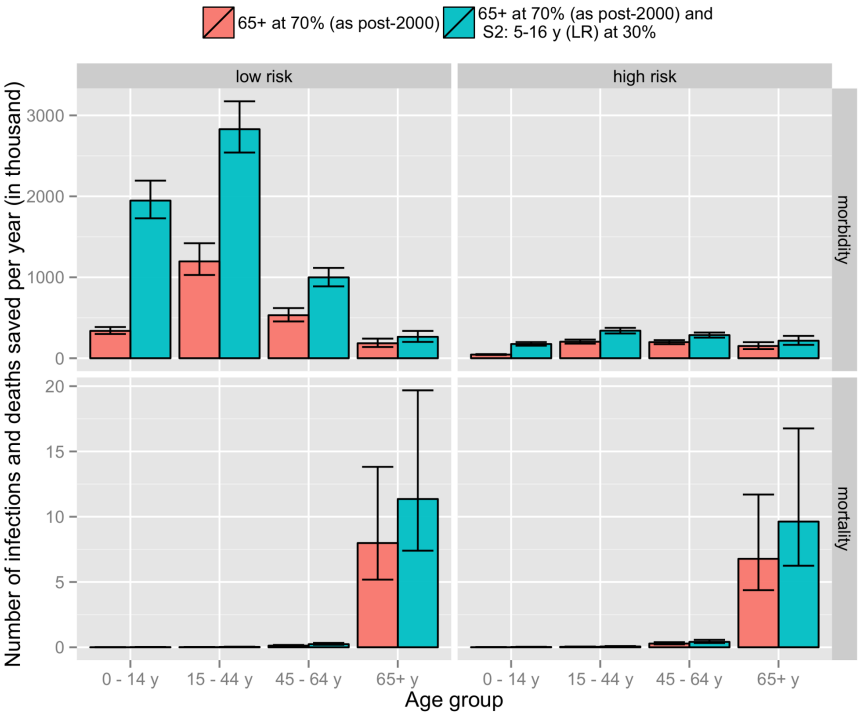

Figure 1.8: Figure from [22]. Comparison between two influenza vaccination strategies: in red 70% vaccination coverage among people 65+, in cyan 70% vaccination coverage among people 65+ and 30% among children aged 5-16. The effects are shown in terms of the number of infections and deaths saved per year.

Another application example can be the introduction of a new vaccine. Respiratory Syncytial Virus (RSV) is a respiratory virus gaining mounting interest, since it is the most common

cause of acute lower respiratory infection in young children [23] and it constitutes a substantial disease burden in older adults [24]. RSV vaccines are still under development, with some of them being at the trials phase [25]. This leads to debates about the necessity of vaccinating people against RSV. In such a complex system, with a lot of interdependent viruses, what can be the result of suppressing RSV? For example, since there is evidence of interference between RSV and influenza [26][27], could a vaccination against RSV cause an increase in influenza cases? A compartmental model can be designed to simulate influenza and RSV co-circulation and quantify the impact of RSV vaccination.

Plan interventions. Compartmental models can guide difficult policy decisions where there is a trade-off between multiple alternative control strategies. An evident example is the COVID-19 pandemic. During the emergency phase, most countries adopted strong interventions: lock-downs, social distancing and face masks. Many studies, based on compartmental models, have simulated the effect of these interventions on the disease spreading, both anticipating future spread to guide government decisions and retrospectively analyzing real epidemics to evaluate the interventions' effectiveness [28].

Chapter 2

Multi-pathogen systems

In chapter 1 we illustrate how to model the spreading dynamics of an infectious disease on a population, through compartmental models. However in real life the same population is subject to many different pathogens co-circulating simultaneously and aiming to the same biological targets. This opens the possibility of interactions among them. To study and understand the dynamics of a pathogen, it may be necessary in certain cases to account for the presence of other pathogens and to model their interaction. In addition, multi pathogens models can be used to describe co-circulating subtypes or strains of the same pathogen to investigate disease evolution.

This work focuses on respiratory viruses epidemics. Despite many studies claiming the presence of interactions among them [4][29], very few mathematical models integrate the presence of different viruses in the same system. Thus, it is still not clear if these interactions are important to describe real data, to quantify the public health impact of these infections and to plan interventions. The aim of this project is to try to fill this gap.

In this chapter we provide an overview on the respiratory viruses co-circulating in the human population and causing seasonal epidemics (section 2.1) as well as on their interactions (section 2.2). Then, two different approaches to tackle multi-pathogen and multi-strain systems are exposed with their strengths and limitations (section 2.3, section 2.4).

2.1 Subtypes, strains and different viruses

There are many different respiratory viruses co-circulating. They aim at the same biological targets (upper and lower respiratory tracts), their symptoms are very similar (cough, sore throat, fatigue, ..) and they attack the same age cohorts (children, elderly people). Taken together these viruses cause the *Influenza Like Illness* (ILI), a syndromic condition often studied aside from its etiology [30]. During recent years, with the evolution of the multiple PCR test, it is becoming possible to test patients for different viruses with a single test [4][29], improving the investigation of ILI composition and the monitoring of viral spreading in the population. Common viruses contributing to ILI include influenza virus, respiratory syncytial virus (RSV), human parainfluenza virus (HPIV), coronaviruses, human metapneumovirus, respiratory adenovirus (AdV) and rhinovirus (RV) [31][32]. An example of the ILI composition found in the study [29] is reported in Figure 2.1.

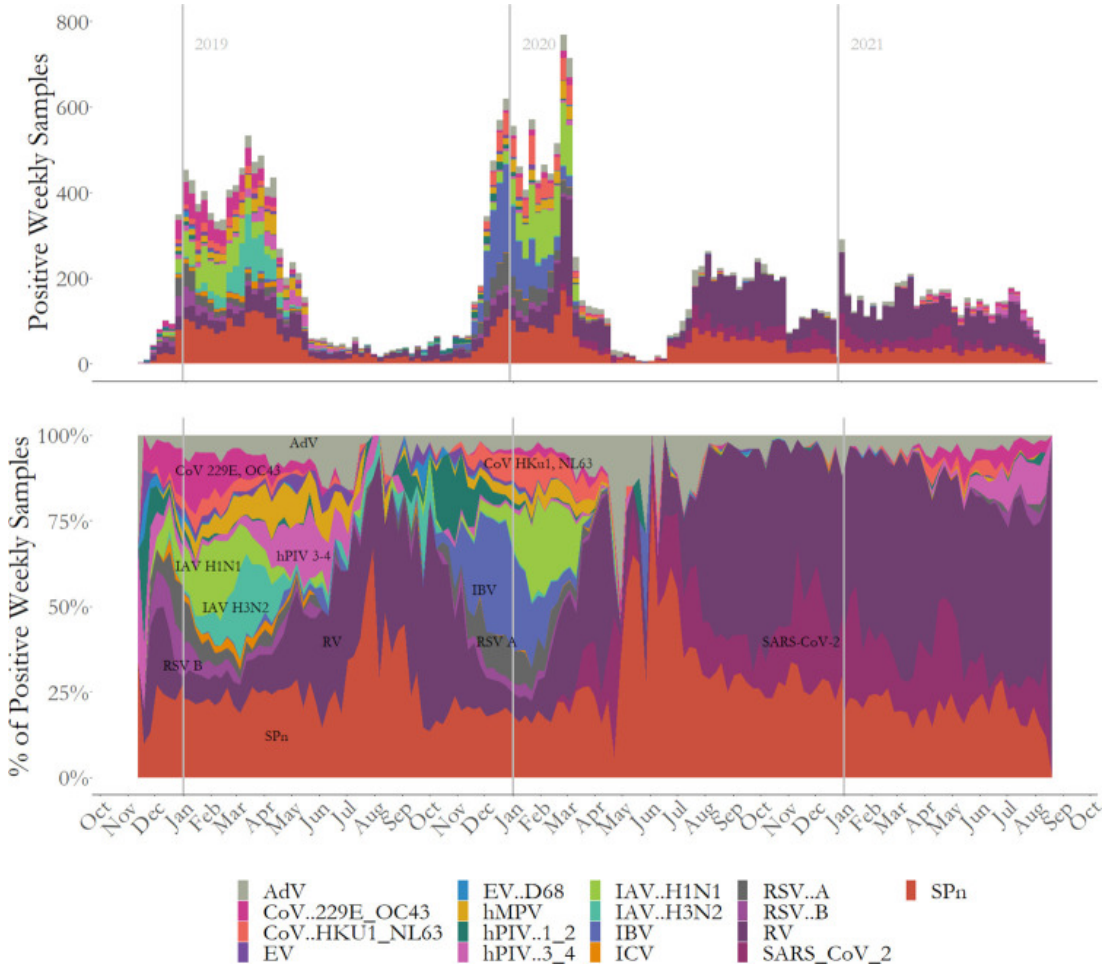

Figure 2.1: The upper panel shows the number of weekly samples positive to ILI viruses and their subdivision into different viruses. The lower panel shows the weekly positive as a percentage of the total. Data and figure from [29].

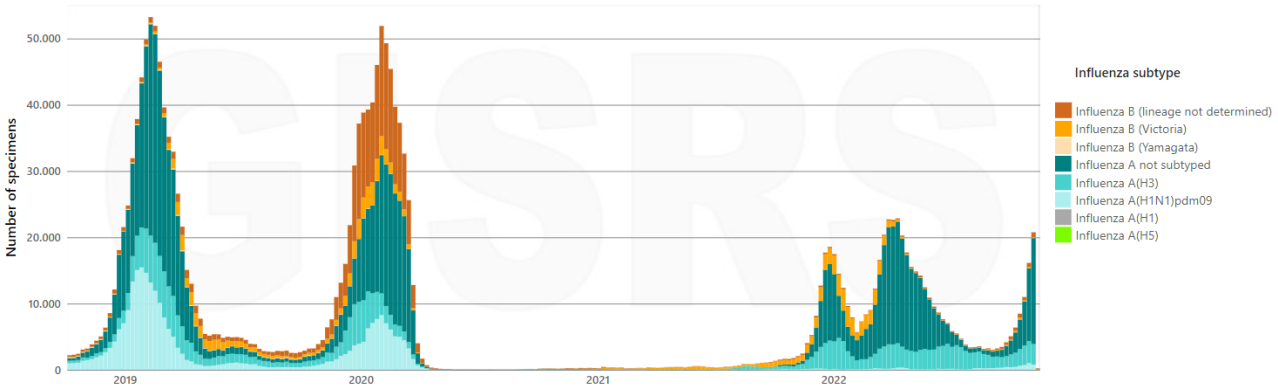

Figure 2.2: Number of influenza detected cases from global surveillance data [33]. Specimens are categorized by influenza A subtypes and B lineages. During COVID-19 pandemic flu reported cases dropped to almost zero.

It is important to highlight that most of the ILI viruses have a further internal subdivision into different subtypes. This distinction is necessary because due to their genetic and antigenic dissimilarities they can be considered as different viruses. We illustrate the case of influenza as an example.

There are four types of influenza viruses, called A, B, C and D [34]. Only types A and B cause the most human illnesses and are responsible for influenza season each year. Whereas influenza C causes mild illness and not human epidemics and influenza D spreads mostly among cattle. About influenza A viruses, they are divided into subtypes, based on different proteins present on the surface of the virus, hemagglutinin (H) and neuraminidase (N). More than 130 influenza A subtypes combinations have been identified in nature and recombinations continuously occur. The subtypes that most frequently circulate among humans are A(H1N1) and A(H3N2) [35]. Regarding influenza B viruses, they are classified into two lineages: B/Yamagata and B/Victoria. Then, there is a further subdivision of influenza A subtypes and B lineages, into genetic clades and subclades. Such subdivision is based on the similarity of genetic mutations and it groups viruses that come from the same common ancestor virus. Since influenza viruses are constantly changing, it is important to always keep track of the variants in circulation to be able, for example, to update the vaccine composition. Seasonal flu vaccines are formulated to protect against influenza viruses known to cause epidemics. By analyzing the composition of the spreading viruses it is possible to forecast the next season circulating viruses, to create specific and more effective vaccines. Figure 2.2 shows the distribution of flu types and subtypes found from recent years global surveillance data. From the situation depicted it is also interesting to note that there is a drop in flu cases reported, corresponding to non-pharmaceutical interventions against COVID-19.

Mutations happen by the means of two different mechanisms: the antigenic drift and the antigenic shift [36]. The *antigenic drift* consists of mutations in the genes of influenza viruses that can result in changes in the surface proteins. These surface proteins are *antigens*: they are recognized by the immune system and can trigger an immune response to generate antibodies to block the infection. Antigenic drifts happen continuously with the replication of the virus among hosts. If these changes are small, the new *variant* of the virus is similar to the parental one, maintaining similar antigenic properties. This means that the antibodies created by the immune system after an infection, are capable to recognize and respond also to antigenically similar viruses. Otherwise, when a change happens in an important location of the proteins or many mutations accumulate, the resulting virus can have very different properties or behavior. In this case the new variant is called *strain*. This can lead to a change also in the antigenic properties and therefore a new strain can trigger a weaker response from the original virus's antibodies. The second mechanism is the *antigenic shift*: it refers to strains recombination that generates new subtypes having a mixture of the original strains' antigens. This includes also the situation in which a flu virus circulating and attacking animals gains the ability to infect humans. The resulting viruses are completely new for the population and their immune systems are not prepared to defeat them. Indeed, when antigenic shift occurs, there can be huge consequences on the population. This was the case of influenza A(H1N1) in 2009 that originated from swine flu and made a jump between species [37]. Human population was not protected from such a virus and the spreading was quick and massive causing a pandemic. Type A viruses undergo both antigenic drift and shift and are the only flu viruses known to cause pandemics, while flu type B viruses change only by the more gradual process of antigenic drift.

Most of the other ILI viruses are structured in subtypes and are subject to mutations, leading to a large number of viral agents co-circulating in the human population and with them a

variety of interactions occurring among them. It is necessary to distinguish ILI viruses and to study their co-circulation as different viruses to better anticipate seasonal epidemics and plan targeted interventions.

The next section is dedicated to the interactions that can occur among respiratory viruses from both the points of view of biology and epidemiological modeling.

2.2 Respiratory viruses interactions

When speaking about 'interaction' in the context of infectious diseases, the reference is to any process caused by an infection by one pathogen that affects infections by other pathogens. Generally interactions can be in both directions: they can be either competitive or synergistic, depending if one infection obstructs or facilitates the other. In the case of respiratory viruses they seem to be mainly competitive, there are many studies showing negative interactions among them, for example: influenza-influenza, influenza-RSV, influenza-RV, influenza-HPIV, coronaviruses-RV [4],[29],[38]. Moreover they can occur at several scales and can be summarized in [3]:

- cellular-level interactions: direct interactions between viral products, altered receptor presentation, modification of release of immune system mediators, competition for host resources among the pathogens;
- host-level interactions: change of transmissibility due to symptoms, individual variation in commensal microbiota, effect of symptomatic responses to infection, tissue damage, competition for host resources, immune cell-mediated interaction, immune signaling-mediated interaction, antibody-mediated interaction;
- population-level interactions: behavioral responses to disease, medication, vaccination.

Interactions at any level can have effects on epidemic spreading at the entire population scale and on public health projections, affecting spatial and temporal spreading patterns. Thus, even if it is still not known the precise biological nature of each single virus-virus interaction, when such a system has to be described with compartmental models, the effects of the interactions can be modeled in a general way at the population-level. To simplify the description of this scenario particular assumptions about conferred immunity and interactions are made.

The mechanism through which after an infection with a virus, the host is protected from a further infection by related strains or even different viruses is called *cross-immunity*. The immunity gained after the first infection protects against viruses other than the one infecting first. Of course, the level of such cross-protection depends on the epidemic context under study. For example, if the viruses considered are different strains from the same parental virus, *i.e.* mutations that are genetically similar but they behave differently, or they are different subtypes, the cross-immunity level should be high, due to their similarities. On the contrary, if the two pathogens are completely different viruses, such protection should be lower or even not present at all. When many cross-reactive diseases are circulating within the population, different assumptions about the nature of the cross-interaction have different effects on the modeling outcome. Below are presented some of the assumptions that can be made when describing interactions between two viruses [5].

A complete cross-immunity between two viruses confers lifelong immunity to both. Modeling both spreading dynamics with SIR models, the two recovered compartments have to coincide.

Based on the previous explanation, such a high level of cross-protection suits the situation of two different strains or subtypes from the same virus. What happens is that since both strains are competing for the same limited resources, the susceptible hosts, only one strain will dominate. The dominating one is the one which is able to use resources more efficiently. The one having the highest R_0 leads the other to extinction. This has evolutionary implications: in theory, any mutation that generates new strains with larger R_0 is favored in such a system. Thus, both the transmissibility β and the infectious period μ should increase and the mortality, if applicable, should decrease. However in reality there is a trade off between transmission and virulence (more transmissible means more pathogen particles and more harmful to the host) and between transmissibility and duration of infectious period (highly transmissible pathogens are often of short duration).

In some cases there is no cross-immunity, but co-infections can not occur. Such an assumption is plausible considering a reduced number of contacts when ill. In this case pathogens can coexist as long as their basic reproductive ratio is greater than 1. This applies to pathogens that are not supposed to be related through immune-mediated interactions. However, if they attack the same population, it is possible that a change in hosts behavior leads to competition and interference between pathogens dynamics. An example, among childhood diseases, could be the measles and whooping cough dynamics [39]. The first one is caused by a virus and the second by a bacterium thus they are considered to be unrelated. However they target the same age cohorts. For this reason the quarantine during the infection and the following convalescent period render a portion of hosts temporarily unavailable to the second disease, leading to pathogens dynamics interference.

Eventually, partial cross-immunity describes the case in which after recovering from an infection caused by a pathogen, an individual gains a partial protection against other infections. It is the most common modeling framework because it can be adapted to suit various viral systems. The strength of the cross-protection depends on the viruses and can be translated in the amount of hosts that gain immunity and its duration. The more the viruses are similar, the stronger the cross-protection: after the first infection the immune system has already prepared a specific antibody response for that virus that can be applied also to similar viruses. This is the case of different strains or subtypes of a virus. On the contrary, if the viruses considered are very different the interaction should be weaker and shorter. Furthermore, the effects of the protection can be modeled as a reduced susceptibility or transmissibility. This can be applied to only a portion of individuals or there can be an homogeneous response on the population. These characteristics make the partial cross-immunity a versatile framework, able to fit systems populated with various levels of interaction.

2.3 Host-view model for any-type interaction among pathogens/strains

As a first multi-pathogen compartmental model we illustrate the one that follows the theory exposed in the book by Keeling and Rohani [5]. In its most general form, this formulation uniquely identifies the entire infection history of individuals within each compartment and their immunological status with respect to the various pathogens under consideration. The sum of the individuals present in all the compartments gives the total population. This approach is called host-view. Such a complete model quickly becomes difficult to study when we have more than 2 pathogens: due to the large number of degrees of freedom the number of compartments

grows exponentially with the number of pathogens.

A basic representative scheme of the spreading dynamics for a 2 pathogen system is shown in Figure 2.3. Susceptible individuals can be infected by pathogen 1 or 2. If during the first

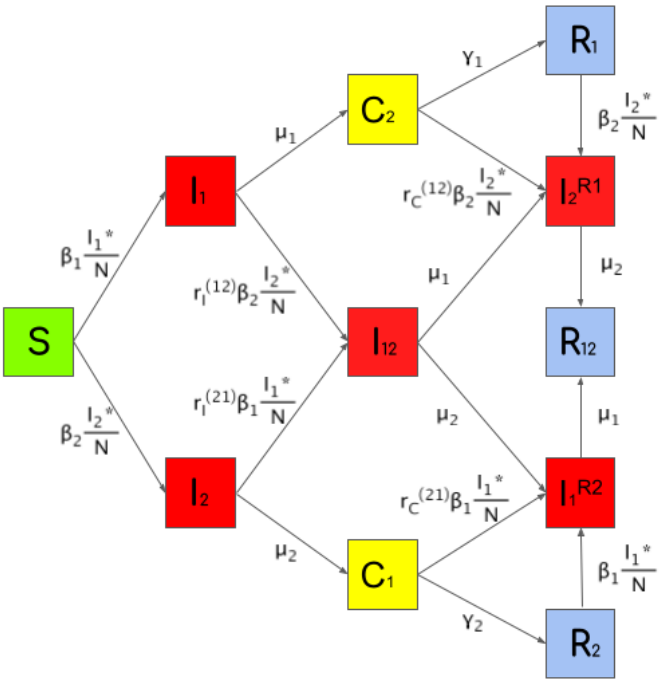

Figure 2.3: Basic scheme of the host-view compartmental model for a 2-pathogen system.

infection they also catch the second disease, then they recover first from one and then from the other. Otherwise, if they are not co-infected, when they heal from the first they stay in a convalescent status that protects them. Then they are finally recovered from the first disease and they can catch the second. Once the infectious period has elapsed, they become recovered to both pathogens.

This type of model is suitable to describe the cross-protection induced by infections from different viruses, thanks to the presence of the convalescent compartment C_i $(i = 1, 2)$. After the infection by j, individuals stay in a convalescent status where they are temporarily protected by further infections. Meaning that this type of immunity is short-term and it is different from the one gained after the infection by i. Compartments C_i and R_i are now differentiated since the original infections are different. This makes this model flexible and able to accommodate weaker interactions as the ones potentially at play between different viruses. In reality, this mechanism could be the result of a post-disease change of behaviors in the individuals, *e.g.* they stay at home to completely recover, or the temporary immune response of their organism that can also affect the susceptibility to other diseases.

The interdependence between pathogens is encoded in the parameters r_I, r_C that act like a reduction in the probability of co-infection or secondary infection after convalescence, respectively. For example, if these parameters are set to 1, the pathogens are independent, the second pathogen can infect independently of the health situation with respect to the first one. While, the lower r_I and r_C, the more the individual is protected to the second pathogen when infected by the first one or convalescent to it. Such mechanism derives from evidence that respiratory virus-virus co-infections are mostly negatively correlated [29].

The dynamics of this model is described by the set of differential equations in Equation 2.1.

$$
\begin{aligned}
\dot{S} &= -(\beta_1 \frac{I_1^*}{N} + \beta_2 \frac{I_2^*}{N})S \\
\dot{I_1} &= \beta_1 \frac{I_1^*}{N} S - (\mu_1 + r_I^{(12)} \beta_2 \frac{I_2^*}{N}) I_1 \\
\dot{I_2} &= \beta_2 \frac{I_2^*}{N} S - (\mu_2 + r_I^{(21)} \beta_1 \frac{I_1^*}{N}) I_2 \\
\dot{I_{12}} &= r_I^{(12)} \beta_2 \frac{I_2^*}{N} I_1 + r_I^{(21)} \beta_1 \frac{I_1^*}{N} I_2 - (\mu_1 + \mu_2) I_{12} \\
\dot{C_1^{(2)}} &= \mu_2 I_2 - (\gamma_2 + r_C^{(21)} \beta_1 \frac{I_1^*}{N}) C_1^{(2)} \\
\dot{C_2^{(1)}} &= \mu_1 I_1 - (\gamma_1 + r_C^{(12)} \beta_2 \frac{I_2^*}{N}) C_2^{(1)} \\
\dot{I_1^{(R_2)}} &= r_C^{(21)} \beta_1 \frac{I_1^*}{N} C_1^{(2)} + \beta_1 \frac{I_1^*}{N} R_2 + \mu_2 I_{12} - \mu_1 I_1^{(R_2)} \\
\dot{I_2^{(R_1)}} &= r_C^{(12)} \beta_2 \frac{I_2^*}{N} C_2^{(1)} + \beta_2 \frac{I_2^*}{N} R_1 + \mu_1 I_{12} - \mu_2 I_2^{(R_1)} \\
\dot{R_1} &= \gamma_1 C_2^{(1)} - \beta_2 \frac{I_2^*}{N} R_1 \\
\dot{R_2} &= \gamma_2 C_1^{(2)} - \beta_1 \frac{I_1^*}{N} R_2 \\
\dot{R_{12}} &= \mu_1 I_1^{(R_2)} + \mu_2 I_2^{(R_1)}
\end{aligned}
\tag{2.1}
$$

. Where S, I_i, C_i, R_i written with the capital letter are the number of individuals present in each compartment. Here we use absolute numbers instead of densities because they are more amenable to stochastic simulations. Blue terms represent the incidence of pathogen 1, and:

$$
\begin{aligned}
I_1^* &= I_1 + I_{12} + I_1^{(R_2)} \\
I_2^* &= I_2 + I_{12} + I_2^{(R_1)}
\end{aligned}
$$

represent the total number of individuals that are infectious with pathogen 1 or 2 respectively.

In Appendix A we illustrate the 2 pathogens complete host-view model and its adaptation to 3 pathogens systems, to give an idea of the amount of mechanisms involved and then to compare them to the model developed during this thesis project.

This model is complete and flexible, thanks to the mechanisms described it is able to account for any-type interaction and can suit systems composed of strains and/or different viruses. However, as highlighted, it can not be scaled to many pathogens. Thus, in the next chapter we present an alternative scalable model present in literature.

2.4 Pathogen-view model for long-lasting interaction among strains

Another framework used to model multi-pathogen systems is the one developed by Gog and Grenfell in [6]. The authors introduce an innovative compartmental modeling framework by adopting the pathogen's view. This means that the population is subdivided into compartments that reflect the individuals' health situation with respect to a single pathogen at a time. The focus is on the current immune state of the host rather than the complex immune history of all his exposures. Thus it is a status-based approach, opposite to the history-based approach previously described with the host-view. Overall the pathogens present in the system are treated independently. Thanks to this characteristic the model can be easily upscaled to n different pathogens co-circulating in the same system, since the number of compartments scales linearly with n. This represents a great improvement on the model complexity with respect to the exponential scaling of the host-view model.

For each of the pathogens $i = 1, .., n$, $j \neq i$ the dynamics scheme in Figure 2.4 is repeated. The sum of the individuals in each i-compartment is N, the total population.

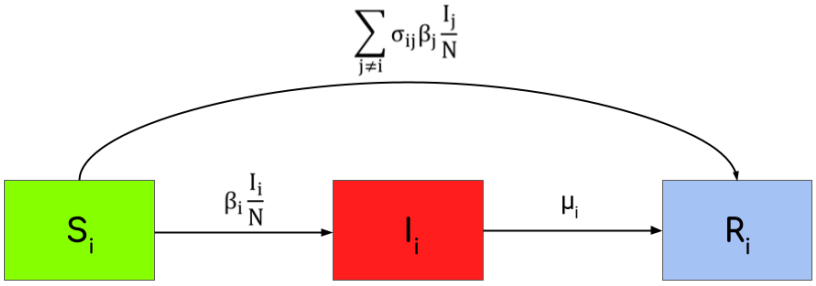

Figure 2.4: SIR scheme of the pathogen-view compartmental model introduced by Gog and Grenfell in [6]. It is valid for each pathogen $i = 1, .., n$, $j \neq i$.

The respective system of ordinary differential equations for each i is:

$$\begin{aligned} \dot{S}_i &= -\sum_{j \neq i} \sigma_{ij} \beta_j \frac{I_j}{N} S_i - \beta_i \frac{I_i}{N} S_i \\ \dot{I}_i &= \beta_i \frac{I_i}{N} S_i - \mu_i I_i \\ \dot{R}_i &= \mu_i I_i + \sum_{j \neq i} \sigma_{ij} \beta_j \frac{I_j}{N} S_i \end{aligned} \tag{2.2}$$

. The basic dynamics described is a simple SIR model, with transitions from susceptible to infected and from infected to recovered. In addition, the interactions between the different pathogens co-circulating are taken into account in the transition from the susceptible to the recovered. When a fraction of the individuals is infected by j, a portion of individuals is removed from the susceptible to the other pathogen i for any $i \neq j$ in the system. The removal of the infected occurs according to the probability σ_{ij} of acquiring immunity to i following an

infection from j. A high value of σ_{ij} means a high cross-protection between i and j, *i.e.* i and j are antigenically similar and j induces strong immune response against i. This immunity mechanism is called *polarized immunity*, it takes into account both the partial cross-immunity and the reduced transmissibility assumptions mentioned in section 2.2: only some hosts gain total immunity when exposed to a partially cross-reacting virus, the others do not gain any immunity from the same exposure. Within the pathogen view, the result is that all hosts have the same chance of gaining immunity to a pathogen, independently of their history of exposure. As a consequence, only one variable is needed to describe susceptibility of a host with respect to each pathogen, greatly simplifying the system and the model.

These removed individuals are moved to the compartment of the recovered from i, meaning that they gain the same immunity as they would have been infected directly by i. This type of immunity has a long-term duration without distinction on the original viral source of the infection. Such assumption is strong and can hold when dealing with pathogens that are very similar and for which it is possible to approximate the immune response of one with the other. This is the case of different strains of the same virus. For example it could suit systems composed of SARS-CoV-2 variants or influenza clades. Indeed, the work in [40] uses a model similar to the one just presented, to describe the co-circulation of three SARS-CoV-2 variants in New York City during winter 2020: the Alpha (B.1.1.7), the Epsilon (B.1.427/B.1.429) and the Iota (B.1.526). During November 2020 the Iota variant was identified for the first time and quickly became a predominant variant in the NYC area. Thus the authors simulate a system considering the already circulating variants of concern (mainly Alpha and Epsilon) and include this emerging one, to characterize its epidemiological properties in relation to the others'.

However if the purpose is to describe systems with different viruses, the model approximation does not hold anymore: the cross-immunity duration is expected to be shorter than the immunity gained after the reference virus infection. This is the main reason for which this model is not suitable to describe multi-pathogen systems, even if it is computational efficient.

Chapter 3

Pathogen-view multi-pathogen model

In the previous chapter we highlight that multi-pathogen models present in literature have some limitations, when used to model co-circulating seasonal respiratory viruses. In particular, one model is flexible, it can include both short-term weak interactions and long-lasting strong ones, but it is not scalable. The other is scalable but not flexible, *i.e.* it models only long-lasting interactions. Thus, for our purposes, we develop a multi-pathogen model that is both scalable and flexible. In particular, we extend the pathogen-view framework to the case of short-term interactions as the ones present among different pathogens. For this reason we will refer to this model as the pathogen-view multi-pathogen model.

In section 3.1 we present our basic model and then we complete it to suit the characteristics of seasonal respiratory viruses epidemics. In section 3.2 the complete model is compared to the host-view one, to check whether the results are coherent between the two views or not. Then, in section 3.3, we retrieve within our framework the incidence of an influenza-like virus, to use it as a reference dynamics in the analysis.

3.1 Pathogen-view multi-pathogen model

The goal of this thesis project is to try to fill the gap of multi-pathogen modelling framework, overcoming the scalability issue and at the same time enabling flexible interaction terms. A model with this purpose needs to be able to capture the key interacting mechanisms between viruses, keeping a simple structure to be scalable. This is the reason behind the choice of adopting the pathogen view for the proposed model. Again, the basic dynamics rely on a standard SIR scheme for each of the viruses present in the system. But the innovative element is to introduce a convalescent compartment C_{ij} as in host-view model. When individuals are infected by j they are removed with a certain probability from the susceptible to any other pathogen $i \neq j$ and they stay in the convalescent compartments C_{ij} for a temporary period q_{ij}^{-1}. Thus, for every i there are $(n-1)$ C_{ij} compartments. The portion of individuals that due to the infection to j are protected from i is regulated by the parameter σ_{ij} and proportional to the incidence of j: $\sigma_{ij}\beta_j \frac{I_j}{N}\frac{S_j}{N}$. This mechanism models the polarized interaction assumption explained in section 2.4. Therefore, the interaction strength between different pathogens is ruled by the σ matrix, which entries are $\in [0, 1]$. When $\sigma_{ij} = 0$ the pathogens are completely independent: the presence of a pathogen j does not affect the dynamics of the others i, *i.e.* individuals infected by j are not removed from the susceptible to i. Whilst as σ_{ij} increases, a

larger portion of the infected by j gains (temporary) immunity also to i, thus the presence of j affects the spreading dynamics of i. In this way the correspondence between the key interaction parameters of this model and the host-view one is straightforward: $1-\sigma$ captures the effect of r_I and r_C of the host-view model, while q is related to γ through: $q^{-1} \sim \gamma^{-1} + \mu^{-1}$.

The basic scheme of this model is shown in Figure 3.1, that has to be repeated for each pathogen $i = 1, .., n$ circulating in the system, with $j \neq i$. It is mathematically represented by Equation 3.1:

$$\begin{aligned}
\dot{S}_i &= -\beta_i \frac{I_i}{N} S_i - \sum_{j=1, j\neq i}^{n} \sigma_{ij}\beta_j \frac{I_j}{N}\frac{S_j}{N} S_i + \sum_{j=1, j\neq i}^{n} q_{ij} C_{ij} \\
\dot{I}_i &= \beta_i \frac{I_i}{N} S_i - \mu_i I_i \\
\dot{R}_i &= \mu_i I_i \\
\dot{C}_{ij} &= \sigma_{ij}\beta_j \frac{I_j}{N}\frac{S_j}{N} S_i - q_{ij} C_{ij}
\end{aligned} \tag{3.1}$$

. The total number of compartments is $[3 + (n-1)]n$: they scale quadratically with n and this is an improvement with respect to the exponential scaling of models that adopt the host-view [6].

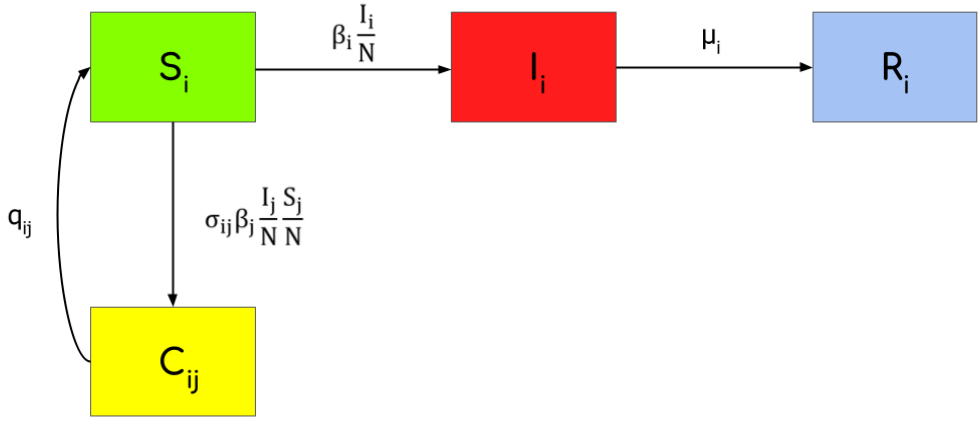

Figure 3.1: Basic scheme of the pathogen-view multi-pathogen compartmental model introduced in this work. It is valid for each pathogen $i = 1, .., n$, $j \neq i$.

3.1.1 Application to respiratory infections

The one just presented is the model with its fundamental basic mechanisms. To make it suitable for real respiratory viruses epidemics a few more aspects have to be added: the seasonality of the viruses, the waning of immunity and the immigration term.

Seasonality

First of all, respiratory viruses are characterized by seasonality. They do not spread uniformly over the year, they generate waves according to environmental variables such as the weather conditions and human behaviors. Dry and cold temperatures cause these viruses to thrive [41]. Moreover, cold weather changes the way human bodies respond to disease and by adding an increased time spent indoors, the result is a higher risk of infection during winter period. For example, in temperate regions viruses like influenza, RSV, HPIV and the human coronaviruses have high spreading ability during winter while they are weaker in summer [4],[42],[43],[44].

Since the parameter β_i encodes the *transmissibility* of the virus i, it is expected to vary over the year. An effective way to model seasonal variation in transmission is to assume β following a sinusoidal with a one year period (Equation 3.2):

$$\beta_i(t) = \beta_{0,i}\left(1.5 + \sin\left(\frac{2\pi}{365}t\right)\right) \tag{3.2}$$

. The parameter $\beta_{0,i}$ tunes the overall level of transmissibility and it is different for different viruses. The center of the sine is set to be 1.5 in order to obtain values of the basic reproductive number ($R_0(t) = \frac{\beta(t)}{\mu}$) in a meaningful range for seasonal respiratory viruses.

Waning of immunity

In SIR model, after acquiring an infection with a pathogen i, once elapsed an average infectious period μ_i^{-1} individuals are no longer infectious. They gain immunity that keeps them protected from a further infection with the same pathogen (recovered compartment). Usually, in the context of respiratory viruses, this immunity does not last forever and its duration can vary from weeks to years, depending on the virus. To model the waning of immunity it is more appropriate to use a SIRS dynamics: to the model it is added a transition that occurs at rate λ_i at which recovered individuals lose immunity and return to be susceptible to i. This leads to an endemic circulation, differently to the SIR dynamics. As anticipated before, in general this type of immunity is longer and stronger with respect to the cross-immunity induced by another virus j, which instead is partial, because developed only by a fraction σ_{ij} of the infected, and of shorter duration. If the virus i tries to attack a recovered host again, it will be defeated by the specific immune response developed by the previous i-infection. On the other hand, the same infection with i can lead to partial cross-protection effects from other j-infections. In this case the mechanism is the one explained in the previous section with the temporary stay in the convalescent C_{ji} compartment. The duration of the two respective immunities, λ_i^{-1} and q_{ji}^{-1}, is different because of the specificity of the immune response.

Immigration term

Seasonality makes the epidemics oscillate over the year and one of the causes is weather conditions. For this reason, the incidence peak occurs at different moments of the year according to different climatic zones. With people traveling across the world the spread of the diseases is favored. This guarantees a global circulation of the viruses even during months with lower incidence. The continuous circulation can be introduced in the model by adding a small constant immigration rate per day: m infected individuals enter the system from the outside and

m susceptible are removed to keep the total population unchanged. In this way, especially with stochastic scenarios, the epidemics persist and do not get extinct.

Complete model

By integrating these three new mechanisms to the basic model, the resulting complete scheme is the one in Figure 3.2.

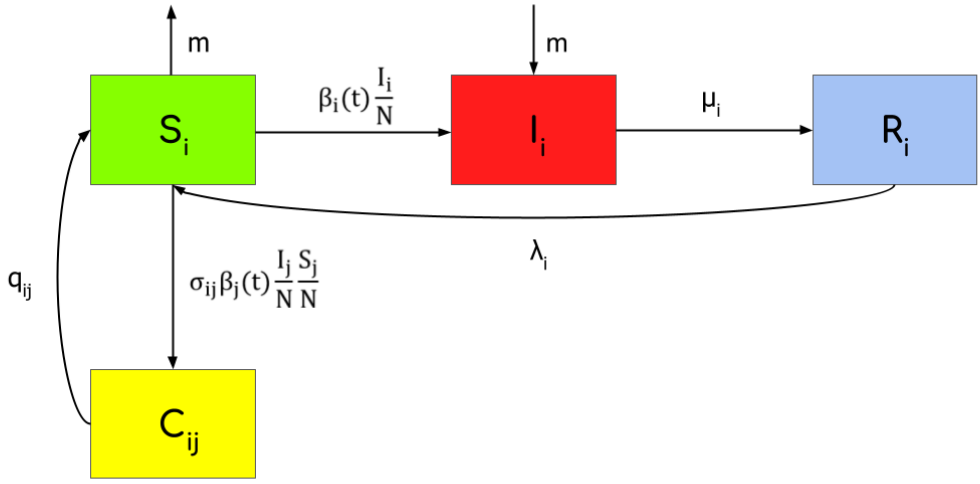

Figure 3.2: Complete scheme of the pathogen-view multi-pathogen compartmental model introduced in this work. It is valid for each pathogen $i = 1, .., n$, $j \neq i$.

The entire spreading dynamics modeled is the following: individuals susceptible to pathogen i (S_i) can be infected by it (I_i) with a probability $\beta_i(t)\frac{I_i}{N}$; after an average infectious period μ_i^{-1} they recover from i, *i.e.* they gain immunity against that virus (R_i). Then, after an average period λ_i^{-1} they lose immunity and become again susceptible to i. The influence of the other infections j on i dynamics is taken into account by removing susceptible to i and temporary protect them by staying in the convalescent compartment C_{ij}. This transition happens according to the rate $\sigma_{ij}\beta_j(t)\frac{I_j}{N}\frac{S_j}{N}$ and lasts for a period q_{ij}^{-1}. These mechanisms are repeated for any other pathogen j present in the system.

Since all the pathogens are handled independently thanks to the pathogen view, in this model it is not possible to keep track of co-infections. The consequences of this aspect will be further investigated in section 3.2.

The dynamics of the entire system is described by the set of Equations in 3.3, for each $i = 1, .., n$,

$j \neq i$:

$$\begin{aligned}
\dot{S}_i &= -\beta_i(t)\frac{I_i}{N}S_i - \sum_{j=1,j\neq i}^{n} \sigma_{ij}\beta_j(t)\frac{I_j}{N}\frac{S_j}{N}S_i + \sum_{j=1,j\neq i}^{n} q_{ij}C_{ij} + \lambda_i R_i - m \\
\dot{I}_i &= \beta_i(t)\frac{I_i}{N}S_i - \mu_i I_i + m \\
\dot{R}_i &= \mu_i I_i - \lambda_i R_i \\
\dot{C_{ij}} &= \sigma_{ij}\beta_j(t)\frac{I_j}{N}\frac{S_j}{N}S_i - q_{ij}C_{ij}
\end{aligned} \tag{3.3}$$

. To study the system dynamics we will monitor the incidence of each pathogen, that by definition is the number of new infections per day. Thus, for this pathogen-view model the incidence of i is:

$$\text{incidence}(t) = \beta_i(t)\frac{I_i(t)}{N}S_i(t) + m$$

3.1.2 Stochastic simulations

Real world epidemics are not deterministic processes. The incidence peaks height varies from year to year because of the variability induced by random effect in transmission. Hence a stochastic version of the model would better resemble real data.

To switch from a deterministic compartmental model to its stochastic version, the transitions between compartments have to be drawn from a binomial distribution. For example, the number of susceptible that gets infected follows a binomial $Bin(S_i(t), \beta_i(t)\frac{I_i(t)}{N})$ and the number of infected getting recovered follows $Bin(I_i(t), \mu_i)$. The immigration term is fundamental to avoid epidemic extinction when dealing with stochasticity. In this case it is no longer represented by a constant rate m, but it is drawn from a Poisson distribution with mean $m = 1$ person/day.

By definition, stochastic results vary from realization to realization. For this reason, to capture a trend in stochastic scenarios, results need to compute the statistics over many years and different realizations. In the setup used for this work, simulations are run for 60 years and repeated for 200 realizations. Since the meaningful part of the dynamics to be analyzed is the stationary one, the first 10 years of transient dynamics are discarded.

During all the simulations performed, the total population considered for the system is $N = 60 \times 10^6$ people and the initial conditions for each compartment are:

$$\begin{cases}
I_{0,i} = 0.05 \times N \\
R_{0,i} = 0.2 \times N \\
Q_{0,ij} = 0 \times N \\
S_{0,i} = N - I_{0,i} - R_{0,i} - \sum_{j\neq i} Q_{0,ij}
\end{cases}$$

for every $i = 1, .., n$ and $j \neq i$. The choice to start simulations with the 20% of population already recovered is made to speed up the convergence to the stationary dynamics. Analogously to the pathogen-view model introduced by Gog and Grenfell, the sum over all the i-compartments gives the total population N, and this is true for every pathogen $i = 1, .., n$.

The parameters that need to be parametrized based on real viruses characteristics are the transmissibility $\beta_{0,i}$, the duration of the infectious period μ_i^{-1} and of the immunity λ_i^{-1} and the parameters related to interactions σ, q. The choice of their values is discussed in detail later.

3.2 Comparison between host-view and pathogen-view model

As mentioned in section 3.1, the pathogen-view multi-pathogen model developed by us (Figure 3.2) can easily scale with n, the number of viruses co-circulating in the system. This is possible since the different dynamics are treated independently. As a consequence, co-infections can not be tracked. This is clearly a convenient simplification but it could also lead to an overestimation of the convalescents with a resulting underestimation of the incidence. For example, when considering more than two pathogens, if a fraction of individuals is co-infected with both pathogens j and k, they could be counted twice in the amount of people that would be protected from i: both in C_{ij} and C_{ik}. This leads to a larger fraction of people temporarily removed from the susceptible to i. Consequently, with less susceptible present there would be less infected individuals, thus the incidence of i would be underestimated. To gain some insight on the consequences of this bias, we compare our pathogen-view model to the standard host-view one, that takes into account every transition and compartment possible, to check if the two results are coherent, despite the differences highlighted above.

The comparison done is not complete and exhaustive, but rather it tests the two models in some specific scenarios under the same assumptions. The two models are compared in the case of systems with two and three pathogens. One pathogen is taken as reference and the other/others are taken with similar, greater and smaller transmissibility. The remaining parameters are kept fixed and equal. Then, incidence peaks height of the reference pathogen are monitored while varying the interaction strength (parameters σ, $r_i = r_c$) and the duration of the cross-protection (parameters q^{-1}, γ^{-1}). All the details are reported in Appendix A. Here there is a quick comment on the results found.

For all the cases tested, the mean incidence peak height of the reference pathogen is computed and then it is done the relative difference between the results from the two models. The differences are always orders of magnitude of 10^{-2} or smaller. This suggests a high similarity between the results of the two models under the cases tested. We can say that the pathogen-view model introduced in this work is able to describe multi-pathogen systems even if it does not track co-infections. This is probably due to the fact that co-infections are rare in real life [29] and thus the previously mentioned incidence underestimation caused by co-infections does not strongly affect our simplified model. Indeed, there are other works suggesting that co-infections have a small impact in modeling these viruses' dynamics. One of them is [45], where the authors build a 2-pathogen host-view model to describe some of the ILI viruses' incidence data, in particular RSV and HPIV. Their model has the same cross-protection mechanism as the one in Figure A.1, but it does not consider co-infections. Despite this, they show that their model can well explain real incidence dynamics of RSV and HPIV together, suggesting a minor role of co-infections.

3.3 Flu-like pathogen

Once having tested that our model recovers similar dynamics with respect to the host-view model, the aim is to systematically analyze the dynamics resulting from the model when simulating different scenarios. In particular, we want to describe multi-pathogen systems reproducing the qualitative properties of co-circulating seasonal respiratory viruses. Given that influenza is the most studied infection we take it as focal infection and see how the presence of other viruses alters in the model its dynamics under different assumptions.

We describe below in detail how the model is parametrized in order to reproduce realistic scenarios. For influenza parameters we use available data and information in the literature. But since more limited information is available for the other viruses, we assume their transmissibility (β_0) and duration of immunity (λ^{-1}) to be drawn from LogNormal distributions:

$$\beta_0 \sim \text{LogNormal}(\beta_{0,mean}, \beta_{0,std})$$
$$\lambda^{-1} \sim \text{LogNormal}(\lambda^{-1}_{mean}, \lambda^{-1}_{std})$$

. We choose as mean and standard deviation for the distributions, values that can represent the heterogeneity of seasonal respiratory viruses and we make hypotheses on their relation with respect to flu parameters. As shown in Figure 3.3, we decide to have a resulting β_0 distribution that is symmetric with respect to flu transmissibility $\beta_{0,mean} = \beta_{0,flu}$ and $\beta_{0,std} = 0.05$, to reflect the variety of real viruses. While the λ^{-1} distribution is asymmetric with respect to the value of flu, it peaks at lower immunity duration, $\lambda^{-1}_{mean} = 1$ year. The majority of the real viruses have duration of the order of magnitude of months, but to replicate a general scenario, assuming $\lambda^{-1}_{std} = 0.1$ year, we obtain a long-tailed distribution that covers values over several years.

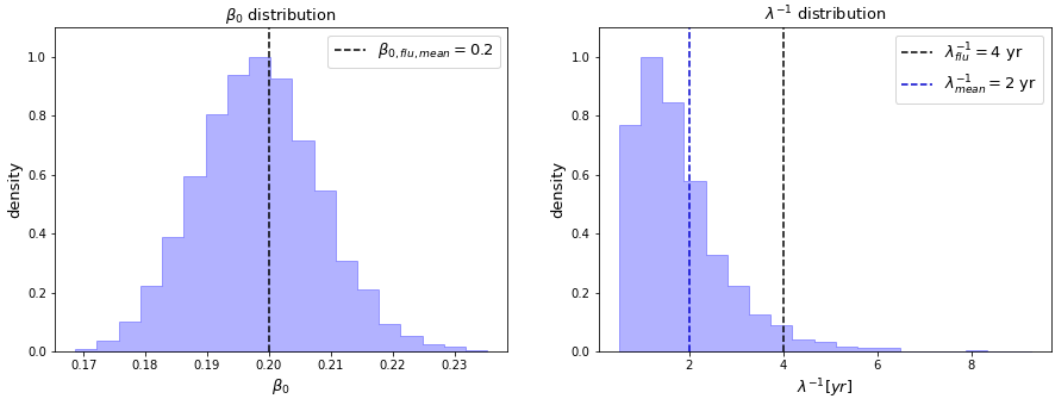

Figure 3.3: LogNormal distributions for β_0 (on the left, LogNormal($\beta_{0,flu}, 0.05$)) and λ^{-1} (on the right, LogNormal($1yr, 0.1yr$) of the generated pathogens circulating in the system.

Then, we select flu mean peak height as the characteristic of flu incidence dynamics that we want to reproduce with our model, to have a well defined quantity to monitor during the analysis. According to literature, the characteristics that a 'flu-like' pathogen has to have are:

- incidence peak in winter [4];
- infectious period: $\mu^{-1}_{\text{flu}} = 4.5$ days [30];

- duration of immunity: $\lambda_{\text{flu}}^{-1} \in [2, 8]$ years [41];
- average weekly incidence peak of 1%. This value comes from the analysis of Réseau Sentinelles data from France [46]: a network of volunteer general practitioners that collects disease cases among their patients, for surveillance and forecasting purposes. These results may not be the same for other countries.

Once set the infectious period and the duration of immunity in the meaningful range, the goal is to find the couple of parameters (β_0, λ^{-1}) that leads to the same incidence mean peak height as flu, when co-circulating with 10 pathogens generated according to Figure 3.3. The choice of considering the co-circulation of 10 pathogens is done to replicate the order of magnitude of the co-circulating seasonal respiratory viruses. Indeed, probably even the values reported in literature or from surveillance data are affected by the others circulation. The interactions among them are set to have $\sigma_{ij} = 0.8$ and $q_{ij}^{-1} = 21$ days for each $i, j = 1, .., n$, with $j \neq i$: a protection of the 80% for a duration of three weeks after each infection. These values are reasonable and compatible with other sources, even if there is limited information about them.

After simulating dynamics under these conditions, the mean weekly incidence peaks height of influenza are collected. The chosen parameters couple is $\beta_{0,\text{flu}} = 0.2$ and $\lambda_{\text{flu}}^{-1} = 4$ years, that leads to a mean peaks incidence around 1%. The resulting basic reproductive parameter stays between $R_{0,\max} = 2.25$ and $R_{0,\min} = 0.45$, which is below the endemic threshold. An example of the flu-like weekly incidence is shown in Figure 3.4: in the upper panel there is the pathogen-view model result, the blue line is one stochastic realization; in the lower panel there is the real flu weekly trend in France, collected by Réseau Sentinelles and available at [46]. As can be seen, the stochastic trend is coherent with the real world dynamics. For example, the maximum values are below the rate of 2000 in both cases and the mean incidence peak height is 1000 people per 100000 people.

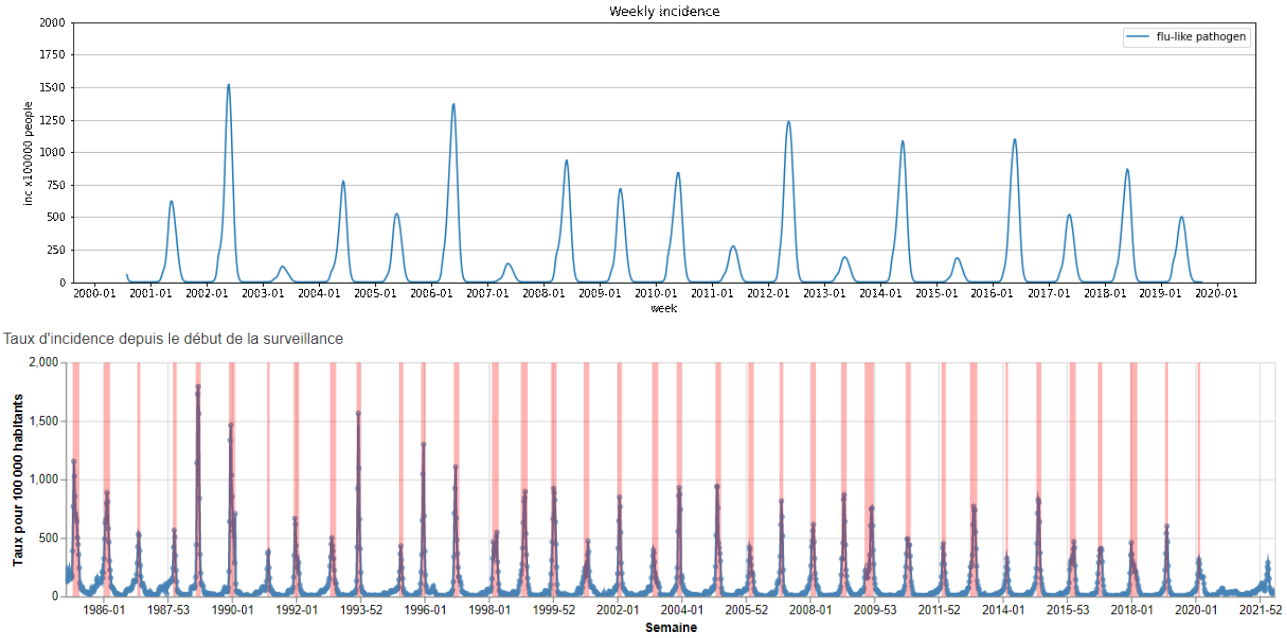

Figure 3.4: Upper panel: flu-like weekly incidence with stochastic pathogen-view model ($\beta_{0,\text{flu}} = 0.2$, $\lambda_{\text{flu}}^{-1} = 4$ years, $\mu_{\text{flu}}^{-1} = 4.5$ days). Lower panel: flu weekly incidence trend in France, available at [46].

In Figure 3.5 there are some examples of stochastic realizations of the incidences resulting

from the model, to better understand the impact of other co-circulating pathogens we compare hypothetical scenarios with an increasing number of pathogens. It is interesting to notice the resulting variability in pathogens' peaks height and position. Even if the initial conditions are the same for all the viruses, stronger pathogens (the ones with higher transmissibility and higher peaks) occur earlier in time: they dominate the other viruses. Then, weaker viruses' peaks follow, with decreasing heights.

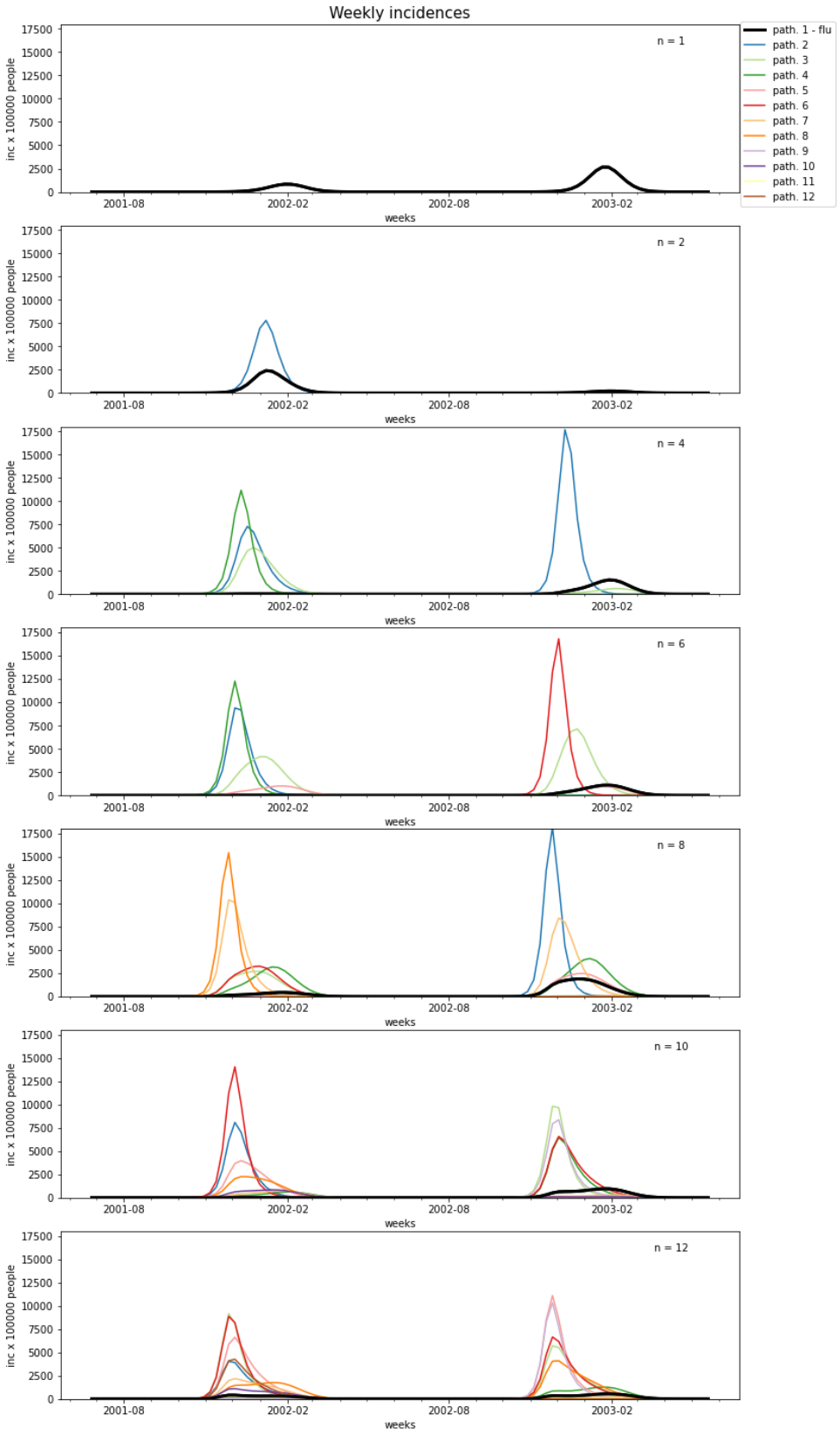

Figure 3.5: Example of stochastic realizations of the pathogens' weekly incidences varying the number of pathogens n co-circulating in the system. The pathogen highlighted in black is flu, the others are generated according to Figure 3.3.

Chapter 4

Analysis

This chapter presents the analysis performed on the complete pathogen-view multi-pathogen model introduced in chapter 3. We take one focal pathogen, *i.e.* influenza, as reference and we address some practical questions:

- What is the effect of the presence of other respiratory viruses on influenza dynamics? (section 4.1)
- What is the effect of the presence of other respiratory viruses on influenza parameters estimation? (section 4.2)
- Is there a relation between the interactions strength matrix σ and the matrix of the correlations between the incidences? (section 4.3)

4.1 Effect of other pathogens on flu dynamics

What is the effect of the presence of other pathogens on influenza dynamics? The dynamics of a pathogen like flu is taken as reference and it is monitored how the incidence peaks are affected by the presence of the others.

Let's take as an example the dynamics produced by the pathogen-view multi-pathogen model in the case displayed in Figure 3.5, where the interactions are assumed to be $\sigma_{ij} = 0.8$ and $q_{ij}^{-1} = 21$ days for each ij pair. In Figure 4.1 we present the respective flu mean peaks height and its peaks height distribution, when it co-circulates with $n-1$ different pathogens. By looking at the plot on the left, it can be seen that assuming a protection of the 80% from 10 pathogens, for a duration of three weeks after the infection, flu incidence mean peaks height is lowered by 60%. From the plot on the right, it is evident the effect of the presence of other viruses on the distribution of the peaks height: starting from the blue curve when flu is alone, by adding pathogens the distribution squeezes towards smaller heights until the pink curve with $n = 12$. This effect of the presence of other respiratory viruses on influenza dynamics is quite relevant and deserves further investigation.

The same analysis is applied to systems with a different number of pathogens and different interaction strengths and duration. Multi-pathogen systems are simulated with $n = \{1, 2, 4, 6, 8, 10, 12\}$, where pathogens' parametrization is always the one described in Figure 3.3.

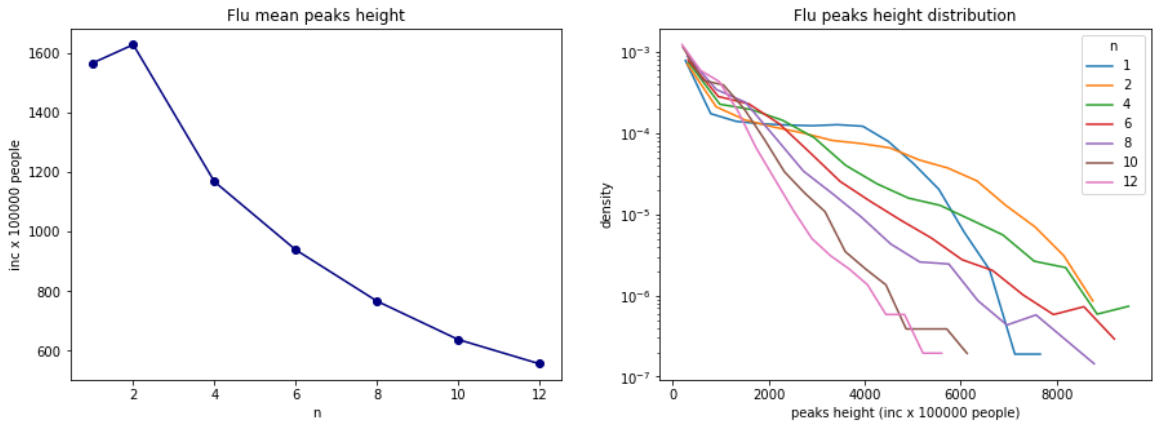

Figure 4.1: Flu mean peaks height and peaks height distribution varying the number of pathogens n co-circulating in the system. Case $\sigma = 0.8$ and $q^{-1} = 21$days.

One pathogen is 'flu-like' and the other $n-1$ have parameters drawn from the LogNormal distribution. About the interaction strength, the entries of σ and q matrices are set to be equal for each couple of pathogens $i, j = 1, .., n, i \neq j$. The tested values are:

$$\sigma_{ij} \in \{0, 0.2, 0.4, 0.6, 0.8, 1\}, \quad q_{ij}^{-1} \in \{7, 14, 21, 30, 60\}\text{days}.$$

The results are presented with the heatmaps in Figure 4.2 for the case varying (σ, n) and in Figure 4.3 when varying (q^{-1}, n): the left plot is the average peaks height of flu, the right plot its coefficient of variation $(= \frac{\text{std}}{\text{mean}})$.

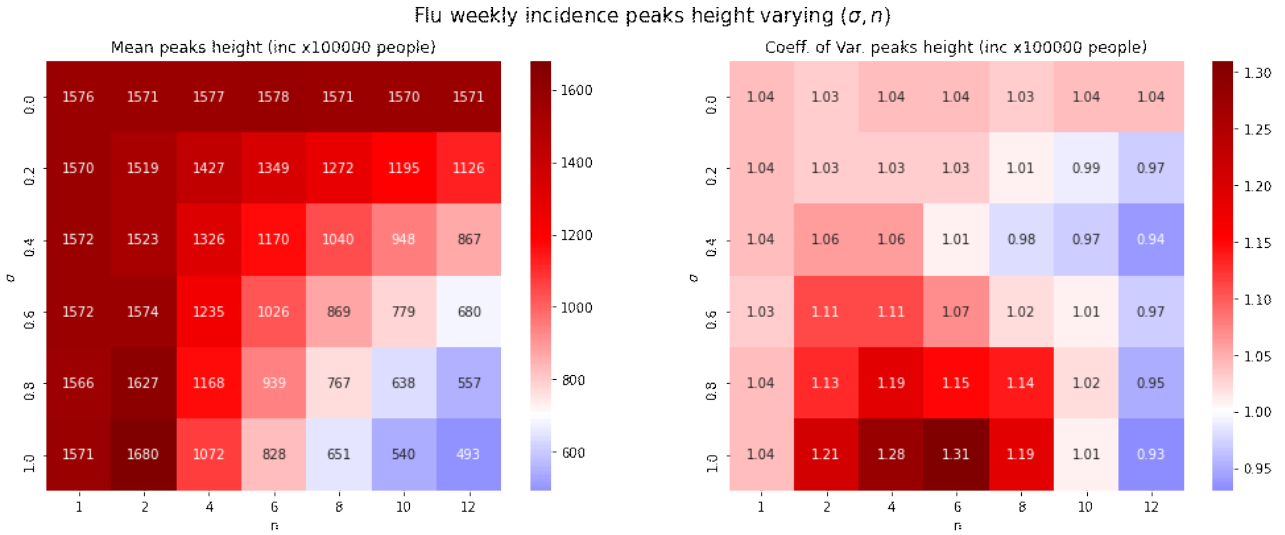

Figure 4.2: Flu weekly incidence peaks height, varying σ and n, fixed $q^{-1} = 21$ days. On the left the mean and on the right the coefficient of variation.

For a clear visualization of the trend varying n, σ, q^{-1}, we also show the curves generated from the heatmaps results. Looking at Figure 4.4, flu weekly mean incidence peaks height is reported as a function of the number of pathogens co-circulating. On the left different curves represent different σ, on the right they represent different q^{-1}. The trend is that, from two pathogens, by increasing the number of pathogens co-circulating, flu incidence peaks lower. There is an

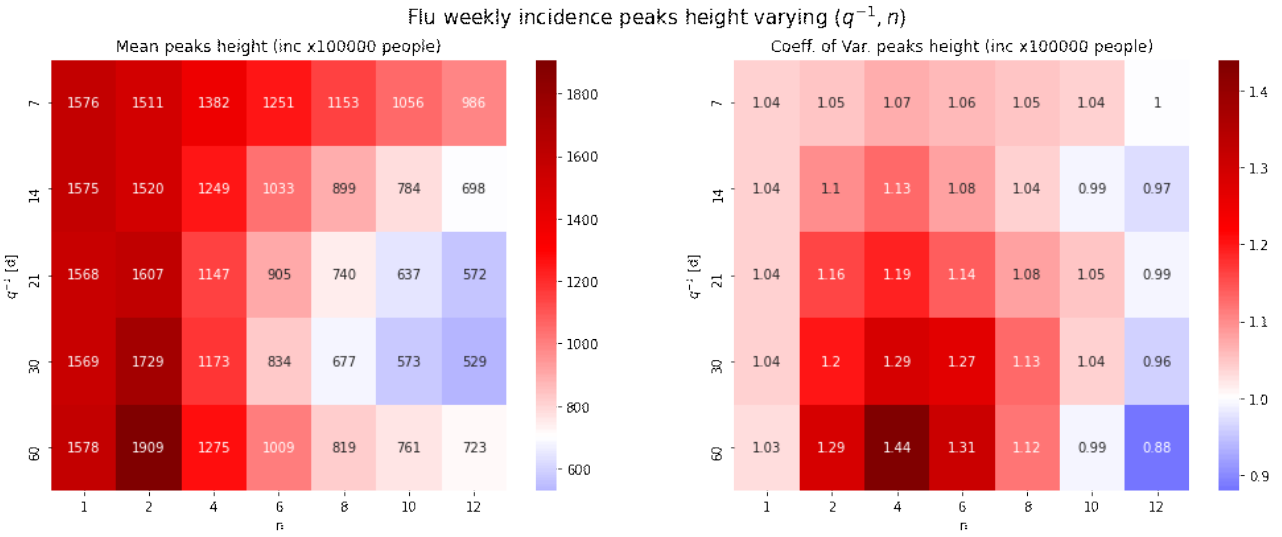

Figure 4.3: Flu weekly incidence peaks height, varying q^{-1} and n, fixed $\sigma = 0.8$. On the left the mean and on the right the coefficient of variation.

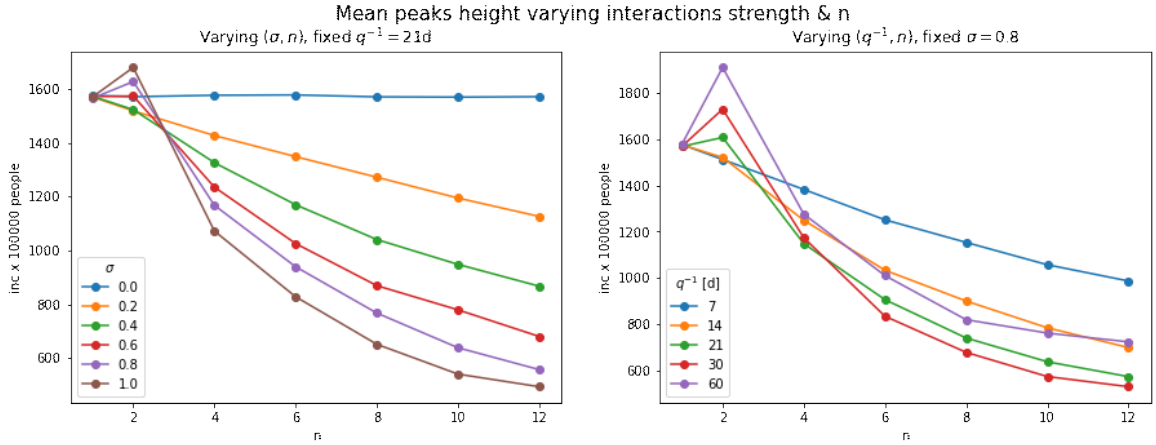

Figure 4.4: Mean peaks height of flu weekly incidence varying the interaction strength and n. On the right σ varies and $q^{-1} = 21$d, on the left q^{-1} varies and $\sigma = 0.8$.

unexpected increase in the peak incidence when passing from one to two pathogens considered. When increasing the strength of interaction (σ) and the duration of the cross protection (q^{-1}) flu incidence peaks lower. Naturally, when $\sigma = 0$, meaning that the pathogens are independent, the mean peaks height remains constant across different n. The coefficient of variation indicates the variability of the peaks height from year to year, thus it is related to the predictability of incidence peak height. We find that it is maximum at intermediate values of n in both Figure 4.2 Figure 4.3.

The overall results of this analysis are expected and natural, apart from the maximum peak height at $n = 2$ and the maximum coefficient of variation at $n = 4$. These lasts are unexpected and should be further investigated. Anyway the magnitude of these effects remarks the importance of considering real systems as multi-pathogen, to have a complete understanding of what really happens.

4.2 Effect of other pathogens on flu parameters estimation

In literature it is possible to find very different estimations of the duration of flu immunity: the usual range goes from 2 to 8 years [41] [47]. Of course this variability can affect whatever analysis or forecast in which this parameter is involved. Thus, it would be very helpful to understand which is the cause of this discrepancy. Since in the previous analysis it has been highlighted that the effect of the presence of other viruses on influenza dynamics can be relevant, it is natural to ask if this could also lead to a bias in parameters estimation. Indeed, to estimate such parameters the presence of other viruses in the same system is normally ignored. This is the question that the framework introduced in this work will try to solve: what is the effect of other pathogens on flu parameters estimation?

When dealing with stochastic realizations, a possible estimator to look at to compare different dynamics can be the peaks height distribution. The goal of the following analysis is to retrieve the peaks height distribution of the 'flu-like' pathogen when it is in a system composed of 10 pathogens (brown distribution in Figure 4.1), with a single-pathogen-system peaks height distribution. To find which (β_0, λ^{-1}) this pathogen has to have to achieve such a distribution, these parameters' space is explored to search for the region with the minimum distance between the two distributions. As distance it is used the Hellinger one, that for two discrete distributions P, Q is defined as:

$$H_{\text{dist}}(P,Q) = \sqrt{\frac{1}{2}\sum_{i=1}^{k}(\sqrt{p_i} - \sqrt{q_i})^2} \qquad p_i, q_i \sim P, Q.$$

The explored parameters region is:

$$\beta_0 \in [0.13; 0.42], \quad \text{step} = 0.002$$
$$\lambda^{-1} \in [100; 6600]\text{d}, \quad \text{step} = 100\text{d}$$

. For each of the points on this grid the single pathogen dynamics is simulated. Each simulation is run for 200 realizations and from them it is retrieved the peaks height distribution. Then it is computed the Hellinger distance between each of these distributions and the one of flu with $n = 10$. The resulting distances are shown in Figure 4.5, where it is superimposed a contour plot to identify the high-interest region, *i.e.* the minimum distance in dark-blue. All the distances are of the order of magnitude of 10^{-2}. The couple of parameters leading to the minimum Hellinger distance is highlighted with a black point, while the original flu parameters are marked with a black x. The peak height distribution of these 'best-match' parameters (blue line) is compared in Figure 4.6 to the $n = 10$ distribution (red line) and $n = 1$ distribution of the respective flu alone (gray line).

From these figures it is clear that the two couples of parameters do not coincide. If the flu-like parameters with $n = 10$ are used in a single pathogen system ($\beta_0 = 0.2, \lambda^{-1} = 4yr$), the peaks height distribution found (gray line) differs from the original $n = 10$ one (red line). The most similar distribution, *i.e.* dynamics, is the one with $\beta_0 = 0.186, \lambda^{-1} = 5yr$. Meaning that if the aim is to retrieve the dynamics of flu with $n = 10$ by considering only one pathogen, the parameters have to be changed. This suggests a possible bias in real flu parameters estimation

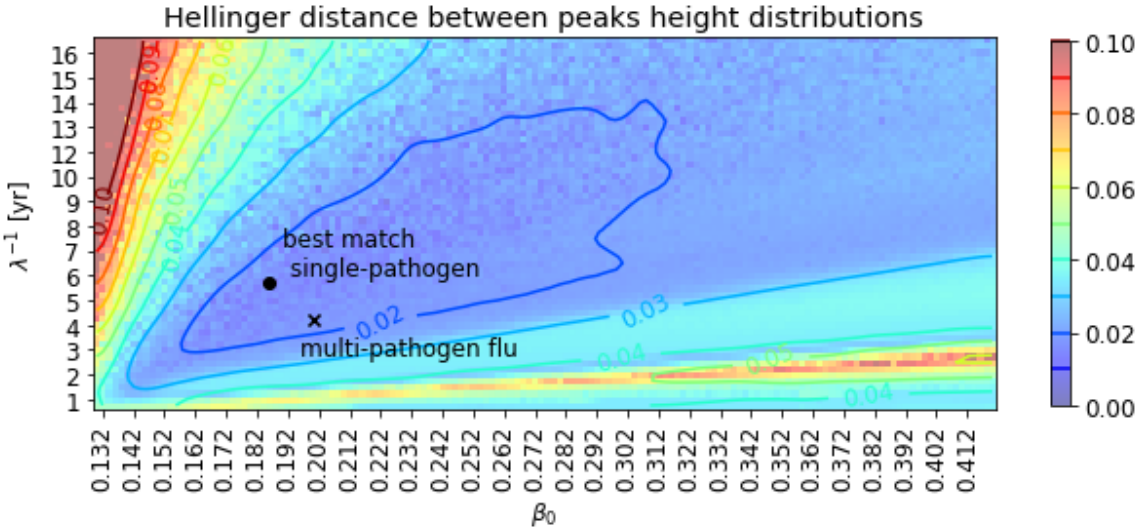

Figure 4.5: Hellinger distance between the peaks height distribution of flu with $n = 10$ pathogens and the peaks height distribution of a single pathogen with (β_0, λ^{-1}). The black dot is the couple of parameters that leads to a minimum distance while in the black x there are flu parameters.

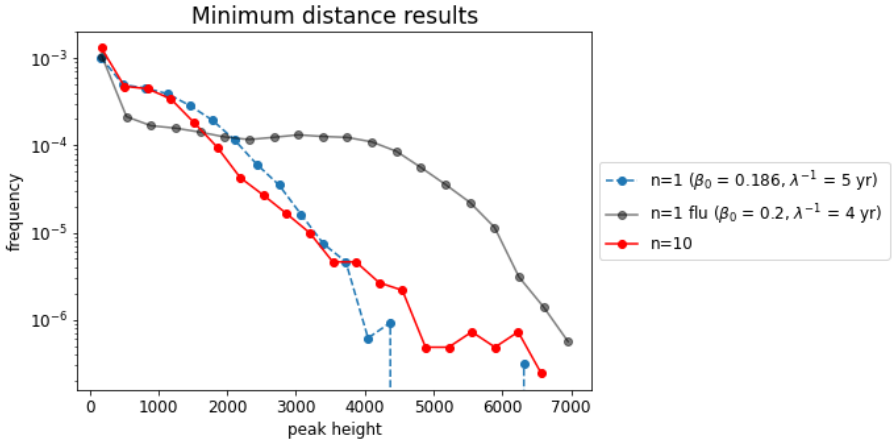

Figure 4.6: Flu peaks height distribution with $n = 10$ pathogens (red), the single-pathogen distribution with the minimum Hellinger distance (blue) and the original flu distribution (gray).

especially for the duration of immunity, as introduced at the beginning of this section. However the two couples of parameters are both inside the minimum-distance region. Indeed this region is quite wide, it comprehends values of $\beta_0 \in [0.160; 0.312]$ and duration of immunity over many years, $\lambda^{-1} \in [3; 14] yr$. Thus, the a posteriori effect on the dynamics of a single pathogen or of a pathogen in a populated system is the same as long as the values of β_0 and λ^{-1} are rescaled.

4.3 Correlation between incidences

In the study conducted by *Nickbakhsh et al.* [4], they start by examining virological diagnostic data from over 40,000 episodes of respiratory illness over a 9-year period. Each of the patients is tested for 11 virus groups, thanks to multiplex-PCR methods. This provides a data source about the co-circulating and co-infecting viruses. From these data they use different approaches to infer virus-virus interaction starting from the viruses prevalence time series: from the simpler Spearman's rank correlation coefficient to multivariate Bayesian framework and hierarchical autoregressive model. Thus, they basically assume that it is possible to retrieve the presence of an interaction between viruses by looking at their dynamics correlation.

However there are some other studies like [48], which is applied to microbial network interactions, where they test whether such correlation-based methods are reliable for inferring interaction networks. By simulating bacterial communities they show that the information about interactions carried by the correlations in abundance is limited.

The reason for the following analysis is to explore such context through the pathogen-view multi-pathogen model. With a simplified setup, the question becomes: is there a relation between the interaction strength matrix σ that is given in input to the model and the correlation matrix ρ, that is computed between the output incidences? For example, for a three pathogens system the relation searched is between:

$$\sigma = \begin{pmatrix} 0 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & 0 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & 0 \end{pmatrix} \overset{?}{\longleftrightarrow} \begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{pmatrix} = \rho$$

. To address this question, for the first time in this work it is introduced a σ matrix that is symmetric but has different entries, since we would like to relate each σ_{ij} to its relative ρ_{ij}. As correlation estimator we use the Pearson correlation coefficient, to see if even in a simplified setup like this one, it is possible to extract some information from the incidences.

We use our model to simulate systems with 2,3,5 pathogens. For the cases of 2 and 3 pathogens, σ off-diagonal elements are all the possible combinations of pairs and triplets composed with $\{0.1, 0.3, 0.5, 0.7, 0.9\}$. For the 5 pathogens case, 100 combinations are chosen at random from the same possible values. For each combination the simulation is run for 100 stochastic realizations. At each realization, pathogens parameters are drawn from the distributions described in Figure 3.3. For the resulting incidences, pairwise correlations are computed and averaged over the realizations. The results from all these simulations are reported in Figure 4.7 as pairwise Pearson correlation coefficients between i and j incidences as a function of the initial σ_{ij}. Each point corresponds to the mean value (over the stochastic realizations) of ρ_{ij} when the original interaction among i and j is σ_{ij}. The bars are the respective standard deviations. When multiple points are present, it means that the same interaction σ_{ij} leads to different values of ρ_{ij}, depending on the other values present in the σ elements combination.

By looking at the most left plot, case of 2 pathogens, the trend is clear: the stronger the interaction the lower the correlation. Suggesting an inverse relation between σ and ρ. Moving to the case of 3 or 5 pathogens, the trend seems to be the same but at each value of the interaction corresponds a range of correlation values. This means that by adding pathogens to the system, the correlation becomes less informative about the pathogens input interactions. To have a confirmation on the trend we compute the Kendall's Tau distribution, a measure to

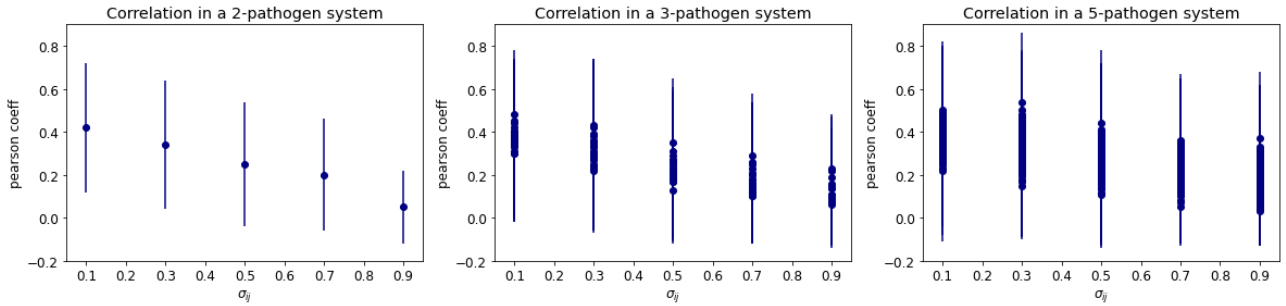

Figure 4.7: Pearson correlation coefficient of pairwise incidences as a function of the respective initial σ_{ij}. Case of 2,3,5 pathogens in the system.

evaluate the correspondence between the two rankings: in this case, between the combinations of σ values and the respective ρ.

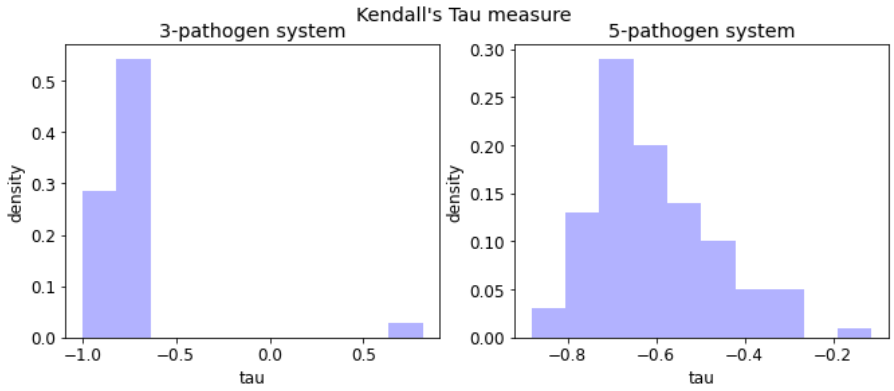

Figure 4.8: Kendall's Tau distribution between correlation and interaction strength rankings. Case of 3,5 pathogens in the system.

The values found in Figure 4.8 for the case of 3 pathogens are negative and mostly close to -1, indicating strong disagreement between the two rankings, *i.e.* the higher the σ_{ij} the smaller the correlation between incidences ρ_{ij}. However, for the case of 5 pathogens on the right, the values found are noisier and widely spread in the negative semi-axis. It can be concluded that with such simplified setup, when there are few pathogens in the system, the interactions strength and the correlation maintain roughly an inverse relation. Though by adding pathogens to the system, the Pearson correlation coefficient cannot be used anymore to retrieve information about the initial interaction between the viruses.

To better illustrate indirect effects that arise when more than two pathogens circulate we consider a specific example. We simulate a system with three pathogens: the pairs of pathogens 1-2 and 2-3 are independent ($\sigma_{12} = 0 = \sigma_{23}$) and the pair 1-3 have a different level of interaction at each simulation:

$$\sigma = \begin{pmatrix} 0 & 0 & \sigma_{13} \\ 0 & 0 & 0 \\ \sigma_{13} & 0 & 0 \end{pmatrix}$$

with $\sigma_{13} \in \{0, 0.4, 0.8, 1\}$. Simulations are repeated for 100 stochastic realizations, with pathogens parameters drawn at random each time. We compute the Pearson correlation coefficients between the resulting incidences and their mean (over the realizations) is reported

in Figure 4.9 as a function of σ_{13}. Each curve corresponds to a pair of pathogens. The trend shown in the plot is that the stronger the interaction between 1-3, the lower the correlation between all the incidences pairs. For 1-3 this is expected as already discussed above, while for 1-2 and 2-3 it is not, since these pathogens are kept independent during all the simulations: the correlation between independent pathogens decreases as the strength of the interaction between the third couple 1-3 increases. Thus, the presence of interactions affects the correlation between all the pathogens, even if they are independent and we expect them not to affect each other. Concluding, when considering systems with interacting pathogens, Pearson correlation between incidences is not a reliable indicator of the level of pathogens interaction.

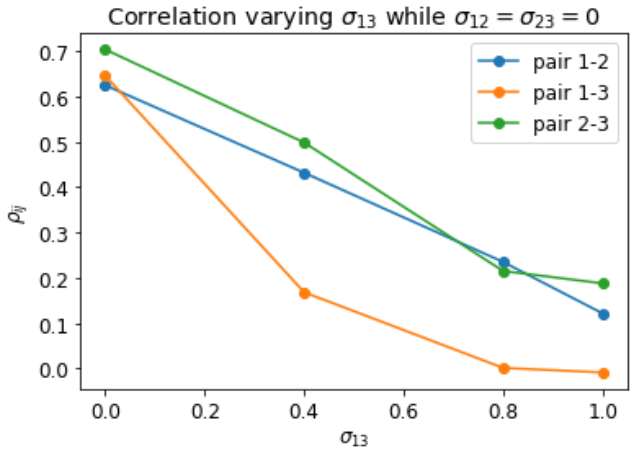

Figure 4.9: Pearson correlation coefficient between incidences in a three pathogens system where pathogens 1-2 and 2-3 are kept independent ($\sigma_{12} = 0 = \sigma_{23}$) and the pair 1-3 has interaction $\sigma_{13} \in \{0, 0.4, 0.8, 1\}$. For all the curves, the higher σ_{13} the lower ρ_{ij}: even correlations between the incidences of independent pathogens are affected by the presence of the third pair's interaction.

4.4 Conclusions

Thanks to the characteristics explained in chapter 3, our model can overcome the actual limitations in modeling interdependent epidemics caused by respiratory viruses. We aim at highlighting the effects of including within our framework the presence of other pathogens on a single pathogen's dynamics. For illustrative purposes, we take influenza as a reference pathogen, we simulate different scenarios and we monitor the response of our model on influenza incidence.

We test our model by changing the number of pathogens, the strength of their interactions and the duration. The effects on influenza incidence are that the peaks get lower by increasing the number of co-circulating pathogens when $n \geq 2$ and by increasing the extent of the interactions both in strength and duration. From this analysis emerges that the presence of a pathogen can slightly perturb the dynamics of another. However, when considering many other pathogens, the overall impact on the focal pathogen can become relevant. If we apply these considerations to ILI context introduced in chapter 2, we obtain that distinguishing the viruses and treating them like a multi-pathogen system helps to clarify the impact of each one of them on the others' dynamics.

This has important implications also in parameter estimation. Considering a pathogen as alone and neglecting the fact that it is co-circulating with many others could lead to a bias in the

estimated parameters. To illustrate this we perform a numerical experiment, we try to reproduce the dynamics of influenza co-circulating with 10 other pathogens with a single-pathogen model, *i.e.* disregarding the presence of other pathogens. The explored parameter space is the (β, λ^{-1}). As a result we find that the dynamics of influenza with many other pathogens can be retrieved with a single pathogen having different transmissibility and immunity duration. This indicates that the estimate may be affected by the presence of the others and causes a rescaling of the parameters.

Finally, we aim at finding a relation between the correlation of the incidences, output of the model, and the viruses' interactions, given in input. We use the Pearson coefficient as a correlation estimator. We find that the general trend is an inverse relation: viruses with stronger interactions have less correlated incidence dynamics and weaker interactions produce more correlated dynamics. This shows that the incidence dynamics carries to some extent information about the virus-virus interactions governing such a dynamics. However, when the number of viruses increases this trend becomes less clear. This highlights the challenge of identifying biological interactions from population level spreading dynamics alone.

Chapter 5

Conclusions

In this work we illustrate the complexities behind systems composed of multiple respiratory viruses, analyzing the consequences of viruses heterogeneities and their interactions. We introduce the epidemiological technique of compartmental models that can be adapted to describe such a complex system. We present two compartmental models that have been developed for modeling multiple-pathogen circulation: the host-view [5] and the pathogen-view [6]. Since we want to describe systems populated by different viruses, these models are not optimal to suit this purpose. The first is not scalable but it can describe different levels of temporary immunity, thus it can suit multi-pathogen scenarios. The second is built for multi-strain systems and as such it considers an acquired immunity that is long-term. However it has the advantage that it can easily scale with the number of pathogens under consideration. To fill this gap, we propose a new framework: due to the pathogen-view it is scalable and thanks to the convalescent compartments the cross-immunity is temporary and appropriate for our multi-pathogen system.

From our analysis emerges that the overall effect of many pathogens on another pathogen's dynamics can be relevant and thus they have to be taken into account to provide an accurate description of the reality. In particular, neglecting the presence of other pathogens in the system can lead to biases in parameter estimation. Especially we underline the case of immunity duration. Then, we find that a simple estimator of the correlation between incidences, as the Pearson coefficient, can not give reliable insights on the strength of the interactions among the viruses. By adding pathogens to the system this parameter becomes more noisy and can not be taken as an indicator of the interactions. This question should be further investigated with more sophisticated techniques.

A future implementation of our framework is simulating systems with real pathogens parameters to have a realistic scenario and properly investigate practical questions. For example, we want to apply this model to the influenza-RSV debate anticipated in chapter 1. We would like to study the response of the system after the introduction of the RSV vaccine, and the role played by the presence of other pathogens. Since there is evidence of interference between ILI viruses [4], what could be the result of suppressing one of them? In particular, would there be more benefits in the number of people protected by the vaccine against RSV or in the number of people that, thanks to the interference between influenza and RSV, are protected against influenza? Also, our framework could be used to understand the co-circulation of pathogens perturbed by the COVID-19 emergency [1].

Appendix A

Comparison between views

The pathogen-view multi-pathogen model (section 3.1) has been introduced in this work to overcome the host-view model (section 2.3) that due to his completeness is impossible to upscale to systems with high number of viruses co-circulating. As their names suggest they adopt two different views: the pathogen and the host one, respectively. The first one can treat the viruses independently, by considering the hosts' health situation with respect to each virus at a time. The latter keeps track of the hosts' health with respect to every virus at the same time, in each compartment, leading to an exponential growth of the number of compartments needed. Both models are built in a way to take into account virus-virus interactions through the convalescent compartments and the reduced probability of secondary infections after the first one, as already explained in the respective sections.

In this chapter it is presented the complete 2-pathogen host-view model (section A.1), that adds to the basic version presented in section 2.3 seasonality, waning of immunity and immigration term. Then, the version extended to 3-pathogen systems (section A.1). Finally the two different views models are tested in different situations under the same assumptions, to compare the two resulting trends. The models are compared only for both the cases of 2 and 3 pathogens in the system.

A.1 Host view model - two pathogens

In Figure A.1 is the scheme of a complete 2-pathogen host-view model. It takes into account single infections (compartments I_1, I_2), co-infections (I_{12}), the cross protection compartments of convalescent (C_1, C_2) and the possibility of a second infection after the recovery from the first one ($I_1^{(R_2)}, I_2^{(R_1)}$). The immunity is lost with a rate λ_i and m represents the immigration term.

The dynamics of this model is described by the set of differential equations in Equation A.1.

$$
\begin{aligned}
\dot{S} &= -(\beta_1 \frac{I_1^*}{N} + \beta_2 \frac{I_2^*}{N})S + \lambda_1(R_1 + C_2) + \lambda_2(R_2 + C_1) - 2m \\
\dot{I_1} &= \beta_1 \frac{I_1^*}{N} S - (\mu_1 + r_I^{(12)} \beta_2 \frac{I_2^*}{N}) I_1 + m \\
\dot{I_2} &= \beta_2 \frac{I_2^*}{N} S - (\mu_2 + r_I^{(21)} \beta_1 \frac{I_1^*}{N}) I_2 + m \\
\dot{I_{12}} &= r_I^{(12)} \beta_2 \frac{I_2^*}{N} I_1 + r_I^{(21)} \beta_1 \frac{I_1^*}{N} I_2 - (\mu_1 + \mu_2) I_{12} \\
\dot{C_1^{(2)}} &= \mu_2 I_2 - (\gamma^{(21)} + \lambda_2 + r_C^{(21)} \beta_1 \frac{I_1^*}{N}) C_1^{(2)} \\
\dot{C_2^{(1)}} &= \mu_1 I_1 - (\gamma^{(12)} + \lambda_1 + r_C^{(12)} \beta_2 \frac{I_2^*}{N}) C_2^{(1)} \\
\dot{I_1^{(R_2)}} &= r_C^{(21)} \beta_1 \frac{I_1^*}{N} C_1^{(2)} + \beta_1 \frac{I_1^*}{N} R_2 + \mu_2 I_{12} - \mu_1 I_1^{(R_2)} \\
\dot{I_2^{(R_1)}} &= r_C^{(12)} \beta_2 \frac{I_2^*}{N} C_2^{(1)} + \beta_2 \frac{I_2^*}{N} R_1 + \mu_1 I_{12} - \mu_2 I_2^{(R_1)} \\
\dot{R_1} &= \gamma^{(12)} C_2^{(1)} + \lambda_2 R_{12} - (\beta_2 \frac{I_2^*}{N} + \lambda_1) R_1 \\
\dot{R_2} &= \gamma^{(21)} C_1^{(2)} + \lambda_1 R_{12} - (\beta_1 \frac{I_1^*}{N} + \lambda_2) R_2 \\
\dot{R_{12}} &= \mu_1 I_1^{(R_2)} + \mu_2 I_2^{(R_1)} - (\lambda_1 + \lambda_2) R_{12}
\end{aligned}
\tag{A.1}
$$

. Where blue terms represent the incidence of pathogen 1, and:

$$
\begin{aligned}
I_1^* &= I_1 + I_{12} + I_1^{(R_2)} \\
I_2^* &= I_2 + I_{12} + I_2^{(R_1)}
\end{aligned}
$$

represent the total number of individuals that are infectious with pathogen 1 or 2 respectively.

A.2 Host view model - three pathogens

The 3-pathogens host-view model takes into account the same mechanisms explained for the 2-pathogens version: single infections (I_i), secondary infections ($I_i^{(R_j)}, I_i^{(R_j R_k)}$), cross-protection in the convalescent compartments (C_i, C_{ij}) and co-infections at each step ($I_{ij}, I_{ij}^{(R_k)}$). In addition, in this case it is possible to have triple infections at the same time (I_{ijk}). There is loss of immunity with rate λ_i and m is the immigration contribution. Again, the interdependence between pathogens is modeled through the reduction parameters r_I, r_C.

The system of differential equations that describes the host-view model with three pathogens is in Equation A.2.

$$
\begin{aligned}
\dot{S} &= -(\beta_1 \frac{I_1^*}{N} + \beta_2 \frac{I_2^*}{N} + \beta_3 \frac{I_3^*}{N})S + \lambda_1(R_1 + C_{23}) + \lambda_2(R_2 + C_{13}) + \lambda_3(R_3 + C_{12}) - 3m \\
\dot{I_1} &= \beta_1 \frac{I_1^*}{N} S - (\mu_1 + r_I^{(1,2)} \beta_2 \frac{I_2^*}{N} + r_I^{(1,3)} \beta_3 \frac{I_3^*}{N}) I_1 + m
\end{aligned}
$$

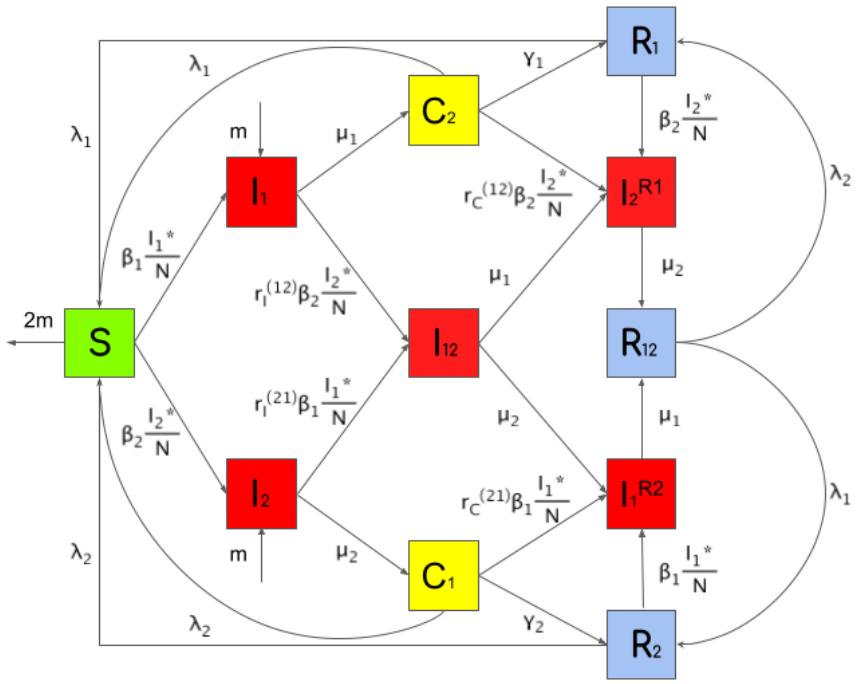

Figure A.1: Complete scheme of the host-view compartmental model for a 2-pathogen system.

$$\dot{I}_2 = \beta_2 \frac{I_2^*}{N} S - (\mu_2 + r_I^{(2,1)} \beta_1 \frac{I_1^*}{N} + r_I^{(2,3)} \beta_3 \frac{I_3^*}{N}) I_2 + m$$

$$\dot{I}_3 = \beta_3 \frac{I_3^*}{N} S - (\mu_3 + r_I^{(3,1)} \beta_1 \frac{I_1^*}{N} + r_I^{(3,2)} \beta_2 \frac{I_2^*}{N}) I_3 + m$$

$$\dot{I_{12}} = r_I^{(2,1)} \beta_1 \frac{I_1^*}{N} I_2 + r_I^{(1,2)} \beta_2 \frac{I_2^*}{N} I_1 - (\mu_1 + \mu_2 + r_I \beta_3 \frac{I_3^*}{N}) I_{12}$$

$$\dot{I_{13}} = r_I^{(3,1)} \beta_1 \frac{I_1^*}{N} I_3 + r_I^{(1,3)} \beta_3 \frac{I_3^*}{N} I_1 - (\mu_1 + \mu_3 + r_I \beta_2 \frac{I_2^*}{N}) I_{13}$$

$$\dot{I_{23}} = r_I^{(3,2)} \beta_2 \frac{I_2^*}{N} I_3 + r_I^{(2,3)} \beta_3 \frac{I_3^*}{N} I_2 - (\mu_2 + \mu_3 + r_I \beta_1 \frac{I_1^*}{N}) I_{23}$$

$$\dot{C_{12}} = \mu_3 I_3 - (\gamma^{(3)} + \lambda_3 + r_C^{(3,1)} \beta_1 \frac{I_1^*}{N} + r_C^{(3,2)} \beta_2 \frac{I_2^*}{N}) C_{12}$$

$$\dot{C_{13}} = \mu_2 I_2 - (\gamma^{(2)} + \lambda_2 + r_C^{(2,3)} \beta_3 \frac{I_3^*}{N} + r_C^{(2,1)} \beta_1 \frac{I_1^*}{N}) C_{13}$$

$$\dot{C_{23}} = \mu_1 I_1 - (\gamma^{(1)} + \lambda_1 + r_C^{(1,2)} \beta_2 \frac{I_2^*}{N} + r_C^{(1,3)} \beta_3 \frac{I_3^*}{N}) C_{23}$$

$$\dot{I_2^{(R_1)}} = r_C^{(1,2)} \beta_2 \frac{I_2^*}{N} C_{23} + \beta_2 \frac{I_2^*}{N} R_1 + \mu_1 I_{12} - (\mu_2 + r_I^{(2,3)} \beta_3 \frac{I_3^*}{N}) I_2^{(R_1)}$$

$$\dot{I_3^{(R_1)}} = r_C^{(1,3)} \beta_3 \frac{I_3^*}{N} C_{23} + \beta_3 \frac{I_3^*}{N} R_1 + \mu_1 I_{13} - (\mu_3 + r_I^{(3,2)} \beta_2 \frac{I_2^*}{N}) I_3^{(R_1)}$$

$$\dot{I_1^{(R_2)}} = r_C^{(2,1)} \beta_1 \frac{I_1^*}{N} C_{13} + \beta_1 \frac{I_1^*}{N} R_2 + \mu_2 I_{12} - (\mu_1 + r_I^{(1,3)} \beta_3 \frac{I_3^*}{N}) I_1^{(R_2)}$$

$$\dot{I_3^{(R_2)}} = r_C^{(2,3)} \beta_3 \frac{I_3^*}{N} C_{13} + \beta_3 \frac{I_3^*}{N} R_2 + \mu_2 I_{23} - (\mu_3 + r_I^{(3,1)} \beta_1 \frac{I_1^*}{N}) I_3^{(R_2)}$$

$$\dot{I_1^{(R_3)}} = r_C^{(3,1)} \beta_1 \frac{I_1^*}{N} C_{12} + \beta_1 \frac{I_1^*}{N} R_3 + \mu_3 I_{13} - (\mu_1 + r_I^{(1,2)} \beta_2 \frac{I_2^*}{N}) I_1^{(R_3)}$$

$$
\begin{aligned}
\dot{I_2^{(R_3)}} &= r_C^{(3,2)}\beta_2\frac{I_2^*}{N}C_{12} + \beta_2\frac{I_2^*}{N}R_3 + \mu_3 I_{23} - (\mu_2 + r_I^{(2,1)}\beta_1\frac{I_1^*}{N})I_2^{(R_3)} \\
\dot{R_1} &= \gamma^{(1)}C_{23} + \lambda_2 R_{12} + \lambda_3 R_{13} + \lambda_2 C_3 + \lambda_3 C_2 - (\beta_2\frac{I_2^*}{N} + \beta_3\frac{I_3^*}{N} + \lambda_1)R_1 \\
\dot{R_2} &= \gamma^{(2)}C_{13} + \lambda_1 R_{12} + \lambda_3 R_{23} + \lambda_1 C_3 + \lambda_3 C_1 - (\beta_1\frac{I_1^*}{N} + \beta_3\frac{I_3^*}{N} + \lambda_2)R_2 \\
\dot{R_3} &= \gamma^{(3)}C_{12} + \lambda_1 R_{13} + \lambda_2 R_{23} + \lambda_1 C_2 + \lambda_2 C_1 - (\beta_1\frac{I_1^*}{N} + \beta_2\frac{I_2^*}{N} + \lambda_3)R_3 \\
\dot{C_1} &= \mu_2 I_2^{(R_3)} + \mu_3 I_3^{(R_2)} - (\gamma^{(23)} + \lambda_2 + \lambda_3 + r_{CC}^{(23,1)}\beta_1\frac{I_1^*}{N})C_1 \\
\dot{C_2} &= \mu_1 I_1^{(R_3)} + \mu_3 I_3^{(R_1)} - (\gamma^{(13)} + \lambda_1 + \lambda_3 + r_{CC}^{(13,2)}\beta_2\frac{I_2^*}{N})C_2 \\
\dot{C_3} &= \mu_1 I_1^{(R_2)} + \mu_2 I_2^{(R_1)} - (\gamma^{(12)} + \lambda_1 + \lambda_2 + r_{CC}^{(12,3)}\beta_3\frac{I_3^*}{N})C_3 \\
\dot{R_{12}} &= \gamma^{(12)}C_3 + \lambda_3 R_{123} - (\beta_3\frac{I_3^*}{N} + \lambda_1 + \lambda_2)R_{12} \\
\dot{R_{13}} &= \gamma^{(13)}C_2 + \lambda_2 R_{123} - (\beta_2\frac{I_2^*}{N} + \lambda_1 + \lambda_3)R_{13} \\
\dot{R_{23}} &= \gamma^{(23)}C_1 + \lambda_1 R_{123} - (\beta_1\frac{I_1^*}{N} + \lambda_2 + \lambda_3)R_{23} \\
\dot{I_1^{(R_2R_3)}} &= r_{CC}^{(23,1)}\beta_1\frac{I_1^*}{N}C_1 + \beta_1\frac{I_1^*}{N}R_{23} + \mu_2 I_{12}^{(R_3)} + \mu_3 I_{13}^{(R_2)} - \mu_1 I_1^{(R_2R_3)} \\
\dot{I_2^{(R_1R_3)}} &= r_{CC}^{(13,2)}\beta_2\frac{I_2^*}{N}C_2 + \beta_2\frac{I_2^*}{N}R_{13} + \mu_1 I_{12}^{(R_3)} + \mu_3 I_{23}^{(R_1)} - \mu_2 I_2^{(R_1R_3)} \\
\dot{I_3^{(R_1R_2)}} &= r_{CC}^{(12,3)}\beta_3\frac{I_3^*}{N}C_3 + \beta_3\frac{I_3^*}{N}R_{12} + \mu_1 I_{13}^{(R_2)} + \mu_2 I_{23}^{(R_1)} - \mu_3 I_3^{(R_1R_2)} \\
\dot{I_{12}^{(R_3)}} &= r_I^{(2,1)}\beta_1\frac{I_1^*}{N}I_2^{(R_3)} + r_I^{(1,2)}\beta_2\frac{I_2^*}{N}I_1^{(R_3)} - (\mu_1 + \mu_2)I_{12}^{(R_3)} + \mu_3 I_{123} \\
\dot{I_{13}^{(R_2)}} &= r_I^{(3,1)}\beta_1\frac{I_1^*}{N}I_3^{(R_2)} + r_I^{(1,3)}\beta_3\frac{I_3^*}{N}I_1^{(R_2)} - (\mu_1 + \mu_3)I_{13}^{(R_2)} + \mu_2 I_{123} \\
\dot{I_{23}^{(R_1)}} &= r_I^{(3,2)}\beta_2\frac{I_2^*}{N}I_3^{(R_1)} + r_I^{(2,3)}\beta_3\frac{I_3^*}{N}I_2^{(R_1)} - (\mu_2 + \mu_3)I_{23}^{(R_1)} + \mu_1 I_{123} \\
\dot{I_{123}} &= r_I\beta_1\frac{I_1^*}{N}I_{23} + r_I\beta_2\frac{I_2^*}{N}I_{13} + r_I\beta_3\frac{I_3^*}{N}I_{12} - (\mu_1 + \mu_2 + \mu_3)I_{123} \\
\dot{R_{123}} &= \mu_1 I_1^{(R_2R_3)} + \mu_2 I_2^{(R_1R_3)} + \mu_3 I_3^{(R_1R_2)} - (\lambda_1 + \lambda_2 + \lambda_3)R_{123}
\end{aligned}
\tag{A.2}
$$

. Where blue terms represent the incidence of pathogen 1, and:

$$
\begin{aligned}
I_1^* &= I_1 + I_{12} + I_{13} + I_1^{(R_2)} + I_1^{(R_3)} + I_1^{(R_2R_3)} + I_{12}^{(R_3)} + I_{13}^{(R_2)} + I_{123} \\
I_2^* &= I_2 + I_{12} + I_{23} + I_2^{(R_1)} + I_2^{(R_3)} + I_2^{(R_1R_3)} + I_{12}^{(R_3)} + I_{23}^{(R_1)} + I_{123} \\
I_3^* &= I_3 + I_{13} + I_{23} + I_3^{(R_1)} + I_3^{(R_2)} + I_3^{(R_1R_2)} + I_{13}^{(R_2)} + I_{23}^{(R_1)} + I_{123}
\end{aligned}
$$

are the total number of individuals that are infectious with pathogen 1, 2 or 3 respectively.

Despite the completeness of this model formulation, which takes into account all the possible mechanisms that occur between three pathogens, its main limitation is the impossibility to be upscaled to higher n. When dealing with two pathogens the number of compartments is

11 while with three pathogens is 33. Such scaling with n is exponential and becomes almost impossible to handle more virus-populated systems.

A.3 Two pathogens comparison

For the two-pathogen system, the reference pathogen is taken with $\beta_{0,\text{ref}} = 0.14$, while for the other pathogen three cases are tested: $\beta_{0,\text{other}}$ equal, smaller, greater than $\beta_{0,\text{ref}}$. For the second and third case the same test is repeated for two different $\beta_{0,\text{other}}$ values: 0.1,0.12 and 0.2,0.15 respectively. The other parameters are set to $\mu^{-1} = 4.5$ d and $\lambda^{-1} = 2 \times 365$ d for all. Each scenario simulation is run for 50 years, for 200 stochastic realizations on both models, varying the duration of the cross protection (γ^{-1} and q^{-1}) and the strength of the pathogens interaction ($r_I = r_C$ and σ):

$$\gamma^{-1} \in \{7, 14, 21, 30, 60\}\text{d}, \quad q^{-1} = \gamma^{-1} + \mu^{-1}$$
$$r_I, r_C, \sigma \in [0; 1], \quad \text{step} = 0.2$$

. Then, the mean peaks height and at its standard deviation are collected and the results for the different cases are reported as heatmaps in Figure A.2, Figure A.3, Figure A.4. When the two pathogens have the same transmissibility parameter (Figure A.2) the mean peak height has the same trend for both models: the stronger the interaction ($\sigma \to 1$, $r_I, r_C \to 0$) or the longer the cross-protection period (smaller q, γ) the lower the peaks. In the case where the reference pathogen is the most transmissible (Figure A.3), the results of the interactions on it are irrelevant (case (a)) or very light (case (b)). The slightly emerging trend is the same found in the case of similar pathogens Figure A.2. In the last case the reference pathogen is the weakest (Figure A.4), when the second pathogen is quite different from it (case (a)), it has a big influence on the reference's peaks heights and the trend observed in the previous cases is inverted. Apart from the 'non-interacting' case, the higher the interaction and the greater q^{-1}, the higher the peaks. While if the pathogen is similar to the reference (case (b)) the trend is the usual one. To conclude, in Figure A.5 the relative differences of the mean peak height show how close are the results of the two models, in each case. All the relative differences have an order of magnitude smaller or equal to 10^{-2}. It can be said that the two models with different views display similar behaviors when subjected to the same conditions, in a 2-pathogen system. Nevertheless, as anticipated at the beginning of section 3.2, the simplification limitations of the pathogen-view model could emerge when dealing with multi-pathogen systems, where the co-infections could play a determinant role. For this reason, to have a more meaningful comparison between these models the analysis just exposed needs to be repeated at least on a 3-pathogen system. It would be optimal to be able to compare the two views also with more populated systems, but already with 3 pathogens the host-view model becomes difficult to handle: this is also the reason why it is important to study a scalable model such as the pathogen-view one.

A.4 Three pathogens comparison

The same simulations already exposed are repeated in the case of three co-circulating pathogens. Three different cases are tested: one where all the pathogens have the same transmissibility parameter ($\beta_{0,\text{ref}} = 0.14 = \beta_{0,\text{others}}$), another where the reference pathogen is the strongest ($\beta_{0,\text{ref}} = 0.14 > \beta_{0,\text{others}} = [0.1, 0.12]$) and then the case in which it is the weakest ($\beta_{0,\text{ref}} =$

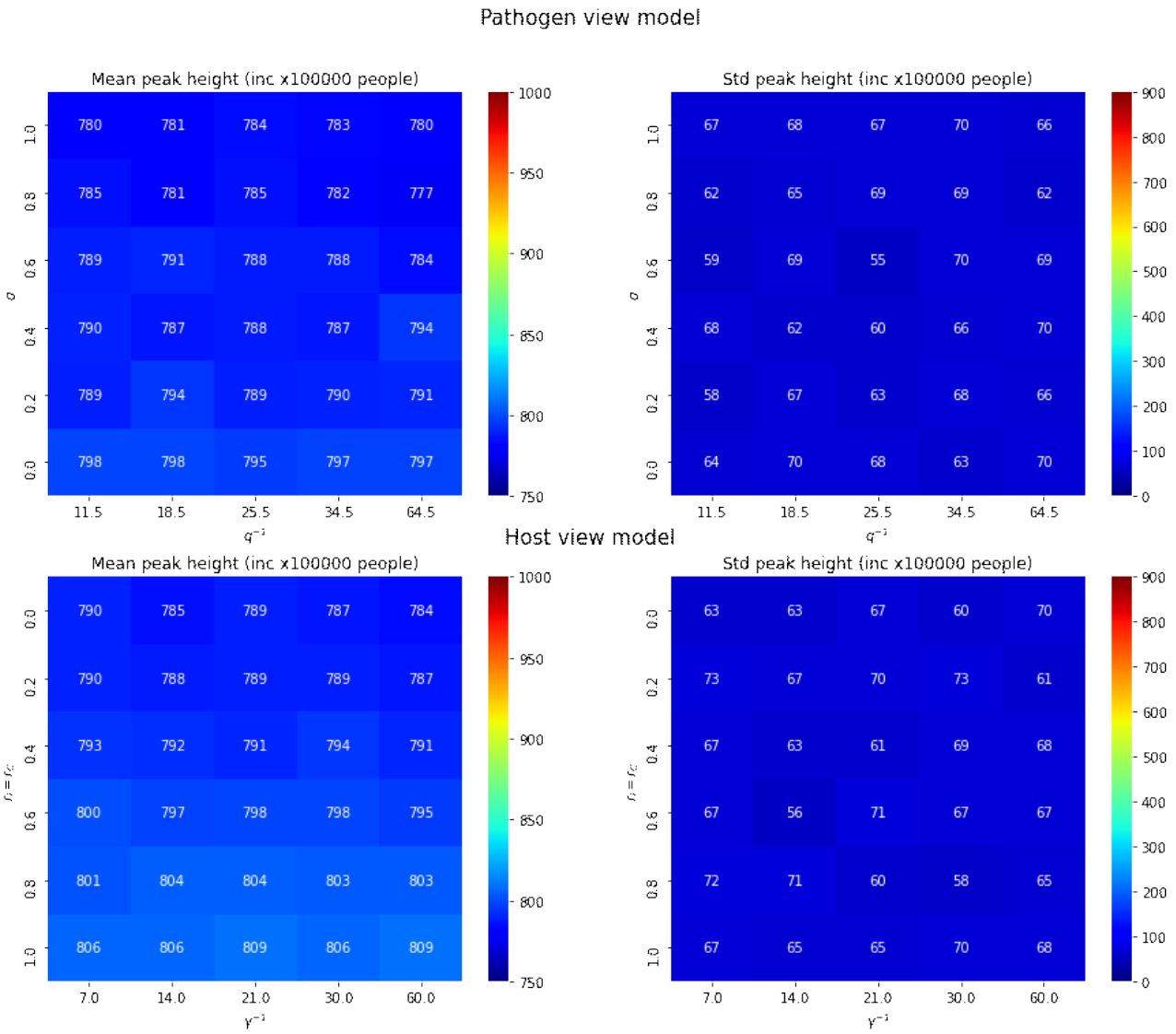

Figure A.2: Weekly incidence peaks on a 2 pathogens system. Case similar pathogens: $\beta_{0,\text{ref}} = 0.14 = \beta_{0,\text{other}}$, $\mu^{-1} = 4.5$ d, $\lambda^{-1} = 2 \times 365$ d.

$0.14 < \beta_{0,\text{others}} = [0.2, 0.15]$). The other parameters are set to $\mu^{-1} = 4.5$ d and $\lambda^{-1} = 2 \times 365$ d.

The resulting trends are the same already highlighted: when pathogens are similar (Figure A.6), the mean peaks height increases with the decreasing of the strength of interaction ($\sigma \to 0, r_I, r_C \to 1$); when the co-circulating pathogens are weaker (Figure A.7) they do not affect the reference's peaks height; while when they are more transmissible (Figure A.8), the trend is the opposite, with the mean height growing with the strength of interaction. In Figure A.9 there are the relative differences between the two models results. Even in this case these differences have an order of magnitude smaller or equal to 10^{-2}, suggesting a coherence between the two views also in the case of three pathogens.

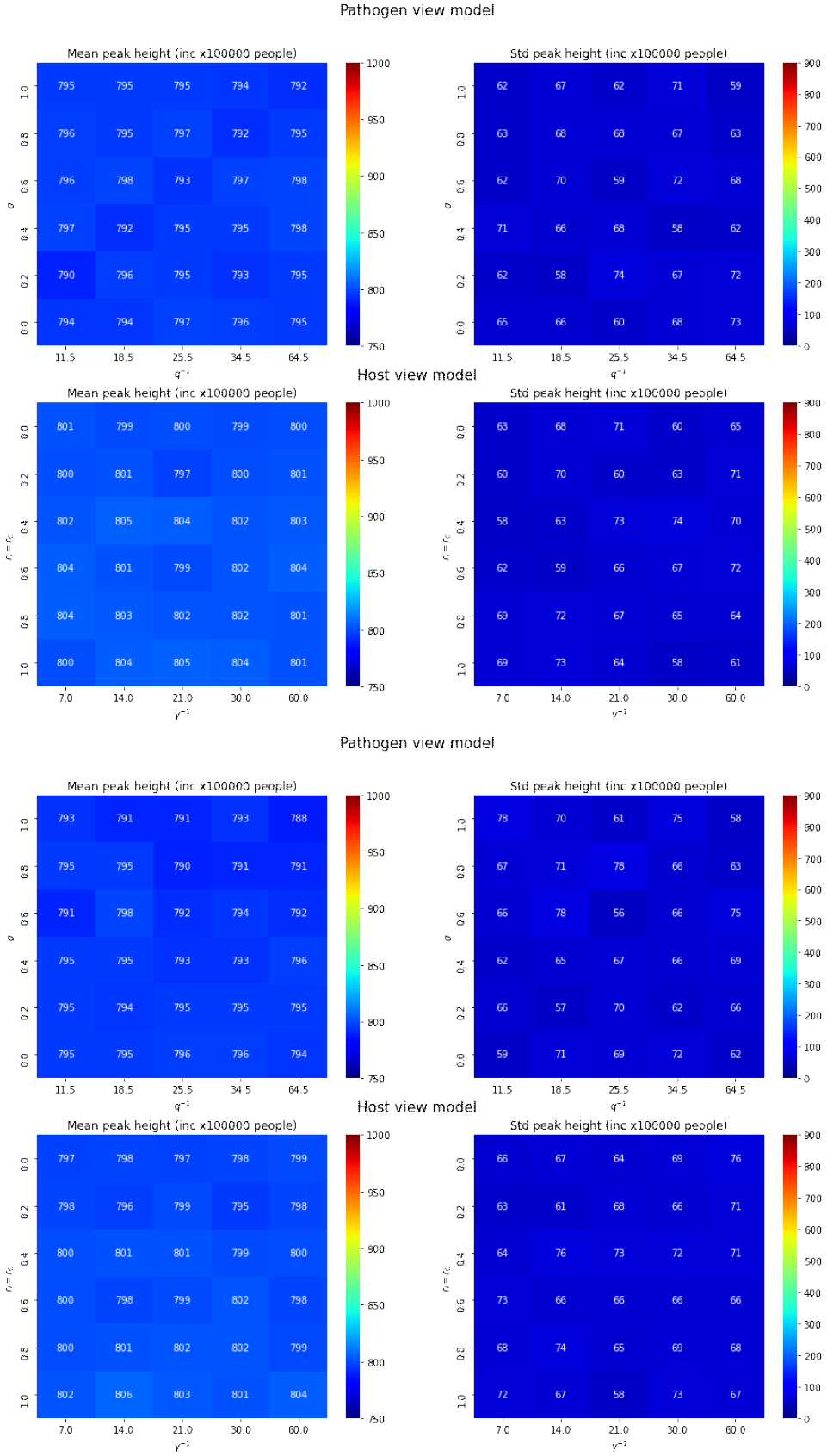

Figure A.3: Weekly incidence peaks on a 2 pathogens system. Case reference pathogen is the strongest one.
(a) $\beta_{0,\text{ref}} = 0.14 > \beta_{0,\text{other}} = 0.1$ (b) $\beta_{0,\text{ref}} = 0.14 > \beta_{0,\text{other}} = 0.12$. $\mu^{-1} = 4.5$ d, $\lambda^{-1} = 2 \times 365$ d.

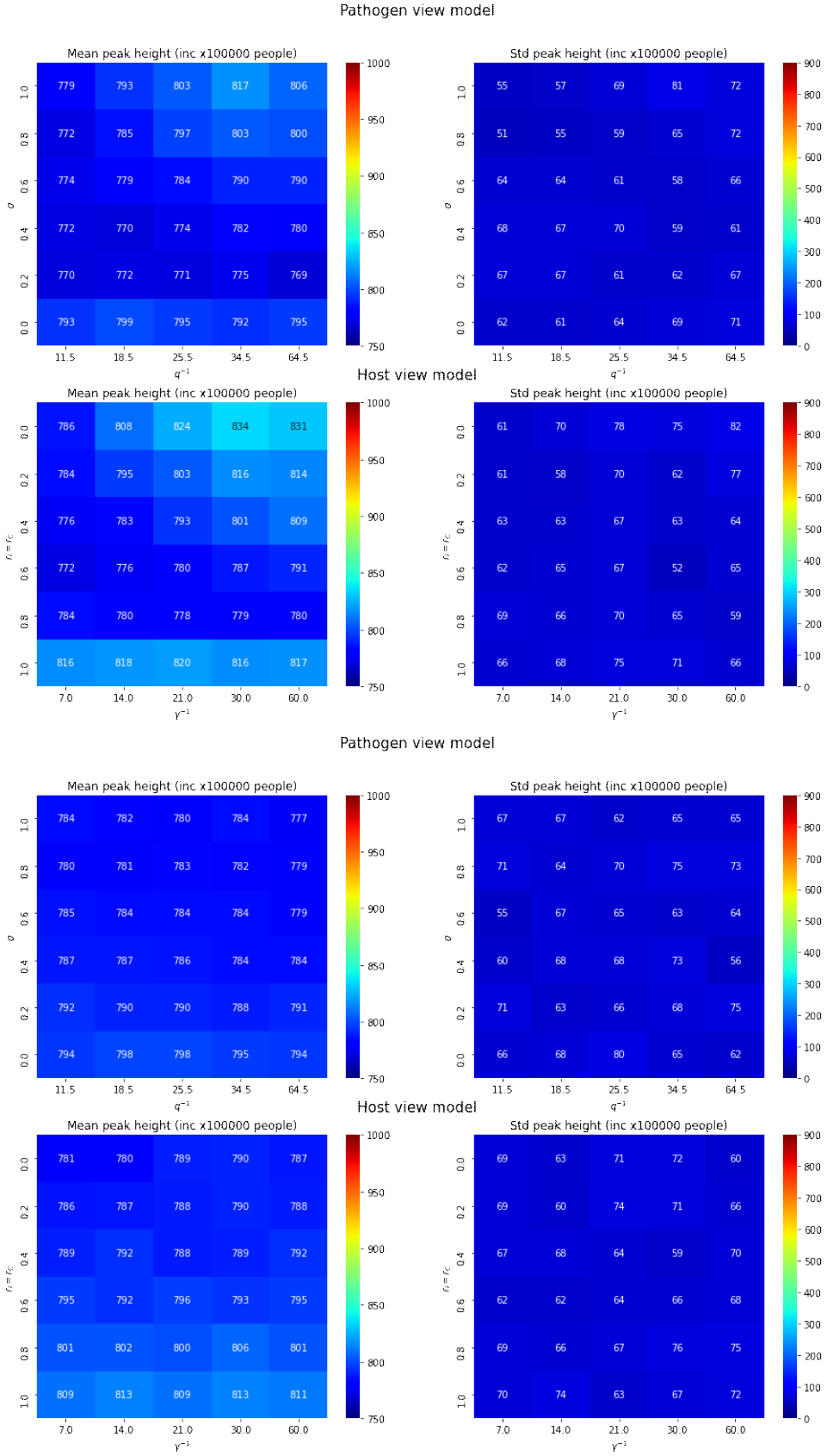

Figure A.4: Weekly incidence peaks on a 2 pathogens system. Case reference pathogen is the weakest one.
(a) $\beta_{0,\text{ref}} = 0.14 < \beta_{0,\text{other}} = 0.2$ (b) $\beta_{0,\text{ref}} = 0.14 < \beta_{0,\text{other}} = 0.15$. $\mu^{-1} = 4.5$ d, $\lambda^{-1} = 2 \times 365$ d.

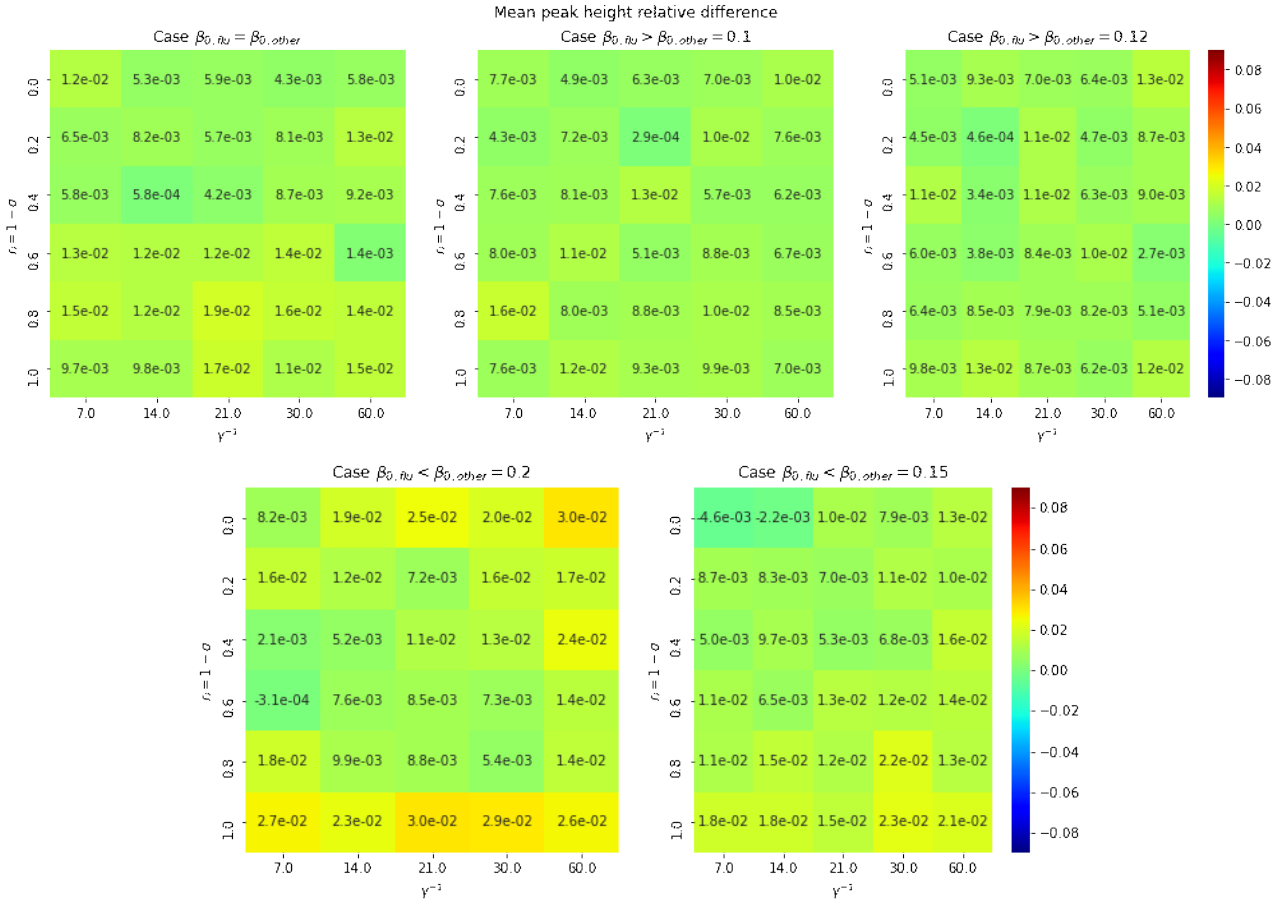

Figure A.5: 2 pathogens system - rel diff $= \frac{\text{mean p.h. host - mean p.h. path}}{\text{mean p.h. host}}$

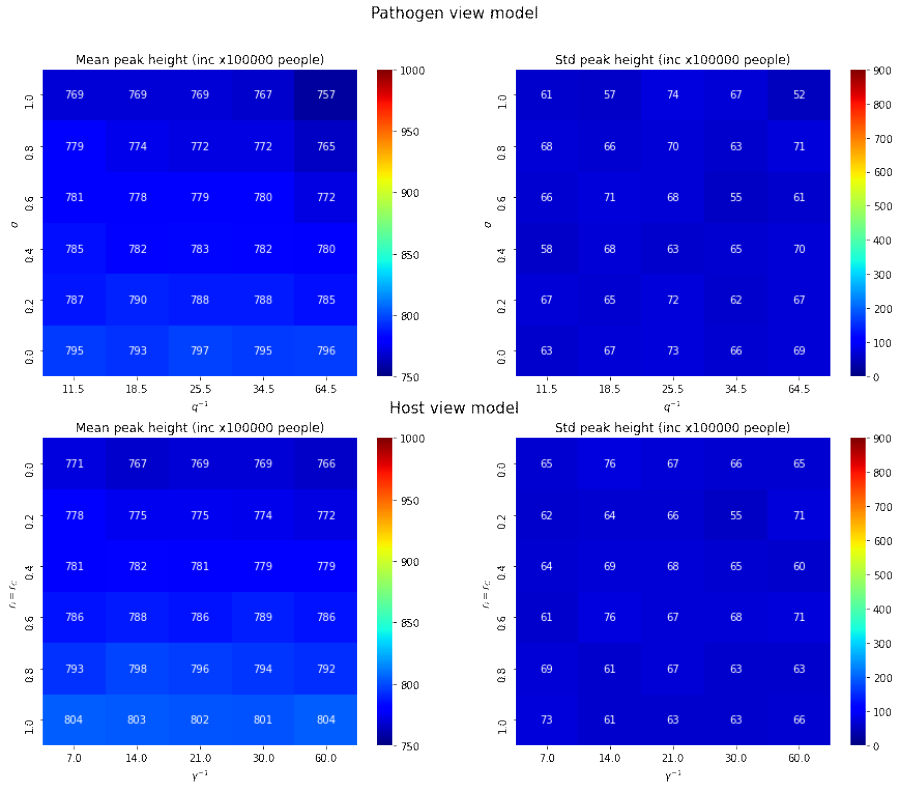

Figure A.6: Weekly incidence peaks on a 3 pathogens system. Case similar pathogens: $\beta_{0,\text{ref}} = 0.14 = \beta_{0,\text{others}}$, $\mu^{-1} = 4.5$ d, $\lambda^{-1} = 2 \times 365$ d.

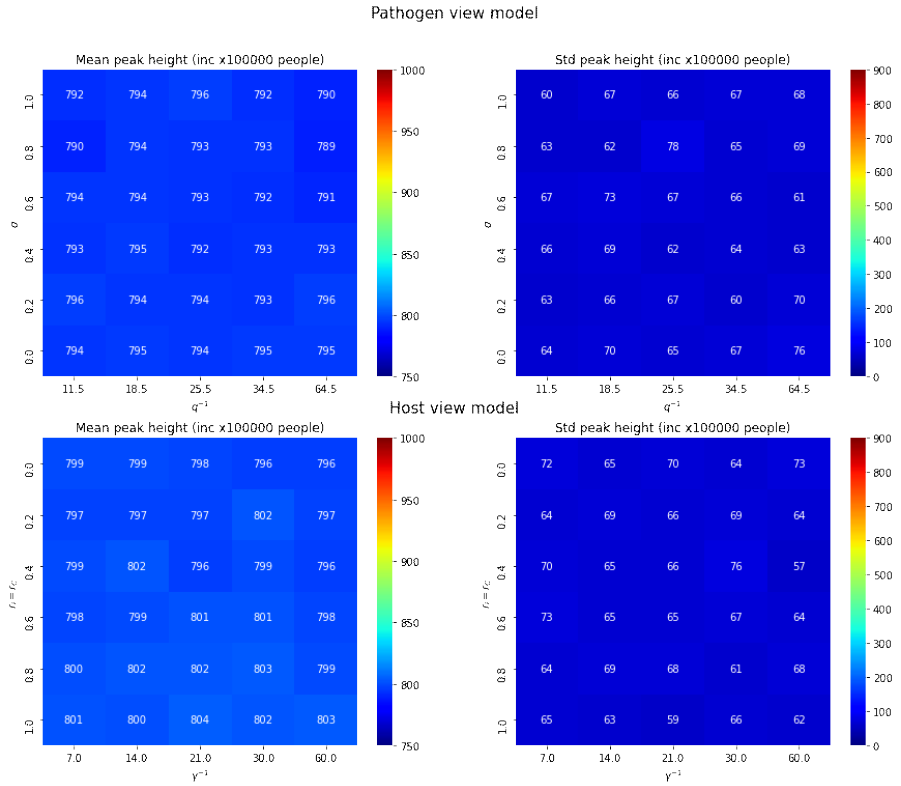

Figure A.7: Weekly incidence peaks on a 3 pathogens system. Case reference pathogen is the strongest: $\beta_{0,\text{ref}} = 0.14 > \beta_{0,\text{others}} = [0.1, 0.12]$, $\mu^{-1} = 4.5$ d, $\lambda^{-1} = 2 \times 365$ d.

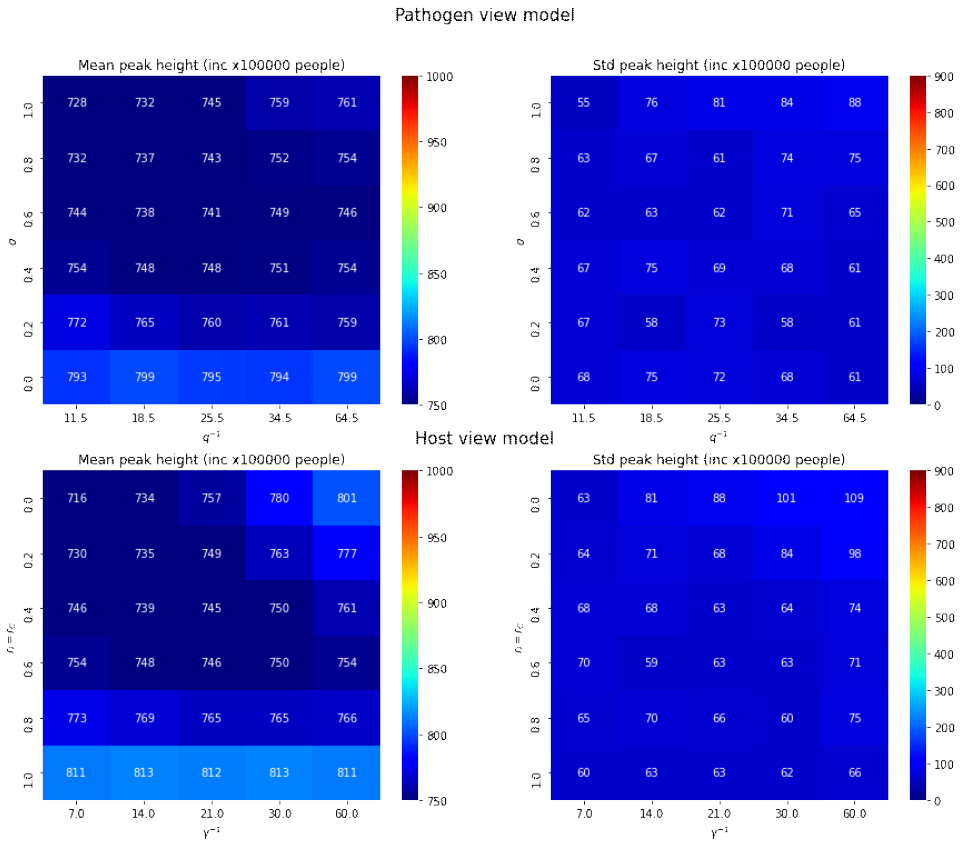

Figure A.8: Weekly incidence peaks on a 3 pathogens system. Case reference pathogen is the weakest: $\beta_{0,\text{ref}} = 0.14 < \beta_{0,\text{others}} = [0.2, 0.15]$, $\mu^{-1} = 4.5$ d, $\lambda^{-1} = 2 \times 365$ d.

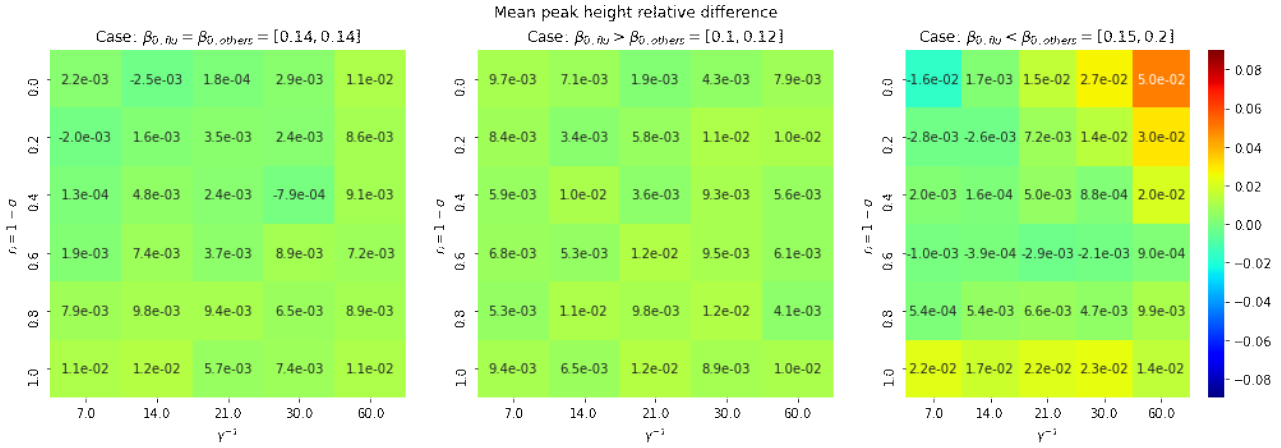

Figure A.9: 3 pathogens system - rel diff $= \dfrac{\text{mean p.h. host - mean p.h. path}}{\text{mean p.h. host}}$

Bibliography

[1] WHO. *Review of global influenza circulation, late 2019 to 2020, and the impact of the COVID-19 pandemic on influenza circulation*. `https://www.who.int/publications/i/item/who-wer-9625-241-264`.

[2] Emi Takashita et al. "Increased risk of rhinovirus infection in children during the coronavirus disease-19 pandemic". In: *Influenza and Other Respiratory Viruses* 15.4 (2021), pp. 488–494.

[3] Lulla Opatowski, Marc Baguelin, and Rosalind M. Eggo. "Influenza interaction with cocirculating pathogens and its impact on surveillance, pathogenesis, and epidemic profile: A key role for mathematical modelling". In: *PLOS Pathogens* 14.2 (Feb. 2018), pp. 1–28.

[4] Sema Nickbakhsh et al. "Virus-virus interactions impact the population dynamics of influenza and the common cold". In: *Proceedings of the National Academy of Sciences* 116.52 (2019), pp. 27142–27150.

[5] Matt Keeling and Pejman Rohani. *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press., 2007.

[6] Julia R. Gog and Bryan T. Grenfell. "Dynamics and selection of many-strain pathogens". In: *Proceedings of the National Academy of Sciences* 99.26 (2002), pp. 17209–17214.

[7] Miquel Porta. *A Dictionary of Epidemiology*. Oxford University Press, 2014.

[8] Daniel Bernoulli. "Essai d'une nouvelle analyse de la mortalite causee par la petite verole". In: *Mem. Math. Phy. Acad. Roy. Sci. Paris* (1766).

[9] John Snow. *Snow on cholera*. London: Humphrey Milford: Oxford University Press, 1936.

[10] Ronald Ross. *The prevention of malaria*. London: John Murray, 1911.

[11] William Ogilvy Kermack, A. G. McKendrick, and Gilbert Thomas Walker. "A contribution to the mathematical theory of epidemics". In: *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 115.772 (1927), pp. 700–721.

[12] M. P. WARD et al. "Estimation of the basic reproductive number (R0) for epidemic, highly pathogenic avian influenza subtype H5N1 spread". In: *Epidemiology and Infection* 137.2 (2009), pp. 219–226.

[13] Robin de Vries et al. "Systematic Screening for Chlamydia trachomatis: Estimating Cost-Effectiveness Using Dynamic Modeling and Dutch Data". In: *Value in Health* 9.1 (2006), pp. 1–11.

[14] R. M. Anderson et al. "The transmission dynamics of gonorrhoea: modelling the reported behaviour of infected patients from Newark, New Jersey". In: *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 354.1384 (1999), pp. 787–797.

[15] Linhua Zhou et al. "Global dynamics of a discrete age-structured SIR epidemic model with applications to measles vaccination strategies". In: *Mathematical Biosciences* 308 (2019), pp. 27–37.

[16] J. Ospina Giraldo and D. Hincapié Palacio. "Deterministic SIR (Susceptible Infected Removed) models applied to varicella outbreaks". In: *Epidemiology and Infection* 136.5 (2008), pp. 679–687.

[17] Paul Diaz et al. "A modified SEIR model for the spread of Ebola in Western Africa and metrics for resource allocation". In: *Applied Mathematics and Computation* 324 (2018), pp. 141–155.

[18] Yingcun Xia, Julia R. Gog, and Bryan T. Grenfell. "Semiparametric estimation of the duration of immunity from infectious disease time series: influenza as a case-study". In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 54.3 (2005), pp. 659–672.

[19] Yunxin Huang and Pejman Rohani. "Age-structured effects and disease interference in childhood infections". In: *Proceedings of the Royal Society B: Biological Sciences* 273.1591 (2006), pp. 1229–1237.

[20] Wan Yang, Eric H. Y. Lau, and Benjamin J. Cowling. "Dynamic interactions of influenza viruses in Hong Kong during 1998-2018". In: *PLOS Computational Biology* 16.6 (June 2020), pp. 1–21.

[21] Cleo Anastassopoulou et al. "Data-based analysis, modelling and forecasting of the COVID-19 outbreak". In: *PLOS ONE* 15.3 (Mar. 2020), pp. 1–21.

[22] Marc Baguelin et al. "Assessing Optimal Target Populations for Influenza Vaccination Programmes: An Evidence Synthesis and Modelling Study". In: *PLOS Medicine* 10.10 (Oct. 2013), pp. 1–19.

[23] You Li et al. "Global, regional, and national disease burden estimates of acute lower respiratory infections due to respiratory syncytial virus in children younger than 5 years in 2019: a systematic analysis". In: *The Lancet* 399.10340 (2022), pp. 2047–2064.

[24] Ting Shi et al. "Global Disease Burden Estimates of Respiratory Syncytial Virus Associated Acute Respiratory Infection in Older Adults in 2015: A Systematic Review and Meta-Analysis". In: *The Journal of Infectious Diseases* 222.Supplement7 (Mar. 2019), S577–S583.

[25] Pfizer. *Pfizer Announces Positive Top-Line Data from Phase 3 Trial of Older Adults for its Bivalent Respiratory Syncytial Virus (RSV) Vaccine Candidate*. `https://www.pfizer.com/news/press-release/press-release-detail/pfizer-announces-positive-top-line-data-phase-3-trial-older`.

[26] Naomi R. Waterlow et al. "Evidence for influenza and RSV interaction from 10 years of enhanced surveillance in Nha Trang, Vietnam, a modelling study". In: *PLOS Computational Biology* 18.6 (June 2022), pp. 1–14.

[27] Naomi R. Waterlow et al. "Competition between RSV and influenza: Limits of modelling inference from surveillance data". In: *Epidemics* 35 (2021), p. 100460.

[28] Lingcai Kong et al. "Compartmental structures used in modeling COVID-19: a scoping review". In: *Infectious Diseases of Poverty* 11 (2022), pp. 2049–9957.

[29] Roy Burstein et al. "Interactions among 17 respiratory pathogens: a cross-sectional study using clinical and community surveillance data". In: *medRxiv* (2022).

[30] Julie A. Spencer et al. "Distinguishing viruses responsible for influenza-like illness". In: *Journal of Theoretical Biology* 545 (2022), p. 111145.

[31] Andrew T. Pavia. "Viral Infections of the Lower Respiratory Tract: Old Viruses, New Viruses, and the Role of Diagnosis". In: *Clinical Infectious Diseases* 52.suppl-4 (May 2011), S284–S289.

[32] Nir Levy, Michael Iv, and Elad Yom-Tov. "Modeling influenza-like illnesses through composite compartmental models". In: *Physica A: Statistical Mechanics and its Applications* 494 (2018), pp. 288–293.

[33] WHO. *Global influenza programme*. `https://www.who.int/tools/flunet`.

[34] CDC. *Types of Influenza Viruses*. `https://www.cdc.gov/flu/about/viruses/types.htm`.

[35] Jerome I Tokars, Sonja J Olsen, and Carrie Reed. "Seasonal Incidence of Symptomatic Influenza in the United States". In: *Clinical Infectious Diseases* 66.10 (Dec. 2017), pp. 1511–1518.

[36] CDC. *How Flu Viruses Can Change: "Drift" and "Shift"*. `https://www.cdc.gov/flu/about/viruses/change.htm`.

[37] Talha N. Jilani, Radia T. Jamil, and Abdul H. Siddiqui. *H1N1 Influenza*. StatPearls Publishing, Treasure Island (FL), 2022.

[38] R.M. Greer et al. "Do rhinoviruses reduce the probability of viral co-detection during acute respiratory tract infections?" In: *Journal of Clinical Virology* 45.1 (2009), pp. 10–15.

[39] P. Rohani et al. "Population dynamic interference among childhood diseases". In: *Proceedings of the Royal Society of London. Series B: Biological Sciences* 265.1410 (1998), pp. 2033–2041.

[40] Wan Yang et al. "Epidemiological characteristics of the B.1.526 SARS-CoV-2 variant". In: *Science Advances* 8.4 (2022).

[41] Jeffrey Shaman et al. "Absolute Humidity and the Seasonal Onset of Influenza in the Continental United States". In: *PLOS Biology* 8.2 (Feb. 2010), pp. 1–13.

[42] Sen Pei and Jeffrey Shaman. "Aggregating forecasts of multiple respiratory pathogens supports more accurate forecasting of influenza-like illness". In: *PLOS Computational Biology* 16.10 (Oct. 2020), pp. 1–19.

[43] Alicia M Fry et al. "Seasonal Trends of Human Parainfluenza Viral Infections: United States, 1990–2004". In: *Clinical Infectious Diseases* 43.8 (Oct. 2006), pp. 1016–1022.

[44] Sema Nickbakhsh et al. "Epidemiology of Seasonal Coronaviruses: Establishing the Context for the Emergence of Coronavirus Disease 2019". In: *The Journal of Infectious Diseases* 222.1 (Apr. 2020), pp. 17–25.

[45] Samit Bhattacharyya et al. "Cross-immunity between strains explains the dynamical pattern of paramyxoviruses". In: *Proceedings of the National Academy of Sciences* 112.43 (2015), pp. 13396–13400.

[46] Réseau Sentinelles. *Weekly flu incidence trend in France*. `https://www.sentiweb.fr/france/fr/?page=epidemies`.

[47] Jeffrey Shaman and Alicia Karspeck. "Forecasting seasonal outbreaks of influenza". In: *Proceedings of the National Academy of Sciences* 109.50 (2012), pp. 20425–20430.

[48] Susanne Pinto et al. "Species abundance correlations carry limited information about microbial network interactions". In: *PLOS Computational Biology* 18.9 (Sept. 2022), pp. 1–21.