

PhD Dissertation



International Doctorate School in Information and
Communication Technologies

DISI - University of Trento

NEW-GENERATION INDIVIDUAL BASED MODELS
FOR INFECTIOUS DISEASES TRANSMISSION

Marco Ajelli

Advisor:

Dr. Stefano Merler

Fondazione Bruno Kessler, Trento

November 5th, 2009

Abstract

Mathematical models are powerful tools for simulating plausible epidemic spread scenarios and for evaluating the impact of control policies. They represent the scientific basis on which public health policy makers should take their decisions on the intervention strategies that should be performed at local, national and international scale. In this context, Individual-Based simulation Models (IBM) have become one of the much relevant approaches.

The crucial point of this thesis project is to override some of the limit of the current generation of IBM. Specifically, highly detailed models of the sociodemography and mobility of the Italian and European population have been developed; a model of individuals and households demographics, which leads the network of contacts among individual to evolve over time, has been introduced; an analysis of the role of different assumptions on the “random” contacts among the individuals of a population on the spread of epidemics has been performed.

Results such as the development, for the first time in literature, of an IBM working on a continental scale and of an IBM suitable for the investigation of endemic diseases represent a crucial improvement for the community of epidemic modelers. Moreover, the achieved results in terms of evaluation of the effectiveness of (individually-targeted) public health control measures have had a practical application. In fact, they have been used by the Italian Ministry of Health for assessing the efficacy of the Italian pandemic preparedness plan and for planning the mitigation strategies for the 2009 A(H1N1) influenza pandemic.

Keywords Agent-based models; computer simulations; epidemiology; control strategies; influenza.

Contents

1	Introduction	1
1.1	The Context	1
1.2	The Problem	2
1.3	The Solution	3
1.4	Innovative Aspects	4
1.5	Structure of the Thesis	5
2	State of the Art	7
3	The Problem	11
3.1	The evaluation of individually-targeted public health intervention for facing an influenza pandemic in Italy	11
3.2	Understanding the impact of different modeling assumptions	12
3.3	Working on a continental scale	13
3.4	Dealing with endemic diseases	14
4	The proposed approach	17
4.1	The Italian model	17
4.1.1	Simulated sociodemographic structure	17
4.1.2	Transmission models	19
4.1.3	Epidemiological parameters	21
4.1.4	Excess mortality	22
4.2	The European model	23
4.2.1	Sociodemographic structure of the European population	23
4.2.2	Mobility of the European population	25
4.2.3	Epidemic transmission details	28
4.2.4	Model parametrization	31
4.3	Modeling endemic diseases	33

4.3.1	Sociodemographic model	33
4.3.2	Epidemic transmission model	36
5	Experimental results	39
5.1	Understanding the impact of different modeling assumptions	39
5.1.1	Definition of the models of the unstructured contacts	39
5.1.2	How to compare different models	40
5.1.3	Results	42
5.1.4	Conclusions	48
5.2	Mitigating an influenza pandemic in Italy	51
5.2.1	Control measures	51
5.2.2	Results	54
5.2.3	Conclusions	59
5.3	Optimized antivirals administration during a pandemic outbreak	64
5.3.1	Discussion	72
5.3.2	Conclusions	75
5.4	Dynamics of an influenza pandemic in Europe	76
5.4.1	Population heterogeneity	76
5.4.2	Human mobility	78
5.4.3	Spatiotemporal spread of a pandemic influenza in Europe	80
5.4.4	Conclusions	84
5.5	Controlling hepatitis A in Italy	86
5.5.1	Results	90
5.5.2	Conclusions	96
6	Related work	101
7	Conclusion	103
	Bibliography	105

List of Tables

5.1	<i>Final attack rate.</i> Final attack rates (with standard deviation) of the different models considered for different G_0 values.	42
5.2	<i>Basic reproductive number.</i> Basic reproductive numbers R_0 (with standard deviation) of the different models considered for different G_0 values.	43
5.3	<i>Peak day.</i> Peak day (with standard deviation) of the different models considered for different G_0 values.	45
5.4	<i>Peak day for different seeding municipality.</i> Peak day (with standard deviation) for different seeding municipality and different values of the first generation index G_0	45
5.5	<i>Probability of a large outbreak.</i> Probability (with standard deviation) of a large outbreak for the different models considered and for different G_0 values.	46
5.6	<i>Mild scenario: Clinical attack rates and peak day, by control measure.</i>	55
5.7	<i>Moderate scenario: Clinical attack rates and peak day, by control measure.</i>	56
5.8	<i>Severe scenario: Clinical attack rates and peak day, by control measure.</i>	56
5.9	<i>Moderate scenario: Clinical Attack rates, peak day, peak daily attack rate and number of courses by combination of control measures.</i>	58
5.10	<i>Severe scenario: Clinical Attack rates, peak day, peak daily attack rate and number of courses by combination of control measures.</i>	60
5.11	<i>Epidemiological parameters.</i>	89
5.12	<i>Effectiveness of vaccination and improvements in hygienic conditions in Campania*.</i>	90
5.13	<i>Effectiveness of vaccination and improvements in hygienic conditions in Puglia*.</i>	91
5.14	<i>Effectiveness of social distancing measures in Campania*.</i>	92
5.15	<i>Effectiveness of social distancing measures in Puglia*.</i>	93

List of Figures

4.1	a Frequency distributions of household size for the different household types (in blue) and frequency distribution of the different household types (in red) considered in the model. b Age distribution from census data (blue) and simulated (red). c Proportion of workers for class of workplace from industry census data (blue) and simulated (red). d Probability density function of travel distances as obtained by using the gravity model (4.1) (in blue) compared with that obtained by using the the distance kernel (4.3) (in red).	19
4.2	a Age structure (year 2007) of Campania region (black line) and age structure simulated by the sociodemographic model (gray line). c Italian household size (black bars) and household size simulated by the model in Campania (gray bars). Note that data on household size are not directly employed in the algorithm. b and d as a and c , but for Puglia.	34
4.3	a Age structure of Campania (black line) and age structures simulated by the sociodemographic model: projections to the years 2012, 2017 and 2022 (from dark gray to light gray). c As a but for frequencies of household size. e As a but for frequencies of employment categories. b , d , f as a , c , e but for Puglia region.	35
5.1	Spreading time from the seeding municipality as a function of the distance for different values of G_0 and different seeding municipalities: Rome (first row), Cagliari (second row), a small isolated village (third row). Model M in orange, M+T in red, L in cyan, L+T in blue, and S in green.	41
5.2	Spatiotemporal dynamics at 40 (left), 50 (center) and 60 (right) days. Infection is seeded in Cagliari (Sardinia island) and $G_0=1.7$. Colored areas (model M in orange, M+T in red, L in cyan, L+T in blue, and S in green) indicate presence of at least one infected, infectious or removed individual.	44

5.3	Number of daily cases for different values of the first generation index and different seeding municipalities: Rome (first row), Cagliari (second row), a small isolated village in the north of Italy (third row). Simulation are initialized with 30 infected individuals, to reduce the stochastic variability observed in first days of the epidemic. Models are: M in orange, M+T in red, L in cyan, L+T in blue, and S in green.	47
5.4	Proportion of the infected population by age for $G_0=1.4$ (top) and $G_0=1.7$ (bottom). Models M in orange, L in cyan and S in green.	48
5.5	Baseline simulations under different R_0 scenarios (blue line: $R_0=2$; red line: $R_0=1.7$; green line: $R_0=1.4$). Bullet points represent the first Italian case and the time elapsed from the first world case.	53
5.6	Spatial spread of pandemic influenza in Italy, $R_0=1.7$. Red areas represent municipalities where at least one case is present.	54
5.7	Clinical AR, by control measure and scenario (Panel A: $R_0=1.7$; Panel B: $R_0=2$): black = baseline results; light blue = 90% air travel restriction; violet = AVP to household contacts; blue = vaccination, administering first dose within 3 months of the first world case for $R_0=1.7$, or within 2 months for $R_0=2$; grey = 90% air travel restriction + vaccination, as reported for the blue line; green = all control measures combined; red = all control measures combined, extending AVP to school/work close contacts.	61
5.8	Cumulative clinical AR, by age and scenario (Panel A: $R_0=1.7$; Panel B: $R_0=2$). Black line represents baseline results; red line represents the standard vaccination strategy (i.e., personnel providing essential services; elderly persons; and 2-18 year-olds); green line represents the effect of limiting vaccination to essential workers and children.	62

5.9 Timing for initial case for $R_0 = 1.4$ (green), $R_0 = 1.7$ (blue), $R_0 = 2$ (red) and $R_0 = 3$ (violet) in the baseline scenarios. Histograms are based on 100 simulations each. **b** Expected case incidence over time (solid lines) and 95% confidence intervals (shaded regions) based on 100 simulations for each scenario. Colors as in **a**. The black time window indicate a reasonable time interval for the availability of a pandemic vaccine. **c** Cumulative clinical attack rate by age (colors as in **a**), compared with data on the 1918-19 pandemic [56] (black line). The vertical dashed lines identify the age classes, namely young, adults and elderly, defined for age-prioritization of the use of antivirals. **d** Expected excess mortality by age classes (colors as in **a**) as obtained by assuming two different age-specific case fatality rates, similar to those estimated for the 1918-19 pandemic in Copenhagen (solid lines) and for the 1969-70 pandemic in Italy (dashed lines). Note that in the latter case, the expected excess mortality in the younger age classes (0-64 years old) is very close to 0 for all the R_0 values considered. 66

5.10 Clinical attack rates, peak day and peak daily clinical attack rate for baseline simulations (green), for antiviral treatment provided to index cases of all age classes (blue) or provided only to specific age classes (cyan, Y=young, A=adults, E=elderly), and for post-exposure prophylactic treatment provided to all age classes (red) or only to specific age classes (orange, Y=young, A=adults, E=elderly). When post-exposure prophylactic treatment is considered, we assume that antiviral treatment is also provided to index cases. 68

5.11 Expected excess mortality as obtained by assuming two different age-specific case fatality rates, similar to those estimated for the 1918-19 influenza pandemic in Copenhagen and for the 1969-70 influenza pandemic in Italy respectively, for baseline simulations (green), for antiviral treatment provided to index cases of all age classes (blue) or provided only to specific age classes (cyan, Y=young, A=adults, E=elderly), and for post-exposure prophylactic treatment provided to all age classes (red) or only to specific age classes (orange, Y=young, A=adults, E=elderly). 69

5.12 Antiviral stockpile required and number of avoided cases divided by the number of persons treated for antiviral treatment provided to index cases of all age classes (blue) or provided only to specific age classes (cyan, Y=young, A=adults, E=elderly), and for post-exposure prophylactic treatment provided to all age classes (red) or only to specific age classes (orange, Y=young, A=adults, E=elderly). When post-exposure prophylactic treatment is considered, we assume that antiviral treatment is also provided to index cases. The horizontal black line represents the Italian antiviral stockpile. 70

5.13 Sociodemographic structures. **(a)** The study area includes 37 countries and accounts for about 515 millions individuals (details are provided in Tab. S1). Colors from yellow to brown indicate increasing density of population. Black labels refer to countries belonging to EU27 while red labels refer to countries which do not belong to EU27. **(b)** Variability in the frequencies of household type at European level. A1 represents single persons, A1_CH single parents with children, CPL_NCH couples without children, CPL_CH couples with children. More than 95% of European households are structured as one of the four above mentioned types. **(c)** Variability in the frequencies of household size. **(d)** Variability in schools size (primary schools in cyan, secondary schools in blue). Horizontal lines identify the percentiles 25 and 75, the points represent the median values. The two boxplots represent the distributions of the average school size in the different countries. **(e)** Age structure curves in the different countries. **(f)** Variability in the employment and school attendance rates. Only individuals aged more than 15 are considered. In the model, individuals aged less or equal than 15 are assumed to attend schools. 77

- 5.14 Population movement patterns. **(a)** Network of yearly airplane travelers across Europe (colors are defined as follows. Green: less than 10,000 travelers, yellow: 10,000 to 100,000 travelers, orange: 100,000 to 1,000,000 travelers, red: 1,000,000 to 10,000,000 travelers, purple: more than 10,000,000 travelers). Each link between two countries is identified by an arc connecting the two capitals. **(b)** as **(a)** but for train travelers. **(c)** Probability density function of travel distances by train (green points), by airplane (blue points) and total (red points). Solid lines represent smooth interpolations of data. **(d)** Model A) (described in the main text; parameters: $\tau_f = 0.57$, $\tau_t = 0.99$ and $\rho = 0.39$): comparison between the observed and the modeled origin-destination matrix. Points compare generic entries of the two matrix and the solid black line represents a smooth interpolation. The model tends to overestimate the number of travelers when the actual yearly number of travelers is less than 1,000; it is in good agreement with the data on the most important links. **(e)** Model A): resulting probability density function of travel distances compared with that resulting from the analysis of the observed data (shown in **(c)**) (red points). **(f)** Internal commuting: probability density function of travel distances to school/workplace (in the model, red points), compared with that proposed in [58] for the radius of gyration of mobile phone users (black points). In the model, students are assumed to attend schools no more than 100Km from home. This results in a change in the slope of the probability curve (blue circle). 79
- 5.15 Spatiotemporal dynamics of a new pandemic influenza ($R_0 = 2$). **(a)** Probable destination of the first case imported in Europe. **(b)** Distributions of the day of the first national case (days since the first world case) in the different countries. **(c)** Distributions of the peak day (days since the first world case) in the different countries. **(d)** Expected number of daily cases per 100,000 individuals in time in Europe (red line) and 95% confidence intervals (shaded area). Green and blue lines (and shaded areas) refer to the expected number of daily cases per 100,000 individuals in Ireland and Bulgaria respectively. These two countries are among those where the impact of the epidemic is expected to be the highest and the lowest respectively. **(e)** Time sequence (in days) of a simulated epidemic. A single simulation with first European case in United Kingdom is shown. Colors from pink to dark red indicate an increasing number of daily cases (dark red indicates more than 10,000 daily cases). 81

5.16	Impact of a new pandemic influenza ($R_0 = 2$). (a) Distributions of the cumulative attack rate in the different countries. (b) Distributions of the peak daily attack rate in the different countries. (c) Average cumulative attack rate as a function of the average household size in the different countries. (d) Distributions of the percentage of cases due to transmission among household members in the different countries. (e) Distributions of the percentage of the population infected during long-distance travels across or outside Europe. (f) Average cumulative attack rate as a function of the fraction of inactive (neither students nor workers) individuals in the different countries.	82
5.17	a Risk factors by age group. Comparison between the National data reported in [94] (black bars) and the baseline simulations in Campania (dark gray bars) and Puglia (light gray bars). Children account for individuals aged 0–14, Young for individuals aged 15–24 and Adults for individuals older than 24 years old. b HAV seroprevalence by age cohort. Comparison between National data as reported in [8] (black bars) and baseline simulations in Campania (dark gray bars) and Puglia (light gray bars). c Age at infection of symptomatic individuals in Campania (black bars) and Puglia (gray bars). d Number of cases over time in three randomly chosen day care centers in Puglia ($R_0 = 3.8$), highlighting the classical “clusters-like pattern”. The numbers refer to the fraction of cases over the number of children enrolled.	94
5.18	Effects of a vaccination programme in Campania. Vaccination coverage is assumed to be 20% among newborns and 80% among 12 years old adolescents. a percentage of susceptible individuals over time (light gray area), recovered individuals (dark gray area), vaccinated individuals (black area) and infected individuals (“small” white area between susceptible and recovered individuals) for a vaccination programme lasting 5 years. b HAV seroprevalence by age cohort for a vaccination programme lasting 5 years. c Age at infection of symptomatic individuals (black bars) and asymptomatic individuals (gray bars) for a vaccination programme lasting 5 years. d, e, f as a, b, c, but for a vaccination programme lasting 20 years. g, h, i as a, b, c, but for a vaccination programme lasting 50 years.	95

5.19	Effects of a vaccination programme in Puglia. Vaccination coverage is assumed to be 20% among newborns and 80% among 12 years old adolescents. a percentage of susceptible individuals over time (light gray area), recovered individuals (dark gray area), vaccinated individuals (black area) and infected individuals (“small” white area between susceptible and recovered individuals) for a vaccination programme lasting 5 years. b HAV seroprevalence by age cohort for a vaccination programme lasting 5 years. c Age at infection of symptomatic individuals (black bars) and asymptomatic individuals (gray bars) for a vaccination programme lasting 5 years. d, e, f as a, b, c , but for a vaccination programme lasting 20 years. g, h, i as a, b, c , but for a vaccination programme lasting 50 years.	96
5.20	a Social distancing measure $d\beta$ applied to all symptomatic individuals: variation of the number of cases (gray bars) and number of symptomatic cases (black bars) in Campania by age cohorts. b As a but for Puglia region.	97

Chapter 1

Introduction

1.1 The Context

At the time of writing (August 6, 2009), a new subtype of influenza A(H1N1) virus is rapidly spreading worldwide, with over 171,000 cases and over 1,500 deaths [39]. Developing measures for controlling the ongoing (and future) influenza pandemics represents a crucial challenge for public health agencies worldwide. In fact, during the last century there were three influenza pandemics, the most devastating in 1918–19 which killed 20 to 50 million people worldwide [80]. Considering that since 1918 the world’s population has more than tripled and that global traveling and urbanization have increased dramatically, it is easy to understand the pessimistic scenarios predicted in many recent studies in case of a future pandemic, with huge human and economic losses [109]. This has motivated intensive pandemic planning in many countries, with several intervention strategies being suggested by the World Health Organization (WHO), involving both medical and public health countermeasures. Measures such as travel restrictions, schools and workplaces closure, antiviral treatment and prophylaxis, vaccination, case isolation, quarantine may all be useful, but the need clearly arises to carefully and timely study their effectiveness and feasibility. Non-pharmaceutical interventions involving social distancing may have negative consequences (e.g., in terms of acceptability by the population) and, in general, can be helpful for slowing down the course of an epidemic rather than to minimize its impact. Antiviral availability will be limited for most of the world’s population and the antiviral stockpile is below the level recommended by WHO also for many occidental countries [121]. Vaccine needs time for the production: its availability is estimated to be of about 4 months at best [111] (other estimates are 6 months at best [128, 112]).

In this context, mathematical models are powerful tools for simulating realistic pan-

demographic spread scenarios and for evaluation of the potential impact of control policies. They may be useful to identify optimal strategies, to set priorities and to establish criteria for deployment and use of antivirals and vaccine and, more in general, for testing the effectiveness of containing/mitigating strategies included in national pandemic preparedness plans [114, 124]. Recent work has illustrated the ability of mathematical models to predict the course of actual epidemics, and to help policy makers in the design of control measures, e.g., in the 2001 foot-and-mouth disease epidemic in UK cattle [48] and in the 2002–2003 global outbreak of severe acute respiratory syndrome (SARS) [116].

Specifically, in the case of person-to-person transmitted diseases, households, schools and workplaces often represents the main places where transmission occurs because individuals have a lot of contacts [105] and spend a lot of time [133] in such contexts. Therefore, from a public health point of view, control measures aimed to reduce the number of contacts in these locations (e.g., schools closure or case isolation) or the administration of drugs to household or school/workplace contacts of a clinical case (e.g., antiviral prophylaxis) can play a key role in the control of an epidemic.

In IBM, individuals and the main places where virus transmission can occur are explicitly represented. Therefore, since they allow the explicit representation of the actual locations where intervention measures will be implemented to reduce transmission, they are currently considered the best tools for testing the effectiveness of control strategies. In this context, the main effort has been done for the evaluation of strategies for containing/mitigating a new influenza pandemic [46, 47, 85, 87, 54, 132, 29, 35, 64, 96], and for testing control measures in response to a bioterroristic smallpox attacks [42, 86, 65, 115].

1.2 The Problem

In this PhD thesis we have faced the following problems. The first is the evaluation of highly realistic individually targeted public health interventions for mitigating an influenza pandemic (as the ongoing A(H1N1) epidemic) in Italy by developing an individual based model. The main aim of this part of the project is giving precise indication on the course of a pandemic influenza to the public health policy makers, in order to minimize the impact of the epidemic in terms of number of cases, severity of the infections, number of individuals simultaneously ill and the economic burden.

The second point is to extend, for the first time, individual based models to a continental scale. Nowadays, IBM has been used for describing epidemic outbreak on a community scale (e.g., see [85]), on a city level (e.g., see [42]) or on a country level (e.g., see [47, 54, 29]). However these models have not been developed at continental scale and

thus it is still how epidemics could spread at this scale and what extent the heterogeneity of mobility patterns and population heterogeneity affect its spread. The main issues we have had to face for working on a continental scale concerns *(i)* the collection of a huge amount of sociodemographic data on each country of the considered geographic area; *(ii)* the development of new more efficient algorithm and data structure for dealing with hundreds of millions of individuals; *(iii)* the development of a model of the highly variable sociodemographic structure of the populations of a continent and *(iv)* the development of a model of their mobility.

As highlighted by Riley [114], the need arises “to develop a simple model of household demographics, so that these large-scale models can be extended to the investigation of long-time scale human pathogens”. In fact, endemic diseases such as tuberculosis, measles, malaria, hepatitis and HIV represent important public health issues worldwide [106]. One of the aim of this project is to extend IBM in order to allow them describing also endemic diseases. To such aim it is required to integrate birth and death processes together with a model of household demography. Again, the use of IBM allows the evaluation of both mass and individually targeted public health interventions aimed at controlling/limiting the impact and severity of endemic diseases.

Finally, since for the transmission of a human-to-human infectious diseases places where individuals have contacts with other individuals could be the source of infections, they can potentially play a relevant role in the spread of an epidemic. Unfortunately, even if information about the number and the duration of such kind of contacts are currently becoming available [105, 133], their explicit representation in the models are still debated [114]. Therefore, we investigate the impact of different assumptions on these kind of “unstructured” contacts on the course of simulated epidemics.

1.3 The Solution

In order to reach the first objective briefly described in Section 1.2, the main ingredient has been the development of a highly detailed model of the Italian sociodemography. This was possible thanks to the collection of detailed sociodemographic data (e.g., on the age-structure, on households composition, on schools and workplaces sizes, on national and international travels). This allows us to test public health interventions and giving precise indications to the policy makers.

For quantifying the impact of the “arbitrary” assumptions on modeling the unstructured contacts we re-implement the main approaches used in literature and we made a systematic comparison of the simulated epidemics.

For extending IBM to a continental scale we collected a huge amount of data, similar to that used for the Italian model, for each considered country. After the data collection, the one of the problems we faced has been the computational cost of the previously developed algorithms and the large RAM memory required by the simulations. Therefore, we have developed new data structure and more efficient algorithms to overcome the computational difficulties. Moreover, from a modeling point of view, we spent a big effort on developing a reliable model of human mobility, which takes into account both daily short-distance travels (e.g., movements between the places of residence and work) and occasionally long-distance travels.

For describing the dynamics of long-time scale human pathogens we have developed a model of households demographics and the network of contacts among individuals has become dynamic. This allows us to investigate endemic diseases through IBM simulations.

1.4 Innovative Aspects

This project contributes to the advancement of the knowledge in modeling infectious diseases dynamics. In particular, it contributes to a better understanding of the consequences of the hypotheses usually made on modeling the random contacts component among the individuals of a human population.

Our evaluation of the mitigation strategies in response to a diffusion of a new influenza pandemic are currently contributing to controlling the A(H1N1) pandemic in Italy. Our evaluation of an age-prioritized use of antiviral drugs certainly represents an original contribution to the fight against influenza pandemics.

From a completely different point of view, thanks to the great improvements in terms of algorithms and data structure, for the first time in literature we have been able to propose a continental scale IBM. The European model, developed within this project, probably represents the more detailed model for describing an epidemic spread on a large-scale. This model allowed us to study the effects of population heterogeneity and human mobility on the spread of an epidemic.

Finally, the development of a dynamic network of contacts and a model of households demography allowed us to extend IBM to the analysis of endemic diseases. This can be considered a crucial step to the identification of more suitable strategies for controlling endemic diseases (e.g., tuberculosis, HIV, hepatitis), which represent important public health issues worldwide [106].

1.5 Structure of the Thesis

This thesis is structured as follows. In the next chapter, a brief review on mathematical models of human infectious diseases transmission is carried out. This is mainly focused on individual based models and it highlights the main assumptions and features of the current “state-of-the-art” IBM. In chapter 3, the problems we have faced during the thesis project are described.

In chapter 4, the approaches used for dealing with the problems are described in details. Specifically, three main types of models are introduced. First, in Sec. 4.1, the model used for testing the effectiveness of different intervention options for mitigating an influenza pandemic in Italy is described. Second, in Sec. 4.2, the continental-scale IBM is introduced. Third, in Sec. 4.3, it is shown how our IBM have been extended in order to investigate endemic diseases.

Chapter 5, the main results of this project are presented. In particular, in Sec. 5.1, it is shown to what extent is the impact of different modeling assumptions on the general (“random”) contacts among the population on the spread of an influenza pandemic. In Sec. 5.2 and 5.3, detailed individually targeted strategies (considered by health policy makers) for mitigating an influenza pandemic in Italy are evaluated. In Sec. 5.4, the role of sociodemographic structure and mobility of the European population on the spread of an influenza pandemic at continental scale. Finally, In Sec. 5.5, the effects of different options for controlling hepatitis A (which ranges from mass vaccination to individually targeted interventions) are shown.

In chapter 6, a comparison with current “state-of-the-art” IBM developed by other authors is performed. Finally, chapter 7 is dedicated to summarize the work made during this thesis project and to highlight possible future directions of IBM for infectious diseases transmission.

Chapter 2

State of the Art

Depending on the diseases and on the intervention options considered, different models can result a valuable choice to describe the spatiotemporal evolution of an epidemic and to assess the impact of containment/mitigation strategies. Classical deterministic compartmental models based on Kermack and McKendrick's work [82] have the main advantage of transparency and simplicity. They are easy to develop and fast to solve, allowing for rigorous sensitivity analysis to explore the dependence of model output on uncertainty of the parameters. They can be used, either in simple and in structured form (considering age structure [37] and/or geographic component [126]), for describing the temporal dynamics of an epidemic and for the assessment of some containment or mitigation strategies, such as mass vaccination [37] or border restrictions [31]. On the other hand, the evaluation of realistic, individually targeted, public health intervention strategies, such as antiviral prophylaxis of household or school/workplace contacts of index cases, in turn requires highly detailed models. Spatially explicit models provide a plausible system in which the precise spatial location of individuals and movement patterns can be employed to evaluate intervention options [114]. Spatial models can be broadly divided into three major categories: patch models, distance-based models, network models. In patch models, the force of infection (FOI) received by an individual living in a patch (e.g., a town) depends on the distance between the patch of residence and the patches of the infectious individuals: all the members of a patch receive the same FOI. In distance-based models individuals are assigned a precise location and the FOI is a decreasing function of the distance between susceptible and infectious individuals. In network models individuals are connected to other individuals by co-locating them into groups (e.g., households, schools and workplaces, etc.). Additionally, the members of a group can be not equally well connected. For instance, in large schools or workplaces subgroups of individuals, representing classmates

or close colleagues, can be more strongly connected. The FOI received by a susceptible individual is non zero when he/she shares an arc with a infectious individual.

To simulate either of the previous models (patch, distance-based, and network) Individual Based Models, also called Agent Based Models, can be used. An IBM is a model for simulating the interaction of individuals (or agents) and it is the common way to simulate a network model (in this proposal, the term IBM refers to network models for infectious diseases where individuals are explicitly represented). In fact, incorporating local contact network structure in classical models is difficult: for instance, households (or schools/workplaces) cannot easily be represented as structures distinct from patches, despite household transmission occurring at a much higher rate than community transmission (which can be considered an unstructured contacts component). Moreover structures as households and schools/workplaces play an important role in the assessment of control measures because they are often considered targets for interventions. The inclusion of this type of population structure together with age structure (coming from realistic demography) and the possibility to model the course of the illness in each individual, makes IBM more suitable.

Making a step back to the main question of responding to the emerging of a new influenza pandemic, network models (and thus IBM) allow the evaluation of intervention measures targeted at individual level, in locations such as households, schools, workplaces, transportations and hospitals which are among the most important routes of influenza transmission. Since households are considered the main route of influenza transmission, as shown in [132], it is not possible to neglect this structure to test efficient mitigation strategies. Many efforts were done in this context. In particular Ferguson et al. propose large-scale (national level) IBM for assessing containment strategies in South-East Asia [46] and mitigation strategies in US and UK [47]. These models are network models (explicitly modeling households, schools and workplaces) where also the distance between individuals is considered (to describe the random component of the transmission). For the first time in [46] was proposed an IBM at national level (accounting for millions of individuals); super-computing techniques were adopted to overcome the computational problems of these models, in terms of RAM memory consumption and computational cost of the algorithm. In [85] a network model is employed for assessing the effectiveness of containment strategies based on antiviral prophylaxis in a typical American community. In this paper a higher level of details on the sociodemographic structure of the population is reached (e.g. local neighborhood and playgroups). This work was extended later to the entire US population [54], so the authors were able to consider also the connections among communities and a realistic national level epidemic was simulated. Another milestone is

the model proposed by Longini et al. in [87], where a network model is employed for assessing the chance of containing at the source an influenza pandemic mostly with antiviral agents. Some of the above mentioned approaches, and other modeling choices, were compared in [2] to understand to what extent different modeling choices for unstructured contacts can affect pandemics prediction and control options.

IBM are commonly employed to describe spatiotemporal evolution of smallpox outbreaks (mainly related to bioterrorism attacks) and to evaluate the impact of containment strategies [65, 115]. From the point of view of public health, it would be useful to evaluate with these models control options for other diseases such as SARS, Marburg, Ebola and Chikungunya.

Going more into detail of the current generation of IBM, networks among individuals are implemented on the basis of data on the sociodemographic structure of the population considered. There exists the need to collect detailed data on the spatial and age distribution of the population, households types and their composition, travel data (both for incorporating commuting destinations and occasional trips), international travels data (for determining incoming cases from abroad), etc. In principle, in IBM it would be possible to characterize the specific contact network, by employing data on every type of contact (e.g. contacts on neglected but important activities, such as leisure time, sport mall, restaurants, etc.) and time-use data (e.g. time spent in school/workplaces, at home, at the canteen, etc.). Therefore IBM represent the most obvious way to relax assumptions that were considered mandatory in traditional mathematical models of diseases, such as the assumption of homogeneous mixing, which most mathematical models are based on. As a first step in the construction of an IBM, contacts are progressively “structured” by co-locating individuals into the diverse environments where they are expected to have contacts, namely households, school, workplaces, commuting and public transportations and so on. However, one readily realizes that most epidemics initiating from a single focus would die out or would not travel, unless a degree of (pseudo)-randomness in contacts, perhaps small, is allowed. Though largely unknown, the impact of the unstructured component on the epidemics dynamics is not necessarily small: in fact it often prevents epidemics to die out and allows more structured contacts to amplify the epidemics themselves, thus leading to potentially devastating outbreaks.

Current generation of IBM have some disadvantages: they are computationally intensive; they do not allow continental/world scale simulations; they are not suitable for describing endemic diseases; they are much harder to parameterize and validate than classical models (e.g. the sociodemographic structure of the population is thorough, but hard to implement and update during a simulation). Some of these limitations have been

overtaken during this thesis project.

Chapter 3

The Problem

3.1 The evaluation of individually-targeted public health intervention for facing an influenza pandemic in Italy

As it is proved by the ongoing A(H1N1) pandemic, facing an influenza pandemic is a real threat. Therefore, obtaining reliable evaluation of the effectiveness of individually targeted public health interventions for controlling an influenza pandemic is crucial for public health policy makers worldwide. Recently, the best solution for this problem has been represented by IBM [114], which have been largely used for predicting the spatio-temporal spread of pandemic influenza and for assessing the efficacy of strategies for containing or mitigation [95, 2, 87, 46, 85, 54, 47, 132, 55, 29, 35, 64, 96] the epidemic. Therefore, one of the problem we faced during this thesis project was to develop an IBM at national (Italian) scale. The main aim of this model is to give precise indication on the dynamics of a pandemic influenza to the policy makers, in order to limit the impact of the epidemic in terms of number of cases and death, severity of the infections, number of individuals simultaneously ill and the economic burden.

Basically, the strategies which would be tested include vaccination, antivirals treatment and/or prophylaxis, social distancing measures and travel restriction. As regards the vaccination, it is relevant to understand which is the optimal target population; e.g., it can be provided to specific age classes, to at risk individuals, to essential workers or it can be a massive. However, it is crucial also to evaluate the effects of the timing of the vaccination: e.g., a late vaccination would results in a very low effect on limiting the epidemic, at the cost of a relevant the economic burden. Antivirals, used for both treat-

ment of cases and/or prophylaxis, can be used for mitigating the spread of an influenza pandemic as long as a pandemic vaccine is not available. However, in some countries the antiviral stockpile exceeds the number actually required for the treatment of all cases [121]. Therefore, the problem of prioritizing the use of antivirals would be a crucial issue. Non-pharmaceutical interventions, such as travel restrictions and social distancing measures, can delay the epidemic arrival and peak, but without reducing the overall impact of the epidemic. However, understanding which non-pharmaceutical intervention it is better to perform and when their impact is much relevant is another point that the model should be able to address.

For developing a model able to reach such targets, we need to collect highly detailed data on sociodemographic structure (e.g., age-structure, households size and composition, school size) and mobility (e.g., number of international airplane passengers, distance between place of residence and work) of the Italian population. Another issue we have had to address was the computational burden required (both in terms of RAM memory and number of CPU cycles) for simulating epidemics. In this thesis it will be highlighted how such problem were faced and overtaken.

3.2 Understanding the impact of different modeling assumptions

IBM represent a plausible way to relax assumptions that were considered mandatory in traditional mathematical models of diseases, such as the assumption of homogeneous mixing, which most of mathematical models are based on.

The sources of infections for human-to-human infectious diseases are the places where individuals have contacts with other individuals. Therefore, households, schools, workplaces, but also cinemas, malls, prisons, public transport systems, gyms, etc. play a crucial role in the spread of an epidemic. Reliable data on households, schools and workplaces (e.g., on the size or on the composition) are available. For this reason, in some sense, we can consider this contact component as “known” or “structured”. On the contrary, for all the other kind of contacts among persons, we have limited information (even if estimations about the number and the duration of such kind of contacts are currently becoming available [105, 133]).

However, one readily realizes that most epidemics initiating from a single focus would die out or would not travel, unless a degree of (pseudo)-randomness in contacts, perhaps small, is allowed. Though largely unknown, the impact of the unstructured component

on the epidemics dynamics is not necessarily small: for instance, many current flu models are based on the explicit assumption that unstructured contacts account for about 1/3 of the total risk of infection per unit time [46]. Moreover, unstructured contacts prevent epidemics to die out and allow more structured contacts to amplify the epidemics themselves, thus leading to potentially devastating outbreaks. The currently available generation of infectious diseases IBM have achieved a sophisticated descriptions of the structured component of contacts, but the unstructured component continues to be loose because of scarce information on unstructured ones which are however the sustaining factor. What concretely happens is that unstructured contacts are modeled in a residual way, mainly reflecting the researcher feeling, and at best are left as free simulation parameters.

The purpose of this part of the thesis project is to evaluate to what extent different modeling strategies for unstructured contacts can affect pandemics prediction and control. Motivated by the issue of modeling interventions aimed at containing a national flu pandemic, we provide a comparison of various alternatives to model the unstructured component. These alternatives include the main approaches proposed in the literature and comprehend some new techniques. In particular, we keep the structured component fixed, and we vary the unstructured one, looking at the implications in terms of the major epidemic outputs, as fade-out and extinction probabilities, spatial traveling, attack rates, and proportion of infected individuals by age.

3.3 Working on a continental scale

Another point is to extend, for the first time, individual based models to a continental scale. Nowadays, IBM has been used for describing epidemic outbreak on a community scale (e.g., see [85]), on a city level (e.g., see [42]) or on a country level (e.g., see [47, 54, 29]). The main issues we have had to face for working on a continental scale concerns *(i)* the collection of a huge amount of sociodemographic data on each country of the considered geographic area and *(ii)* the development of new more efficient algorithm and data structure for dealing with hundreds of millions of individuals.

Specifically, the spread of an infectious disease epidemic is driven by the interplay of two factors: the transmissibility of the virus responsible for the infection and the characteristics of the host population. When the role of host is played by a human population, predicting the spread of an epidemic is a tough problem due the complexity of modern human societies. It is well established that the spatial structure of the population has an impact on the diffusion of an epidemic: measles waves in England and Wales, spreading from large cities to small towns, are determined by the spatial hierarchy of the

host population structure [61], and the spatial distribution of farms influences the regional variability of foot-and-mouth outbreaks in United Kingdom [81]. The heterogeneity of the population itself can play an important role in the spread of an epidemic [38]. It is also well known that human mobility patterns affect the spatiotemporal dynamics of an epidemic: the role played by the airline transportation network has been analyzed in [30], and it has been shown that the high degree of predictability of the worldwide spread of infectious diseases is caused by the strong heterogeneity of the transport network [68].

However, Europe has never been analyzed as a whole and thus it is still uncertain how a new pandemic influenza could spread in Europe. Europe comprises countries characterized by completely different social and economical backgrounds that result in different levels of population heterogeneity, in terms of both sociodemographic structure and mobility. In particular, it is still unclear how differences in the sociodemographic structure, which result in different levels of population heterogeneity, and different patterns of human mobility can affect the spatiotemporal spread of an epidemic.

3.4 Dealing with endemic diseases

As highlighted by Riley [114], the need arises “to develop a simple model of household demographics, so that these large-scale models can be extended to the investigation of long-time scale human pathogens”. In fact, endemic diseases such as tuberculosis, measles, malaria, hepatitis and HIV represent important public health issues worldwide [106]. One of the aim of this project is to extend IBM in order to allow them describing also endemic diseases. In fact, “traditional” IBM have been used only for the investigation of epidemic outbreaks (which duration can be consider one year at most). Therefore, the network of contact among individuals can reasonably be considered static. On the contrary, when we deal with endemic diseases, the investigated temporal scale is of several years. This implies that the network of contacts among individual has to be dynamic. Therefore, for extending IBM to the investigation of endemic diseases we have to develop a model for the dynamics of contacts among individuals.

Again, by using such kind of model allows the evaluation of both mass and individually targeted public health interventions aimed to controlling/limiting the impact and severity of endemic diseases.

Among endemic diseases, tn this part of the thesis project, we decide to focus our attention on the dynamics of viral hepatitis A in Southern Italy. Hepatitis A virus (HAV) is the cause of viral hepatitis A infection, which results in an acute form of hepatitis. Patients recover completely and develop full immunity against future HAV infections

[122]. Nevertheless, viral hepatitis A represents an important public health issue, imposing a remarkable economic burden worldwide [15, 34] and in Italy as well [90]. HAV is transmitted via the fecal–oral route by person–to–person contact (direct transmission) or by ingestion of contaminated food or water (indirect transmission). Indirect transmission represents the most important source of infection in countries with high living standards [49, 89]. Specifically, raw mussels and shellfish consumption represents the main source of infection in Italy [94], especially in the most affected regions: Puglia and Campania [24, 25, 118]. Another significant source of infection is represented by travels to high endemicity areas [94], where both direct and indirect transmission can occur. Nowadays, even in the absence of vaccination, hepatitis A is in a decaying phase, mainly determined by improved hygienic conditions derived from economic development and higher standards of living [77, 78]. Although this is true also for Italy as a whole, as documented by both notification [94, 72] and serological data [33], Southern Italy shows a different pattern. HAV infections are still common in Puglia and Campania, two regions in Southern Italy, where a very large outbreak was observed in 1996–1997 [90], despite the improvement of socio–economic conditions.

Thanks to a new generation of mathematical modeling tools, the effectiveness of both pharmaceutical (e.g., treatment or prophylaxis on a contact tracing basis) and non–pharmaceutical (e.g., social distancing) individually targeted intervention measures can be thoroughly investigated. Since hepatitis A is a vaccine–preventable disease, by using our model, we want to evaluate the effectiveness of different vaccination programmes. Moreover, we want to analyze the effects of improvements in standards of living and hygiene and of social distancing measures, such as isolation of symptomatic cases and closure of day care centers and kindergartens.

Chapter 4

The proposed approach

4.1 The Italian model

We have developed an individual based models for the transmission of an influenza like illness. Basically, the model simulates a synthetic population of agents representing every Italian individual. Among this population the infections can be transmitted from agent-to-agent by a stochastic simulated transmission process. Therefore, the model has two main ingredients: a sociodemographic model and a transmission model.

4.1.1 Simulated sociodemographic structure

Households In the national model, individuals were randomly grouped in households to match the 2001 census data (Italian Institute of Statistics: XIV Censimento generale della popolazione e delle abitazioni, 2001. Available at url <http://dawinci.istat.it/MD/>) on age structure and data from a specific 2003 survey (Italian Institute of Statistics: Strutture familiari e opinioni su famiglia e figli, 2003. Available at url http://www.istat.it/dati/catalogo/20060621_03/) on household size and composition. Nine different types of households were considered (e.g., singles or couples, with or without children, with or without additional members, adults living together) and individuals were co-located in households according to specific data on the percentage of the different household types, their size, the age of the household head. Frequency distribution of household sizes for the different household types are shown in Fig. 4.1a, together with the frequency distribution of the household types. The availability of these data allowed us to develop a very realistic model of the mixing of the age classes within households. The resulting age structure of the population is shown Fig. 4.1 and it agrees well with the 2001 census data.

Schools and workplaces The Italian population at 2001 is structured as follows: 20,559,595 workers, 11,360,556 students and 25,084,274 unemployed or retired. Children and young adults were assigned to one of six levels of school (i.e., from day care to university) on the basis of age and specific data on school attendance by age (Italian Ministry of University and Research: *La scuola in cifre*, 2005. Available at url <http://www.miur.it/ustat/documenti/pub2005/>). Italian Ministry of University and Research: *L'università in cifre*, 2005. Available at url <http://www.miur.it/ustat/documenti/pub2005/>). Attendance to school varies widely with age: it ranges from 14% in day care centers, to 90% in kindergartens, approximately 100% in primary and middle schools, 82% in high schools, 31% in university. We used specific data on employment rate by age in Italy to assign an employment to individuals aged more than 15 years. Workers were randomly assigned to one of seven employment categories, defined by the number of employees in the workplace (see Fig. 4.1c) (Italian Institute of Statistics: *VIII Censimento generale dell'industria e dei servizi*, 2001. Available at url <http://dwcis.istat.it/cis/index.htm>). Teachers and school employees were also considered in the model.

Commuting We modeled travel destinations by using specific Italian data on travels between place of residence and place of work or study. Specifically, we used a gravity model [126], in which the probability of commuting from municipality i to municipality j increases with the population sizes and decreases with the distance:

$$C_{ij} = \theta \frac{p_i^{\tau_d} p_j^{\tau_r}}{d_{ij}^{\rho}}, \quad (4.1)$$

where p_i and p_j represent the number of individuals living in municipality i and j respectively and d_{ij} is the distance between the two municipalities. θ is a proportionality constant, $\tau_d = 0.28$ and $\tau_r = 0.66$ tune the dependence of dispersal on donor and recipient sizes and $\rho = 2.95$ tunes the dependence on the distance. Model parameters were optimized as in [29] in order to take into account that the fraction of commuters (individuals traveling outside the municipality of residence for work or study) in Italy varies significantly from South to North of Italy, ranging from 15% in Southern Italy to 60% in Northern Italy. Fig. 4.1d shows the resulting probability density function of travel distances, compared with that obtained by using a model depending only on the distance, namely Eq. (4.3), used for modeling the transmission in the general population.

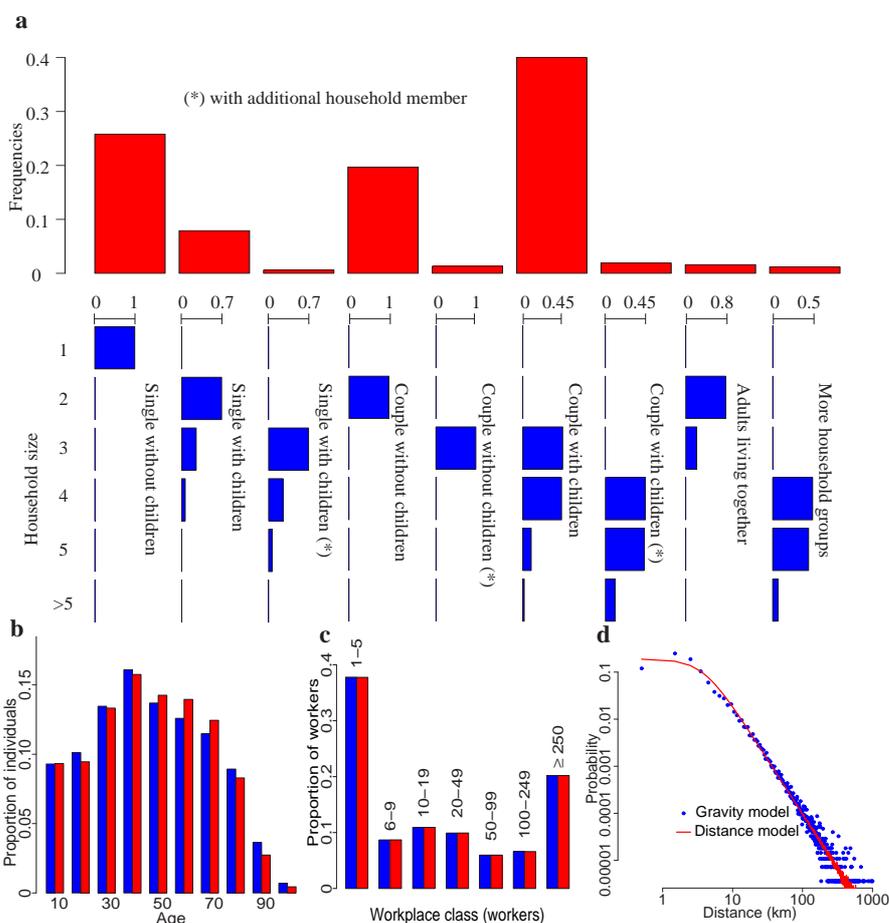


Figure 4.1: **a** Frequency distributions of household size for the different household types (in blue) and frequency distribution of the different household types (in red) considered in the model. **b** Age distribution from census data (blue) and simulated (red). **c** Proportion of workers for class of workplace from industry census data (blue) and simulated (red). **d** Probability density function of travel distances as obtained by using the gravity model (4.1) (in blue) compared with that obtained by using the the distance kernel (4.3) (in red).

4.1.2 Transmission models

As in [47, 54], the seeding of the infection is based on the arrival of infected individuals from abroad. Therefore, the model accounts for two of transmission processes: among the simulated individuals and by the importation of cases from abroad.

Worldwide transmission model

The worldwide spread of influenza pandemic and the consequent importation of cases in Italy were modeled using a deterministic homogeneous-mixing SEIR model. This model was used for determining the number of imported cases in Italy from abroad over time.

Specifically, we coupled the results of the worldwide model with 2003 data on arrivals and departures in Italy's 38 international airports. More in detail, the number of imported cases over time was estimated by sampling a Poisson distribution of parameter $p \frac{E(t)}{N} \Delta t$, where p is the total number of passengers arriving daily in Italy ($\approx 70,000$ on average), $E(t)$ is the number of exposed individuals at time t predicted by the global homogeneous-mixing model, N is the world population and Δt is the time step of the simulation.

National transmission model

The national impact of the epidemic in Italy was predicted using a stochastic, spatially-explicit individual-based model [47, 29]. For each individual i we define:

- H_i as the set of the n_i individuals belonging to the same household of individual i ;
- L_i^j as the set of the m_i^j individuals attending the same school (index $j = 1, \dots, 6$ identifies school types, from day care centers to university) or sharing the same workplace (index $j = 7, \dots, 13$ identifies workplaces of increasing size, see Fig. 4.1c) of individual i ;

Any susceptible individual i , at any time t of the simulation has a probability $p_i = 1 - e^{-\lambda_i \Delta t}$ of becoming infected, where $\Delta t = 0.25$ days is the time-step of the simulation and λ_i is the instantaneous risk of infection. The latter is the sum of the risks coming from the three modeled source of infections, namely contacts with infectious members of the household, contacts with infectious individuals working in the same workplace or attending the same school, random contacts in the population:

$$\begin{aligned}
 \lambda_i &= \sum_{k|H_k=H_i} \frac{I_k \beta_h \kappa(t - \tau_k) [1 + C_k(\omega - 1)]}{n_i^\alpha} \\
 &+ \sum_{j,k|L_k^j=L_i^j} \frac{I_k \beta_p^j \kappa(t - \tau_k) [1 + C_k(\omega \psi_p^j(t - \tau_k) - 1)]}{m_i^j} \\
 &+ \sum_{k=1}^N \frac{I_k \beta_r \kappa(t - \tau_k) f(d_{ik}) [1 + C_k(\omega - 1)]}{\sum_{k=1}^N f(d_{ik})}
 \end{aligned} \tag{4.2}$$

The terms in Eq. (4.2) are defined as follows:

- N is the total population, i.e. $\approx 57,000,000$ of individuals;
- $I_k = 1$ if individual k is infected, 0 otherwise;
- β_h is the within-household transmission coefficient, β_p^j are the within-school/workplace transmission coefficients and β_r is the transmission coefficient for random contacts. The different transmission scenarios were drawn by varying the transmission parameters.
- τ_i is the time in which individual i became infectious and $\kappa(T)$ is a lognormal function describing infectiousness over time. Estimates of incubation period (1.48 ± 0.47 days) and infectiousness period ($\int_0^\infty T \kappa(T) dT$) lead to a generation time $T_g = 2.6$ days (as in [46]);
- $C_k = 1$ for symptomatic cases (we assume 50% of cases to be symptomatic), 0 otherwise. Since $\omega = 2$, the infectiousness of symptomatic cases doubles the one of asymptomatic cases (as in [46]);
- $\alpha = 0.8$ scales the household transmission rates with household size (as in [46]);
- $\psi_p^j(T)$ is a function accounting for induced absenteeism and it is defined as follows: if $T > 0.25$ (the minimum time for recognizing the infection) $\psi_p^j(T)$ is set to: 0.1 for $j = 1, 2$; 0.2 for $j = 3, 4$; 0.25 for $j = 5$; 0.5 for $j = 6, \dots, 13$; 1 otherwise;
- as in [47, 46, 29], we assume that random contacts in the population depend explicitly on the distance d_{ik} between infectious individual k and susceptible individual i . The probability that an infectious individual k infects individual i is weighted by the kernel function

$$f(d_{ik}) = \frac{1}{1 + (d_{ik}/a)^\rho} . \quad (4.3)$$

with $a = 3.8km$ and $\rho = 2.32$ [29].

We assume that 33% of transmission occurs in households, 33% in schools or workplaces and 33% in the general community [46, 29].

4.1.3 Epidemiological parameters

In the worldwide model, we assumed that infectious individuals were all symptomatic and no longer traveling and that exposed individuals were asymptomatic and possibly traveling before the infectious phase. In the national model, infectious individuals were divided

into symptomatic and asymptomatic classes. Once an individual become infectious, the probability of developing symptoms was set to 0.5. In both models, we assumed that the latency period for influenza was the same as the incubation period: duration of 1.5 (\pm 0.5 standard deviation) days. In the national model, we assumed that the duration of infectiousness varied over time, as a lognormal function [47, 46, 29]. Infectiousness peaked at 1.75 days, and its duration was truncated at 10 days. This corresponded to an average generation time of 2.6 days. In the worldwide model, the infectious period was assumed to be constant over time and was set at 1.5 days, to give essentially the same growth rate as the national model [47, 29].

4.1.4 Excess mortality

Though it is not possible to predict death rates in future pandemics (reliable estimates are not available yet for the ongoing A(H1N1) influenza outbreak), it is important to assess the effects of antiviral treatment and prophylaxis under different assumptions on age-specific case fatality rates. We used results presented in [7] on the lethal 1918-19 influenza pandemic in Copenhagen (scenario *EM1918*), where deaths occurred primarily among young persons, and in [117] on the mild 1969-70 influenza pandemic in Italy (scenario *EM1969*), where deaths occurred primarily among elderly (as during inter-pandemic seasons), to estimate age-specific case fatality rates. Basically, we assumed that the estimated age-specific excess mortality rates as reported in [7] were associated to an epidemic with $R_0 = 2$ (authors report estimates of R_0 in 2.2-2.4 for the Summer wave and $R_0 \approx 1.2$ for the Fall wave, due to preexisting immunity in the population) and we estimated age-specific case fatality rates (for symptomatic individuals) in such a way that the age-specific excess mortality rates as obtained by running simulations with $R_0 = 2$ comply with the values reported in [7]. The resulting age-specific case fatality rates were used to estimate age-specific excess mortality in all the considered transmission scenarios. Similarly for the data on the 1969-70 influenza pandemic in Italy, where we assumed $R_0 = 1.4$ (estimated value in the range 1.3-1.6 [63]).

4.2 The European model

We developed an individual-based epidemic simulation model [95] that accounts, at country level, for explicit transmission in households, schools, workplaces (where homogeneous mixing is assumed) and in the general population (where the force of infection is assumed to depend explicitly on the distance). The epidemic can spread from one country to another through cross-borders diffusion and because of long distance travels. The infection is continuously sustained in the study area by importation of cases from countries outside Europe. Sociodemographic data were used to generate a highly detailed synthetic population of individuals, explicitly grouped in households, schools and workplaces, for simulating the populations in the different countries of the study area. Data on air and railway transportation data were used to simulate long-distance travels across the countries of the study area and to simulate importation of cases.

4.2.1 Sociodemographic structure of the European population

Population density

The Gridded Population of the World version 3 (GPW v3) [14] produced by the Center for International Earth Science Information Network (CIESIN) of the Earth Institute at Columbia University was used as the source of population density for the study area.

The study area covers most part of Europe and includes all the member states of the European Union, for a total of 37 countries. The total population of the study area is about 515 million individuals. The most and the less populous countries are Germany (≈ 81 millions) and the Principality of Monaco (≈ 14 thousand), respectively.

The study area covers a surface of about 5 million square kilometers and it is divided into 63,794 cells, whose average surface is about 77 square kilometers. The number of individuals per cell varies from 1 (in many parts of Scandinavia and Scotland) to 1,355,987 (a cell of Paris), with an average of 8,066 individuals. Such thorough grid allows a precise spatial location of households, schools and workplaces, crucial for modeling the spatial spread of the epidemic.

Households

The importance of considering realistic household groups in spatial studies of human diseases, such as influenza, is well known [114]. The explicit representation of the household groups in the model allows testing the effectiveness of intervention options such as antiviral prophylaxis, which is considered one of the key measures for containing/mitigating a

new influenza pandemic [129]. Moreover, since households are characterized by a static geographic location, it is possible to evaluate the effectiveness of spatial interventions (i.e., the administration of antivirals to the individuals living within a certain distance from a symptomatic case), which can be crucial for containing an influenza pandemic at the source [87, 46]. Therefore, it arises the need of developing a realistic model of household groups in the different countries of the study area. We used an heuristic model which matches marginal distributions of household size and population age structure, and maintains realistic generational age gaps within households (by avoiding randomly assigned ages to the households members), respecting as best as possible the actual mix of students, workers and inactive individuals. A sketch of the heuristic model employed is shown below:

1. determine the household type by sampling from the distribution of the frequencies of household type (see Fig. 1b in the main text);
2. assign an age to the household head, a_h , by sampling from the distribution of the age classes (see Fig. 1e in the main text) and by taking into account the following constraints:
 - (a) $a_h \geq 18$,
 - (b) $a_h \leq 65$ if there are children among the household members;
3. determine the number of additional members of the households by sampling from the proper, type dependent, distribution of the frequencies of household size;
4. assign an age to the additional members by sampling from the distribution of the age classes and by taking into account the following constraints:
 - (a) the age of (eventual) wife/husband, a_w , satisfies $a_h - 15 \leq a_w \leq a_h$ and $a_w \geq 18$;
 - (b) the age of (eventual) children, a_c , satisfies: $a_M - 40 \leq a_c \leq a_m - 15$, where $a_M = \max \{a_h, a_w\}$ and $a_m = \min \{a_h, a_w\}$, if there are two adults living in the household; $a_h - 40 \leq a_c \leq a_h - 15$, otherwise.

Schools and workplaces

As regards the schools size, not all the countries in the study area are covered by the PIRLS 2001 and PISA 2000 and 2003 international surveys. In the countries not covered by the surveys we used average values. The values reported in the surveys on the size of the Italian schools were validated by comparing them with detailed data on the location

of all Italian schools. Specifically, we used average values of the schools size, as resulting from the surveys, for generating a synthetic set of primary and secondary Italian schools. We found that the resulting number of secondary schools complies with the actual number of Italian secondary schools. On the contrary, the resulting number of primary schools is largely lower than the the actual number of Italian primary schools. Very likely, this is due to different definitions of school. The Italian data refer to scholastic buildings while the survey data refer to administrative units, which can comprise several scholastic buildings. Consequently, the actual size of the scholastic buildings (the structures relevant to the epidemic transmission) can be largely lower than that reported in the surveys, especially in rural areas. This considered, we allowed the minimal size of the schools simulated in the model to be much lower than the minimal size resulting from the analysis of the survey data.

Schools and workplaces were spatially-distributed proportionally to the population. This assumption was supported by the analysis of the spatial location of the Italian schools.

Data on the workplaces size for all the countries in the study area were not available. We analyzed the distribution of the workplaces size in Italy and the United Kingdom and we did not find significant differences. We used these data to generate a distribution of workplaces size that we used for assigning the size to the workplaces in all the countries of the study area.

4.2.2 Mobility of the European population

Long-distance travels within Europe

The data on the mobility of the EU27 population reveals a relevant flux of individuals traveling among the EU27 countries (more than 360 million travelers per year). Such volume of human movement has to be considered in the epidemic transmission model. Therefore, since the available data refer only to EU27 members, the need arises to develop a model of “long-distance” travels among all the considered countries.

We used three different methods to fit the origin-destination matrix, i.e. to estimate the flows from country i to country j . Specifically, we generated a synthetic population of travelers taking travels according to a gravity model whose masses are given by the normalized GDPs (model A, full description of the model can be found in Sec. 5.4.2), by the population sizes, as in [126] (model B, $F_{ij} = \theta \frac{p_j^{\tau_i} p_i^{\tau_j}}{d_{ij}^p}$, where p_i is the population of country i) and, finally, taking travels by choosing random destinations (model C). We performed a statistical analysis which revealed that the best model for reconstructing the

data is model A.

Internal commuting

As detailed in the main text, students and workers were randomly assigned to a school or workplace, in such a way that probability density function of travel distances complies with a truncated power-law distribution, as proposed in [58] for the radius of gyration of mobile phone users, namely

$$P(r_g) = (r_g + r_g^0)^{-\beta_r} \exp\left\{-\frac{r_g}{\kappa}\right\}, \quad (4.4)$$

where $r_g^0 = 5.8\text{km}$, $\beta_r = 1.65$ and $\kappa = 350\text{km}$.

Here we compare the probability density functions of travel distances as obtained by employing Eq. (4.4) and by employing a gravity model previously developed by the authors for Italy [29] and detailed below.

Commuting destination are assigned in order to fit available commuting data [110]. In particular, the proportion of individuals with age ≥ 15 working or attending school in the same municipality of residence is available for each municipality, together with the number of individuals traveling either to a municipality of the same province they live in, or outside the province but within the same region, or outside the region.

As a starting point, for determining the probability of commuting from municipality i to municipality j , we employed a gravity model of the form [126]

$$C_{ij} = \theta \frac{n_i^{\tau_f} n_j^{\tau_t}}{d_{ij}^\rho}, \quad (4.5)$$

where n_i and n_j are the number of individuals living in municipality i and j respectively and d_{ij} is the distance between the two municipalities. θ is a proportionality constant, τ_f and τ_t tune the dependence of dispersal on donor and recipient sizes and ρ tunes the dependence on the distance.

The proportion of commuters (individuals traveling outside the municipality of residence for work or school) in Italy, however, varies significantly by province. In particular, the proportion of commuters drastically increases from South to North of Italy. The proportion of commuters varies from 15% to 60% and this variability does not depend on the size of the municipalities or on the distance among municipalities. Indeed, it depends on social factors and thus can not be explained by model (4.5).

Thus, we considered model (4.5) with an additional constraint forcing the model to produce in each province the proportion of commuters as resulting from the available data.

The set of model parameters was optimized by searching for the set giving rise to a simulated population of commuters matching the available data on the number of individuals commuting within the province, the region or outside the region of residence. We obtained the following estimates: $\tau_f = 0.28$, $\tau_t = 0.66$ and $\rho = 2.95$ which are very close to those obtained in [126] for modeling travel destinations in the United States at short distances (less than 119 km). We found that the probability density function of travel distances as obtained by employing the gravity model can be approximated by a power-law distribution

$$C_{ij} = \frac{1}{1 + (d_{ij}/a)^b}, \quad (4.6)$$

with $a = 3.8km$ and $b = 2.32$.

We used the same law (Eq. 4.4) for modeling the internal commuting in all the countries of the study area. This is a strong assumption. However, it was originally proposed in [58] as an universal law of human mobility (at least in developed countries, as the European countries). Moreover, our assumption is strongly supported by the good agreement between the results obtained by employing this law and a gravity model developed on specific Italian data of internal commuting.

Travels from/to outside Europe

The number of intercontinental passengers traveling from and/or to EU27 in 2007 were more than 135 millions. Since it is not realistic to assume that a new influenza pandemic will spread only within Europe, the need arises of modeling the worldwide spread of the epidemic. A model based on the number of individual traveling from/to each pair of international airports of the world as in [31] represents the optimal solution. Once the location of the first world case is correctly identified, it is possible to make predictions on timing, volume and location of the cases over time. However, since the aim of this study is to predict the spatiotemporal spread of an epidemic in Europe, it is sufficient to estimate the number of imported cases over time and to identify the more likely final destinations of infected individuals. For accomplishing this task, a much simpler “global” model based on an homogeneous mixing assumption in the world population is sufficient.

Therefore, we used a discrete-time stochastic SEIR epidemic model for describing the dynamics of the worldwide epidemic. In order to estimate the number of imported cases by international flights at each time step Δt of the simulation in any European country c , we sampled from a Poisson distribution of parameter

$$a_c \Delta t [E(t) + (1 - P_s)I(t)] / \tilde{N} \quad (4.7)$$

where $E(t)$ is the number of exposed individuals (predicted by the homogeneous mixing SEIR model at time t); $I(t)$ is the number of infectious individuals (predicted by the homogeneous mixing SEIR model at time t); P_s is the probability of developing symptoms, \tilde{N} is the world population, a_c is the total number of passengers arriving daily in the country c . Finally, imported cases were randomly assigned to a cell of the country (by replacing susceptible individuals) with probability proportional to the population. Let us note that, by employing Eq. (4.7), we are assuming that only exposed and asymptomatic individuals have a chance of arriving in Europe. In fact, we are assuming that symptomatic individuals can not travel because of the influenza symptoms or that they can be recognized as ill and consequently isolated.

The “global” model is parametrized in the same way as the European individual-based model. The probability of developing symptoms is kept fixed to 0.5. Latent and infectiousness periods have the same duration as in the European model. The transmission rate is chosen to give the same reproductive number as for the European model.

4.2.3 Epidemic transmission details

Natural history of influenza

The epidemiological flow of influenza can be schematized as follow. A newly infected person passes through latent and infectious phases (the latter can be characterized by presence or absence of clinical symptoms). After the infectious phase, either individuals recover and acquire (partial or complete) immunity, or die.

As in [87, 46], we assume a fully susceptible population, even if residual immunity derived from other influenza strains was suspected for last pandemics [107]. Moreover, induced mortality is not considered [87, 29, 47, 54]. We consider a scenario accounting for the transmission of a single new influenza strain: the presence of multiple strains [27] or the contemporary presence of other diseases potentially affecting the influenza dynamics [97] are not considered. Spontaneous changes in the population behavior in response to a pandemic and potentially affecting its course [45, 113] are not considered.

The mechanism responsible for the seasonality of influenza epidemics are becoming clearer [84]. In particular a study has highlighted how the vapor pressure is probably the driving factor for this process [119]. However, we do not consider seasonality effects since the whole study area is located in the northern hemisphere. Moreover, it is not clear yet how to model the dependence of the transmission rates on seasonal factors (e.g., the above mentioned vapor pressure). Therefore, in each simulated epidemic the transmission rates are kept constant, though it is possible that the school closure in the Summer period

would delay the spread of the epidemic in Europe, independently on seasonal factors.

According to [46, 47, 29], the probability of developing symptoms is assumed to be 0.5. Different scenarios were drawn by assuming that the probability of developing clinical symptoms ranges from 0.35 to 0.8. Symptomatic individuals are assumed to transmit the infection more than asymptomatic individuals (by a factor 2) [29, 47, 85]. As in [29, 47], we assume that the duration of latent and infectious periods is the same in both symptomatic and asymptomatic individuals. The latent period is assumed to be exponentially distributed with mean 1.5 days. We assume an exponential distribution also for the infectious period (mean 2 days). However, we also performed a sensibility analysis on the duration of the infectious period. Specifically, different scenarios were drawn by assuming that the infectious period ranges from 1.5 to 3 days.

Transmission model

The influenza transmission was simulated by a stochastic spatially structured individual-based model. We considered an epidemic occurring within the European population, accounting for ≈ 515 million individuals, kept constant during a simulation (i.e., without considering neither born/dead nor immigration/emigration processes). Individuals are explicitly represented in the model and are characterized by citizenship, household membership and school/workplace membership (if any). Households, schools and workplaces are located in an explicit geographic location.

Once the population is initialized, at any time t of the simulation (time step $\Delta t = 0.5$ days), any susceptible individual i has a probability $p_i = 1 - e^{-\Delta t \cdot \lambda_i(t)}$ of becoming infected. The probability of becoming infected depends on the instantaneous risk of infection $\lambda_i(t)$, computed at any time of the simulation. The risk of infection for each individual is defined as the sum of the risk factors coming from the different sources of infections considered, namely:

1. contacts with infectious members of the household (first term in Eq. 4.8),
2. contacts with infectious individuals working in the same workplace or attending the same school (second term in Eq. 4.8),
3. random contacts in the population and cross-border diffusion (third term in Eq. 4.8),
4. long distance travels (fourth term in Eq. 4.8).

$$\begin{aligned}
\lambda_i = & \sum_{\{k=1, \dots, N_{S_i} | H_k = H_i\}} \frac{I_k C_k \beta_h}{n_i} \\
& + \sum_{\{k=1, \dots, N_{S_i} | P_k^j = P_i^j\}} \frac{I_k C_k a_k^j \beta_p^j}{m_i} \\
& + \sum_{\{k=1, \dots, N | S_k \in S_i^*\}} \frac{I_k C_k \beta_u f(d_{ik})}{\sum_{\{k=1, \dots, N | S_k \in S_i^*\}} f(d_{ik})} \\
& + \sum_{\{k=1, \dots, N | S_k \neq S_i\}} \frac{I_k C_k \beta_t P(S_i, S_k)}{N_{S_k}} \tag{4.8}
\end{aligned}$$

The terms in Eq. (4.8) are defined as follows:

- S_i is the index of the country where individual i lives in.
- H_i is the index of the household where individual i lives in.
- P_i^j is the index of the place where individual i works/studies (if i is employed) and j identifies the place type (e.g., school, workplace).
- N is the European population.
- N_{S_k} is the population of country S_k .
- S_i^* is the set indices of the countries bordering country S_i , country S_i included. This set of countries (S_i^*) allows taking into account random contacts in the general community in country S_i and cross-borders diffusion due to random contacts among individuals living in bordering countries ($S_i^* \setminus S_i$).
- $I_k = 1$ if individual k is infected, 0 otherwise.
- $c_k = 2$ for symptomatic cases (we assume the 50% of cases to be symptomatic), 1 otherwise. As discussed before, this choice is consistent with the one adopted in [46, 29, 47].
- a_k^j accounts for sickness-induced absenteeism and it is defined as follows: 0.2 if individual k is a symptomatic student, 0.5 if individual k is a symptomatic worker, 1 otherwise. This is a simplified version of the parameters used in [46, 47, 29] for modeling induced absenteeism. According to the results of a survey presented in [108], for every 100 children followed up during an influenza season, which included 37 school days, an excess 28 illness episodes and 63 missed school days occurred

(average number of schooldays missed: 2.25). According to the results of a survey presented in [4], the average number of workdays missed due to ILI is 1.30 days. These results support the choice of the parameters values.

- $f(d_{ik})$ is the function in Eq. (4.4). It makes the transmission of the epidemic in the general community (and the cross-border diffusion) explicitly dependent on patterns of human mobility (as described in [58]).
- $P(S_i, S_k)$ is the probability of traveling from country S_i to country S_k . It is estimated by using the gravity model A.
- β_h (expressed in day^{-1}) is the within-household transmission rate.
- β_p^j (in day^{-1}) is the within-school/workplace transmission rate, which depends on the type j of the place. Specifically, as in [29, 47], we assume that β_p^j is the same for all the school types and that it is three-times the transmission rate in the workplaces. Sensitivity analysis is shown in Sec. 4.2.2.
- β_u (in day^{-1}) is the transmission rate in the general community.
- β_t (in day^{-1}) is the transmission rate for long-distance travels.

At any time t of the simulation, infected (exposed) individuals enter the infectious phase with probability $\Delta t/T_l$, where T_l is the latent period. When exposed individuals become infectious they develop symptoms with probability 0.5. Finally, infectious individual recover from the infection with probability $\Delta t/T_i$, where T_i is the infectious period.

4.2.4 Model parametrization

The basic reproductive number R_0 is the fundamental parameter in epidemic models. It is defined, essentially, as the average number of secondary infections caused by a typical primary infection in an infinite and completely susceptible population [6]. It depends not only on the transmissibility characteristic of the virus, but also it greatly depends on the structure of the host population. Therefore, it is not possible to set an unique R_0 for the whole study area without changing the values of the transmission rates in the different countries. Since there are neither evidence nor reliable reasons for assuming differences among the European countries in the probability of transmitting the infection in the different social contexts, we fixed the transmission rates to obtain a certain R_0 in the United Kingdom and then we employed the same transmission rates in all the countries of the study area.

As regards the value for the reproductive number, according to recent estimations based on past influenza pandemics [46, 27, 99], we investigated three plausible transmissibility scenarios: namely $R_0 = 1.6, 2$ and 2.4 .

We assumed the same proportion of transmission in the different social contexts as in [47]. As stated in the main text, the model was parametrized in such a way that in the United Kingdom 30% of transmission occurs in households, 37% in schools and workplaces and 33% in the general community. This choice, coupled with the choice of fixing the basic reproductive number in the United Kingdom, allowed us to compare the results of this study with those obtained in [47], which is focused on United Kingdom and United States. Reliable data on the proportion of transmission in social contexts, crucial to the disease transmission, as prisons, leisure places, public transportation systems, hospitals are not available, though some research works are contributing to fill the gap [105, 133]. Our assumption is supported by the findings presented in [105], where the authors analyzed social contacts and mixing patterns in eight European countries. Specifically, they found that living in a larger household size was associated with higher number of reported contacts. Moreover, they found that the dominant feature of the contact matrix data is the strong diagonal element: individuals in all age groups tend to mix assortatively (i.e., preferentially with others of similar age) and this pattern is most pronounced in those aged 5-24 years, i.e. the scholar age. They also found that 58% of all reported contacts occur at home, at work, or at school. This results supports our assumption on the proportion of transmission in the different social contexts (in the model, 67% of transmission occurs in households, schools and workplaces, at least in the United Kingdom).

Finally, we chose parameter β_t on the basis of the following criteria: since the number of passengers traveling among European countries is about 2.7 times the number of passengers from outside Europe, we assumed the number of cases generated during long-distance travels in Europe to be about 2.7 times the number of imported cases from outside Europe.

4.3 Modeling endemic diseases

We have developed an individual-based model with dynamic network of contacts, parameterized by employing sociodemographic and epidemiological data and accounting for millions of individuals. The main ingredients of the model are: *(i)* the sociodemographic model, simulating the temporal evolution of the population; *(ii)* the transmission model, simulating the temporal evolution of the epidemic.

4.3.1 Sociodemographic model

Initialization of the network of contacts The contact network among individuals is adapted from the one introduced in [2]. Basically, each individual is explicitly represented in the model and the network of contacts is generated by co-locating individuals in households, schools and workplaces.

The populations of Campania (5,701,931 inhabitants) and Puglia (4,020,707 inhabitants) are modeled. Individuals are hierarchically grouped within the region of residence by municipalities and provinces, according to the administrative borders of the study area and to the number of individuals by municipality.

Census data on population, household type and size [110] and data on age structure [75] are jointly used with survey data [74] to randomly assign age and co-locate individuals in households. Data on the age structure are specific for the two regions, while survey data refer to an analysis conducted at national level. Nine different household types are considered in the model. Comparison between real and simulated age structure and between real and simulated household size frequencies are reported in Fig. 4.2.

Demographic, school [101, 102] and industry census data [73] are used for generating schools and workplaces, and for randomly assigning an employment category (student, worker, unemployed) to each individual on the basis of age and employment rates by age (which also include rates of school attendance). Six school types, from day care to university, are considered in the model. School employees (e.g. teachers) are located in schools and not in generic workplaces.

Commuting destinations for both workers and students are randomly assigned by employing a gravity model [126], integrated with specific data [110].

All the details on the initialization of the sociodemographic model can be found in [3].

Dynamic network of contacts The vital dynamics of the population has to be considered for modeling endemic diseases. This implies that the network of contacts among individuals has to be kept updated. In particular, in the proposed model, individuals

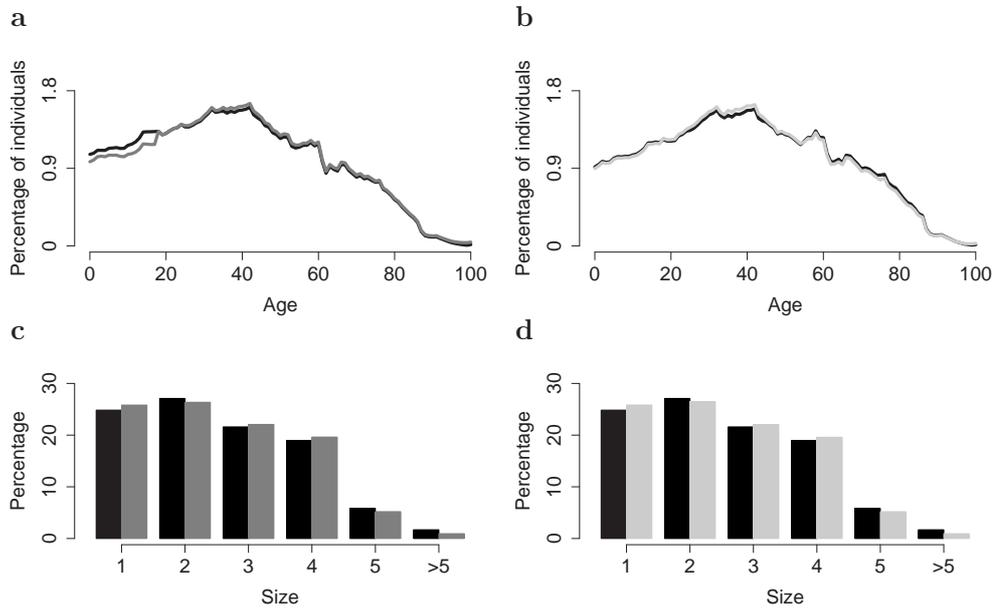


Figure 4.2: **a** Age structure (year 2007) of Campania region (black line) and age structure simulated by the sociodemographic model (gray line). **c** Italian household size (black bars) and household size simulated by the model in Campania (gray bars). Note that data on household size are not directly employed in the algorithm. **b** and **d** as **a** and **c**, but for Puglia.

come to life, grow, can generate new households, can procreate and die; moreover, they can attend school (following the educational career), have an employment and retire.

The population is assumed to be closed, i.e. without immigration and emigration, and with a constant number of individuals, i.e. the number of deaths corresponds to the number of newborns. Despite Italy has one of the lowest fertility rates worldwide, it has experienced a large immigration since the 1990s [69], which has contributed to slightly increase the Italian population. Campania and Puglia followed the completely opposite pattern. The demographic balance is positive, with an excess of births over death, but the two regions have experienced a large emigration. As a consequence, in the last 5 years Campania and Puglia populations have been constant in size: the variations (both positive and negative) have been less than 0.7% per year. Therefore, assuming a constant number of individuals can be considered as a reliable modeling choice.

Births replace dead individuals, which are randomly selected on the basis of the mortality rates by age, specific for the two considered regions [71]. Each newborn is located in an existing household, chosen among those of suitable size and members age. Finally,

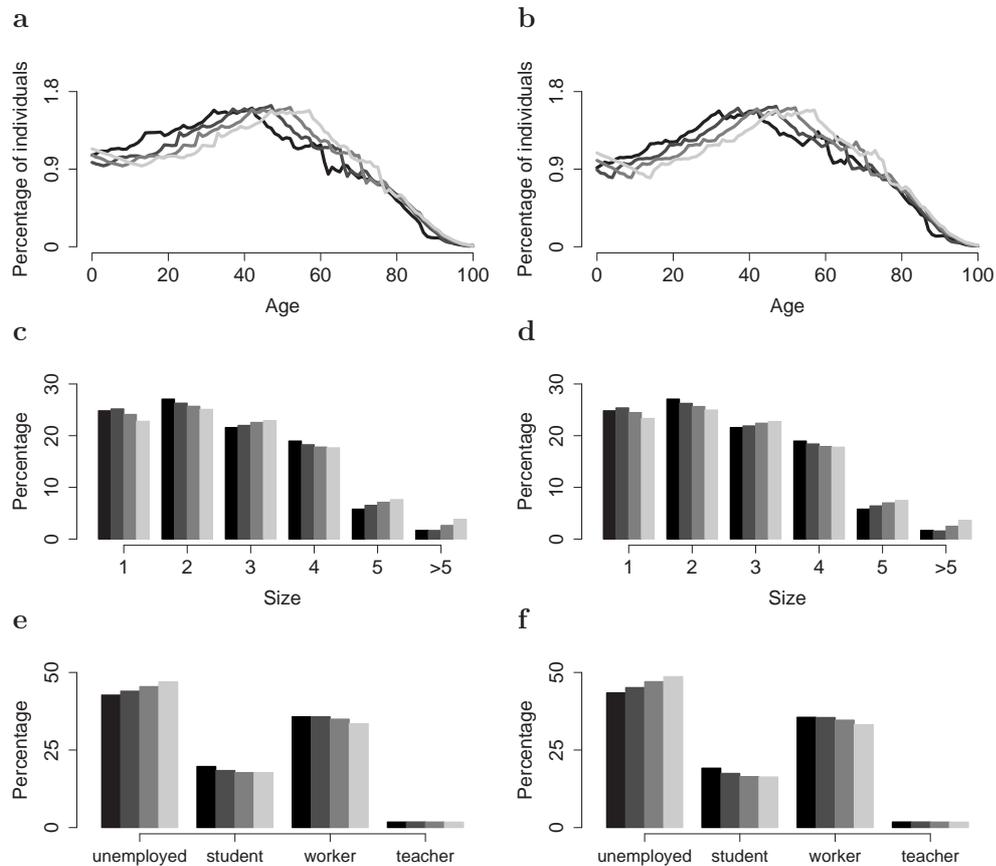


Figure 4.3: **a** Age structure of Campania (black line) and age structures simulated by the sociodemographic model: projections to the years 2012, 2017 and 2022 (from dark gray to light gray). **c** As **a** but for frequencies of household size. **e** As **a** but for frequencies of employment categories. **b**, **d**, **f** as **a**, **c**, **e** but for Puglia region.

it is determined whether the newborn is enrolled in a day care center or not.

Individuals who have not already generated their own family and are at least 18 years old can generate a new household group, possibly a single member household, on the basis of the probability of getting married by age [76].

Every time the population is updated the age of each individuals is increased and an employment category (possibly, “unemployed”) is associated to each individual on the basis of the employment rates by age. If this category is the same that she/he had before the update she/he keeps her/his place, otherwise a new school/workplace is randomly chosen among the existing ones. In the latter case, the commuting destination is determined by employing the same gravity model employed for the initialization of the network of contacts. As regards the students, the school type (day care center,

kindergarten, primary, middle, high school, university) is deterministically assigned on the basis of the age of the individual. Since the age of the individuals is kept updated, the previously described procedure allows students to follow the educational career.

Fig. 4.3 shows the sociodemographic projections for the years 2012, 2017 and 2020 in terms of age structure, frequencies of households size and employment category. These projections show a progressive population aging (Fig. 4.3a and 4.3b) and the consequent decrease in the number of students and workers and the increase in the number of retired individuals (Fig. 4.3e and 4.3f). These demographic projections comply with the Italian situation. In fact, nowadays, Italian population is undergoing a phase of progressive aging. For validating these demographic projections, we initialized the population of both Campania and Puglia with their age-structure in the 2002 and compared them with the age-structure simulated by the model using these data as input. Then we ran the simulations and compared the age-structure in the 2007 with the one predicted by the model. The predicted age-structures for both Puglia and Campania are in good agreement with the real data.

4.3.2 Epidemic transmission model

The epidemic transmission model is adapted from the one proposed in [1], which is specific for describing the hepatitis A transmission in Italy. The model accounts for the three main sources of HAV infection in the region: person-to-person transmission, ingestion of infected food and travels to high endemicity areas.

For each susceptible individual i , the probability of being infected at time τ is given by

$$p_i(\tau) = 1 - e^{-\lambda_i(\tau)\Delta t},$$

where $\lambda_i(\tau)$ is the force of infection for the susceptible individual i at time τ , and Δt is the time step of the simulation (1 week). $\lambda_i(\tau)$ is the sum of the risk factors due to the three considered sources of infection: $\lambda_i(\tau) = \lambda_i^D(\tau) + \lambda_i^S(\tau) + \lambda_i^T(\tau)$, where $\lambda_i^D(\tau)$ represents the direct transmission component, $\lambda_i^S(\tau)$ represents the indirect transmission component and $\lambda_i^T(\tau)$ the component associated to travels to high endemicity areas.

Direct transmission Direct transmission accounts for person-to-person contacts. Since individual-based models allow the explicit representation of the places where transmission can occur, this transmission component is divided into contacts within households, schools and workplaces.

In particular, for each susceptible individual i the risk factor $\lambda_i^D(\tau)$ at time τ is given by:

$$\lambda_i^D(\tau) = \sum_{j \in H_i} \frac{I_j(\tau)\beta_h}{h_i - 1} + \sum_{j \in P_i} \frac{I_j(\tau)\beta_p^{\vartheta(i)}}{p_i - 1}, \quad (4.9)$$

where

- $I_j(\tau)$ is equal to 1 if individual j is infectious at time τ , 0 otherwise.
- H_i is the set of the indices of the h_i individuals living in the same household of i (note that if $h_i = 1$ the first term in Equation 4.9 is 0 by definition).
- β_h is the transmission rate within households.
- P_i is the set of the indices of the p_i individuals working or studying in the same place of i (if $p_i = 0$ or $p_i = 1$ the second term in Equation 4.9 is 0 by definition).
- $\vartheta(i)$ is the employment category of individual i (i.e., one of the seven types described in the sociodemographic model).
- $\beta_p^{\vartheta(i)}$ is the transmission rate within the place of type $\vartheta(i)$.

Indirect transmission Indirect transmission accounts for the infections by ingestion of contaminated seafood. The risk of infection due to this component at time τ is given by:

$$\lambda_i^S(\tau) = \beta_s(a_i)U(\tau),$$

where

- $\beta_s(a_i)$ is the transmission rate associated to the ingestion of infected seafood. Since consumption of raw mussels and shellfish varies by age, β_s is a function of the age of the individual i .
- $U(\tau)$ is the quantity of HAV in seafood at time τ , as a consequence of the excretions of infected individuals.

The variable $U(\tau)$ depends on the number of individuals that were infective prior to time τ , basically because they excrete HAV in the environment during their entire period of

infectivity. For simplicity, we decided to model the dynamics of $U(\tau)$ by the following ordinary differential equation:

$$\frac{d}{d\tau}U(\tau) = \delta \left[\sum_{j=1}^N I_j(\tau) - U(\tau) \right], \quad (4.10)$$

where δ is the decay rate of HAV in the environment and $\sum_{j=1}^N I_j(\tau)$ is the number of infectious individuals in the population, which comprises N individuals, at time τ . A discussion on this model of indirect transmission can be found in [1].

Travels to high endemicity areas Travels to high endemicity areas account for all the infections occurring outside the considered region. This component of the force of infection is modeled as

$$\lambda_i^T(\tau) = \beta_t(a_i),$$

which is constant over time and varies by age.

Chapter 5

Experimental results

5.1 Understanding the impact of different modeling assumptions

The modeling framework adopted for the comparison, particularly for the structured component, is represented by the recently developed IBM used for pandemic prediction and control in Italy (and described in Sec. 4.1). Three main approaches are considered to model the unstructured component: a spatially explicit model depending on a parametric kernel function of the distance, with asymptotic power-law form [46, 47, 29, 96]; a model where random contacts are chosen in the local communities [85]; a model where random contacts are defined on the basis of commuting data [98]. For ease we term the three models as models S, L and M respectively. Moreover, we also included occasional long-distance trips T (as in [54]) in models L and M, called now L+T and M+T respectively.

5.1.1 Definition of the models of the unstructured contacts

Here we define unstructured any contact which is not a household or workplace contact and we consider the following five different models of transmission by unstructured contacts.

- Model S: unstructured contacts through the whole space by a distance-based model. Each individual is in contacts with every other individual in the population, with probability (decreasing with the distance) given by a specific kernel function.
- Model L: unstructured contacts within the municipality the individual lives in.
- Model M: unstructured contacts within the "commuting community" the individual belongs to. In particular, for individuals who study or work in the same municipality

they live in, the social network consists of other inhabitants of the same municipality and those who commute to this municipality. For individuals traveling outside the municipality of residence, the social network consists of the inhabitants and commuters of both depart and arrival municipality.

Moreover, we consider two additional models including occasional long-distance trips [54] in models L and M, called L+T and M+T respectively. In particular, all individuals are assumed to spend in average 10 days (randomly chosen) per year in a community other than that of residence and school/work. In these periods, within household, school and workplace transmission is not allowed.

5.1.2 How to compare different models

In order to perform the comparison, we define the *first generation index* G_0 , as the average number of secondary infections generated by the first infected individual, during his entire infectious period, in a completely susceptible population.

In traditional models the simplest choice would be to fix the basic reproduction number R_0 (see [6, 37]), which can be estimated by approximating the slope of the cumulative number of cases during the exponential growth phase of the epidemic. The difference between first generation index and basic reproduction number lies on the fact that the former is determined only by the first generation of infection while the latter emerges after the underlying next generation operator is applied for a sufficiently large number of generations. Our choice is motivated by the simplicity of the G_0 computation, in opposition to the difficulty in appropriately calculating R_0 for individual based models (see [54, 17]). Moreover, by adopting the first generation index as comparison indicator, all the models are initialized in the same way.

Three different scenarios have been investigated, corresponding to $G_0=1.1;1.4;1.7$. All the simulations are initialized with only one infected individual, yielding a completely susceptible population. Estimate of the transmission rates in the different transmission places (household, school/workplace and community) leading to the chosen G_0 value is done by keeping trace of number and place of the secondary infections. A reference model (the M model for instance) is chosen and transmission coefficients are determined by an additional constraint on the proportions of cases generated by the different sources of infection considered in the model. In particular, the contribution of each of the three sources of infection is set to 1/3. For S and L models, the transmission rates within households and schools/workplaces are kept fixed, while a specific rate is selected for the transmission in the communities, satisfying the above constraint on the proportions of

cases generated by the different sources. Models including long-distance trips M+T and L+T) inherit transmission coefficients from the corresponding basic models.

This choice leads to within households and schools/workplaces transmission slightly smaller than in the respective basic models, because transmission during trips occurs only by random contacts in the population. For our choices of the transmission rates, the final proportion of cases generated by the three sources differs no more than 0.018 from 1/3. For each model and choice of the first generation index, an average of at least 20,000 runs were considered, to guarantee a sufficiently accurate estimate of the relative transmission parameters.

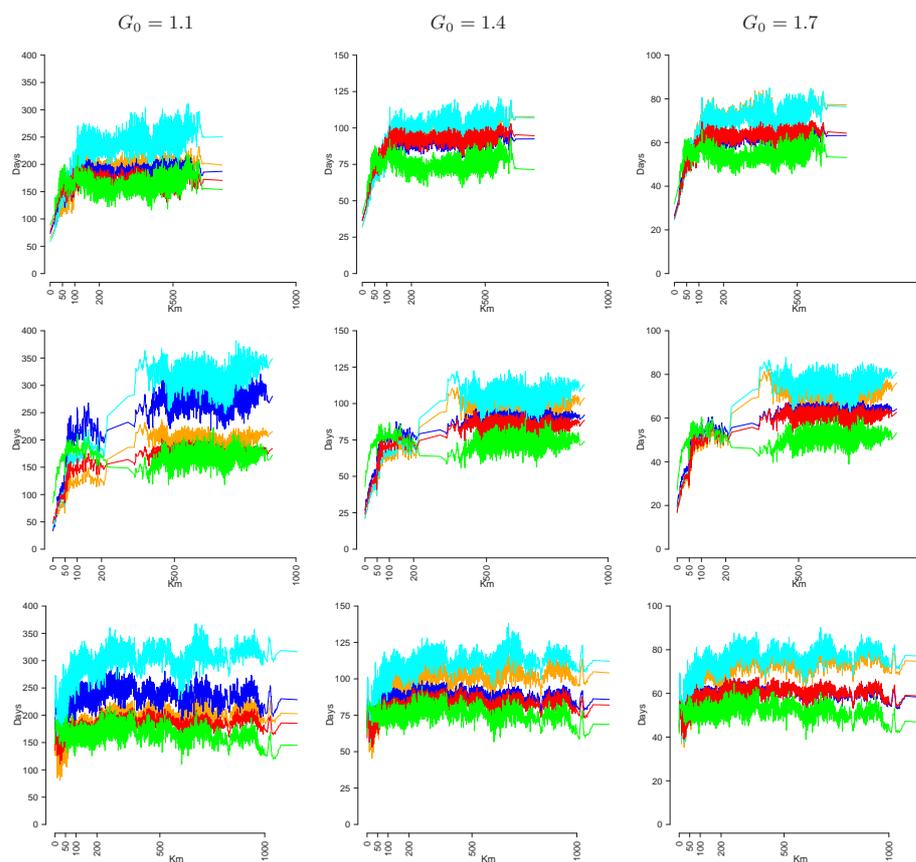


Figure 5.1: Spreading time from the seeding municipality as a function of the distance for different values of G_0 and different seeding municipalities: Rome (first row), Cagliari (second row), a small isolated village (third row). Model M in orange, M+T in red, L in cyan, L+T in blue, and S in green.

5.1.3 Results

More than 1,000,000 experiments were run to evaluate how the different approaches to modeling unstructured contacts can affect the spatiotemporal epidemic dynamics. For each considered model, different model instances were realized by varying the first generation index G_0 defined as the average number of secondary infections generated by the first infectious individual during the entire infectious period in a completely susceptible population (more details are given in Methods): G_0 values of 1.1, 1.4 and 1.7 were considered to simulate low to high transmission scenarios. We compared a variety of summary measures such as the probability of having a large outbreak, the epidemic evolution (attack rate, basic reproductive number, peak day, proportion of infected by age) and the spatial diffusion, i.e., the average distance from the seed area for individuals infected since the start of the epidemic as a function of time. For comparison's sake, all the simulations were seeded with only one infected individual, even though a pandemic influenza in a European country will be very likely sustained by mechanisms of case importation, e.g., by international travels [31, 32, 127, 41]. Different seeding municipalities were chosen to take into account the role played by the demographic size, density and geographic location of the seeding zone; we considered large cities, small/medium size towns, isolated villages, and, as an extreme case of isolated seeding region, islands.

The final attack rate of the considered models is significantly different (see Tab. 5.1) and it ranges from 19.1% to 25.7% for $G_0=1.1$, from 47.8% to 50.7% for $G_0=1.4$ and from 62.4% to 67.8% for $G_0=1.7$. No substantial differences are observed by varying the seeding municipality. The introduction of occasional long-distance trips substantially decreases the final attack rate of both the M and the L models. In fact, in our implementation, transmission is not allowed within household and within school/workplace during long-distance trips.

The basic reproductive number R_0 of the simulated epidemics is calculated as in [54, 28]

Table 5.1: *Final attack rate*. Final attack rates (with standard deviation) of the different models considered for different G_0 values.

Model/ G_0	1.1	1.4	1.7
M	25.7 (0.029)	50.7 (0.014)	64.6 (0.011)
M+T	21.4 (0.040)	47.7 (0.016)	62.4 (0.011)
L	26.9 (0.031)	50.7 (0.016)	67.6 (0.011)
L+T	22.9 (0.035)	47.8 (0.018)	65.7 (0.011)
S	19.6 (0.077)	48.6 (0.039)	64.8 (0.017)

Table 5.2: *Basic reproductive number*. Basic reproductive numbers R_0 (with standard deviation) of the different models considered for different G_0 values.

Model/ G_0	1.1	1.4	1.7
M	1.34 (0.018)	1.78 (0.010)	2.18 (0.011)
M+T	1.29 (0.022)	1.71 (0.010)	2.11 (0.013)
L	1.34 (0.022)	1.73 (0.011)	2.16 (0.016)
L+T	1.29 (0.025)	1.67 (0.011)	2.10 (0.011)
S	1.27 (0.029)	1.72 (0.013)	2.14 (0.008)

(see Methods). The observed R_0 values, among all the considered models, do not vary more than 0.07, 0.11 and 0.08 for $G_0=1.1$, 1.4 and 1.7 respectively (see Tab. 5.2). Note that R_0 is systematically larger than the average number of secondary cases generated by the primary infection in a wholly susceptible population, as observed in [54].

Significant differences can be detected in the spatial diffusion of the epidemic (see Fig. 5.1). For $G_0=1.1$, L models are spread systematically more slowly than the respective M models (with difference of about 100 days to cover 200 km). In fact, the set of unstructured contacts as considered in M models includes individuals living in or traveling to the same municipality where the individuals travel to, thus inducing a higher probability of exporting the epidemic. Due to the specific choice of kernel function and parameters, S models are spread systematically more quickly (with difference of about 100 days to cover 200 km with respect to M models). However, alternative choices of kernel function and parameters can lead to different model outputs. The behavior of L and M models tends to be similar when increasing the first generation index, while the S models are systematically the fastest. Not surprisingly, models including long-distance trips M+T and L+T spread quite faster than the respective M and L models (even though their attack rate is systematically lower), independently from the first generation index and the seeding municipality. Furthermore, the observed pattern of spatial spread strongly depends on the seeding region. For instance, when the epidemic is seeded in a small, isolated village, no clear pattern of diffusion is observable (especially for S models) since the epidemic is more likely to spread towards far, large cities than towards close, small size municipalities (see Fig. 5.1, third row). At a given distance, the variability observed in the time of epidemic arrival is basically determined by the variability in the population size of the arrival municipalities. Trivially, on average, the epidemic is very likely to spread first towards large population municipalities than towards small, isolated municipalities. When infection is seeded in very isolated regions, as Sardinia island, the models behave

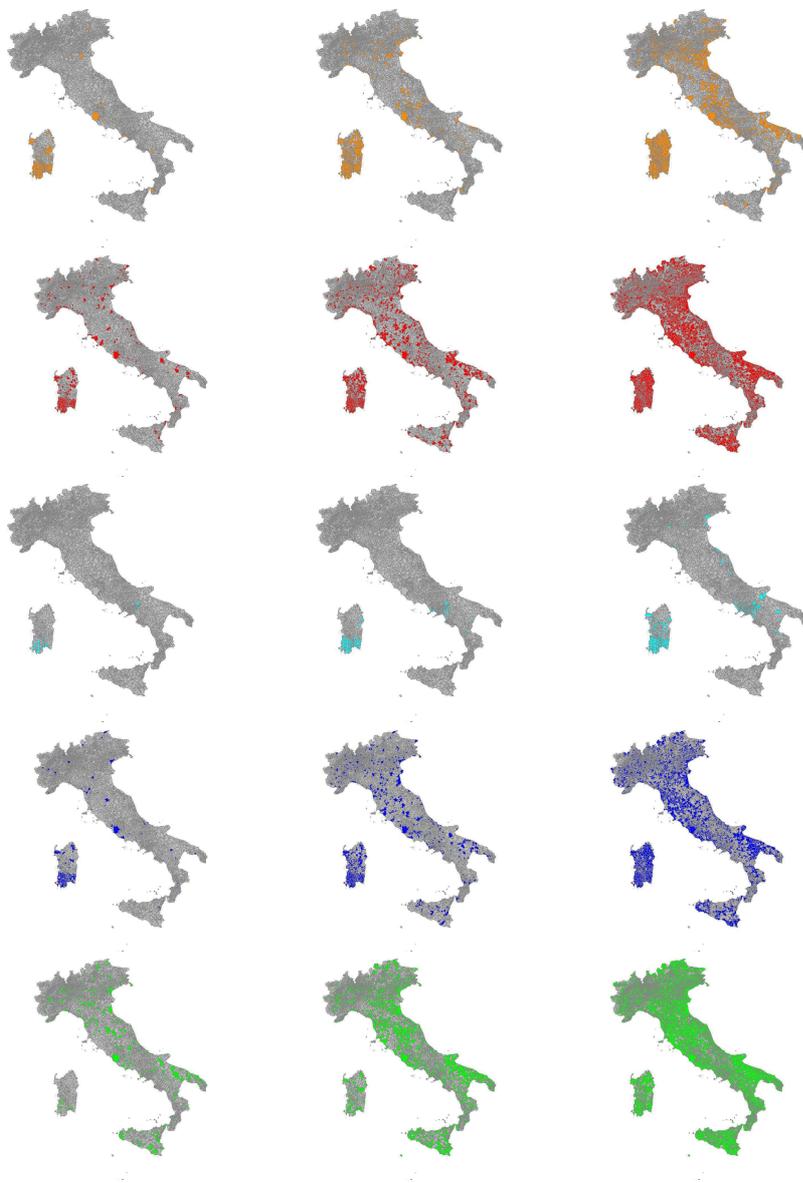


Figure 5.2: Spatiotemporal dynamics at 40 (left), 50 (center) and 60 (right) days. Infection is seeded in Cagliari (Sardinia island) and $G_0=1.7$. Colored areas (model M in orange, M+T in red, L in cyan, L+T in blue, and S in green) indicate presence of at least one infected, infectious or removed individual.

quite differently (see Fig. 5.1, second row). Basically, in M and L models the epidemic is spread on the entire island before being spread out to the rest of Italy. A similar behavior is observed in M+T and L+T models, even though it is not so pronounced, while in S model the epidemic is spread out in the first phase. In fact, only a very small fraction of

Table 5.3: *Peak day*. Peak day (with standard deviation) of the different models considered for different G_0 values.

Model/ G_0	1.1	1.4	1.7
M	287.1 (22.0)	143.7 (11.3)	104.5 (6.7)
M+T	294.9 (29.0)	137.9 (10.2)	98.9 (7.2)
L	439.3 (127.4)	153.6 (11.5)	107.3 (7.1)
L+T	407.6 (163.7)	142.9 (10.7)	99.3 (7.7)
S	302.0 (33.8)	127.4 (10.6)	90.3 (7.7)

Table 5.4: *Peak day for different seeding municipality*. Peak day (with standard deviation) for different seeding municipality and different values of the first generation index G_0 .

G_0	Municipality	M	M+T	L	L+T	S
1.1	Rome	287.2 (20.6)	289.3 (25.8)	429.5 (129.7)	396.3 (153.6)	299.0 (33.6)
	Cagliari	288.4 (27.3)	286.2 (26.2)	448.2 (126.5)	448.2 (174.1)	300.9 (33.7)
	Luserna	286.4 (13.8)	294.6 (26.4)	421.5 (143.7)	380.8 (144.9)	298.3 (38.0)
	Turin	286.2 (20.5)	288.9 (25.4)	427.9 (130.3)	388.3 (148.2)	298.8 (31.3)
	Vieste	287.1 (22.1)	294.9 (29.0)	439.4 (127.5)	407.7 (163.8)	302.1 (33.8)
1.4	Rome	144.9 (12.4)	136.7 (10.2)	154.4 (10.5)	142.1 (10.7)	127.9 (11.8)
	Cagliari	142.5 (9.6)	136.0 (10.2)	153.5 (9.4)	142.0 (11.0)	126.1 (12.3)
	Luserna	145.9 (12.7)	134.2 (8.4)	158.0 (10.3)	141.1 (9.9)	129.2 (11.5)
	Turin	143.6 (11.9)	136.6 (10.2)	155.0 (12.0)	142.8 (11.5)	127.6 (11.3)
	Vieste	143.8 (11.4)	138.0 (10.3)	153.6 (11.6)	143.0 (10.8)	127.4 (10.6)
1.7	Rome	106.1 (6.6)	97.8 (6.2)	108.7 (6.5)	98.5 (7.5)	89.7 (6.4)
	Cagliari	105.5 (5.6)	98.0 (6.2)	109.7 (6.3)	99.7 (7.5)	87.9 (6.1)
	Luserna	106.0 (6.3)	97.3 (5.9)	109.5 (6.0)	98.3 (7.3)	88.6 (6.8)
	Turin	104.9 (6.9)	98.0 (6.8)	107.9 (7.3)	98.9 (7.7)	89.6 (6.2)
	Vieste	104.5 (6.7)	99.0 (7.3)	107.4 (7.1)	99.4 (7.8)	90.4 (7.7)

Roma is the largest city of Italy (2,546,804 inhabitants), located in the central Italy.

Cagliari is a city (164,249 inhabitants) in the Sardinia island.

Luserna is a small isolated village (297 inhabitants) in the northern of Italy.

Turin is a big city (865,263 inhabitants), located in the northern Italy.

Vieste is a small town (13,430 inhabitants) in the southern Italy.

workers and students commutes to or from the island, greatly reducing the set of contacts outside the island in M and L models, while this is not the case for S models. However, note that the kernel parameters of the spatially explicit model were chosen on the basis of commuting data. While this is a reasonable choice for assigning commuting destination, it is unclear whether this is the best choice for modeling the spatial spread of an epidemic through unstructured contacts. Completely different behaviors are to be expected when adopting different kernel shapes. Although the spatially explicit model is flexible, it requires detailed data, both demographic and epidemiological, for choosing the optimal kernel and kernel parameters.

Table 5.5: *Probability of a large outbreak.* Probability (with standard deviation) of a large outbreak for the different models considered and for different G_0 values.

Model/ G_0	1.1	1.4	1.7
M	0.184 (0.050)	0.449 (0.027)	0.572 (0.060)
M+T	0.184 (0.096)	0.446 (0.072)	0.598 (0.055)
L	0.192 (0.047)	0.460 (0.079)	0.628 (0.023)
L+T	0.154 (0.040)	0.464 (0.086)	0.661 (0.013)
S	0.201 (0.047)	0.474 (0.007)	0.662 (0.021)

Significant differences are observed in the peak day (see Tab. 5.3). In particular, for large values of G_0 ($G_0 \geq 1.4$) the epidemic peak of S models occurs systematically earlier than M and L models (with differences of about 15 to 20 days for different values of the first generation index). Since in S models the epidemic is spread much more quickly, new infection foci occur simultaneously in many different regions, thus inducing a spatial synchronization of the epidemic. No substantial differences are observed by varying the seeding region (see Tab. 5.4). For $G_0=1.1$, no significant differences are observed between M and S models, while, on average, the peak day of L models occurs later than M and S models. This is due to the several simulations behaving very differently from all the others (and independently from the seeding region), characterized by a very long initial phase and giving rise to a high standard deviation. In fact, for low values of the first generation index, L models are less likely to spread out the epidemic because of the reduced set of unstructured contacts. Not surprisingly, the introduction of occasional long-distance trips significantly anticipates the epidemic peak in both the M and L models (5 to 10 days earlier than the respective M and L models). See also Fig. 5.3 where the number of cases in time of the different models are reported for different seeding municipalities and different first generation indices.

Differences are also observed in the proportion of infected by age (see Fig. 5.4). In order to compare the different models, the curves are normalized, and we consider the indicator $\frac{a_i}{\sum a_i}$, where a_i is the proportion of infected of age i . Independently from the first generation index, the proportion of infected generated by M models in individuals older than 65 years is lower than for other models, while the opposite behavior is observed for individuals younger than 65 years. In terms of unnormalized proportion of infected, differences of 5% to 10% are observed in the older individuals for $G_0=1.7$. In fact, in M models the set of unstructured contacts of infectious individuals proportionally includes a larger number of traveling individuals (i.e., with age between 3 and 65 years). Consequently, this latter

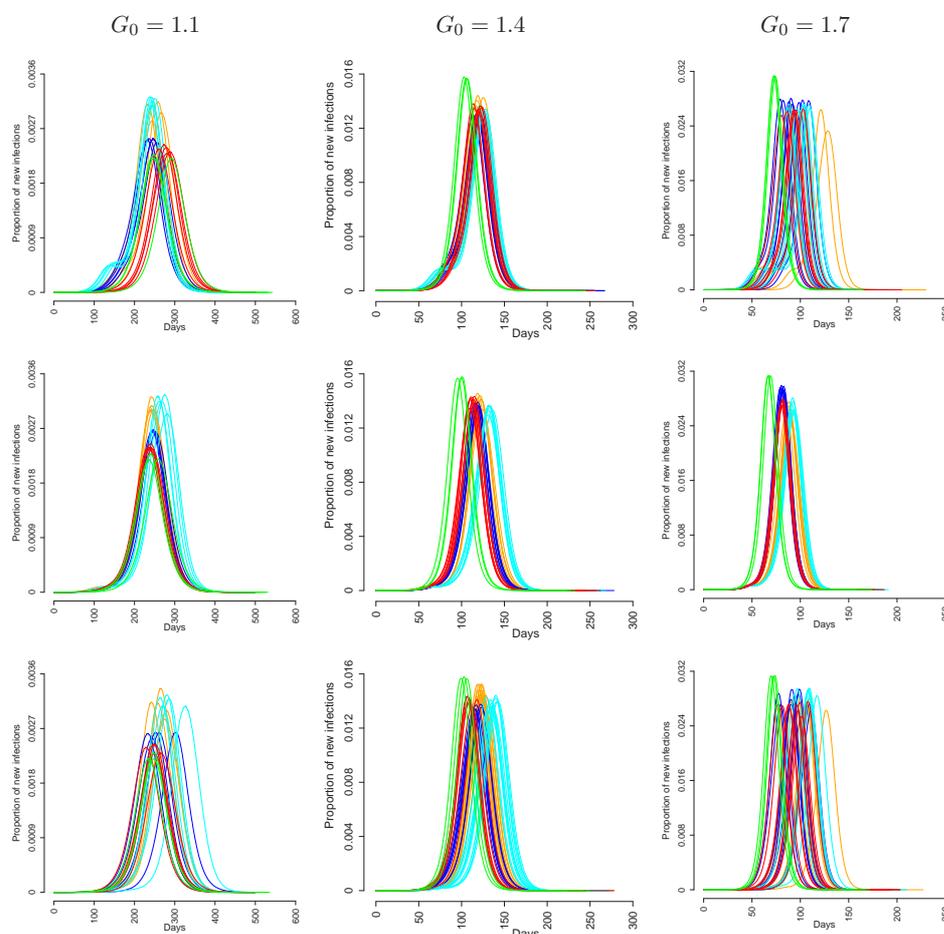


Figure 5.3: Number of daily cases for different values of the first generation index and different seeding municipalities: Rome (first row), Cagliari (second row), a small isolated village in the north of Italy (third row). Simulation are initialized with 30 infected individuals, to reduce the stochastic variability observed in first days of the epidemic. Models are: M in orange, M+T in red, L in cyan, L+T in blue, and S in green.

class of individuals is proportionally more exposed to contacts with infectious ones. This is not the case for age independent unstructured contacts models, as S and L models. In [46, 47, 54], the authors introduce additional parameters to make the unstructured contacts dependent on age, while for M models this is obtained in a natural way. The slightly larger proportion of infected observed in individuals aged 35-45 is due to the structured component of the contacts: in fact, they have a higher probability of living with individuals aged 3-18, the most infected class.

Finally, no substantial differences are detected in the probability of observing a large

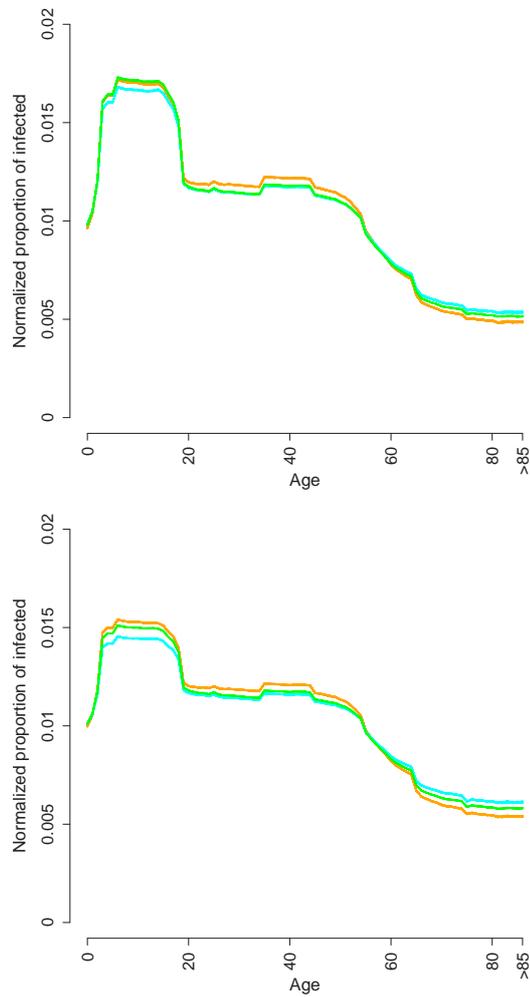


Figure 5.4: Proportion of the infected population by age for $G_0=1.4$ (top) and $G_0=1.7$ (bottom). Models M in orange, L in cyan and S in green.

outbreak for $G_0=1.1$ and $G_0=1.4$ (see Tab. 5.5). For $G_0=1.7$, the probability of a large outbreak in S models is larger than that observed for the other models.

A sensitivity analysis on the effect of varying the number of travel days in models M+T and L+T and on the effect of varying the spatial kernel in models S is carried on as already been performed.

5.1.4 Conclusions

In principle, it would be possible to improve the characterization of structured contacts, for instance by employing data on contacts on neglected but important activities, such as

leisure time, sport mall, restaurants, etc. and time-use data to provide useful information for parameterizing IBM. Such information starts being available (see the EU project Polymod) and it will be mandatory to integrate it in the next generation of IBM. However, it is not possible to take into account all the possible sources of infections. In fact, this would mean tracing all the possible contacts (which in turn requires to model all the places where these contacts occur, how much time is spent in each place, etc.), to establish the "type" of contacts (e.g., skin to skin or indirect) of each individual, which is unfeasible. It is thus required to consider in the models a source of infection accounting for pseudorandom contacts.

The scenarios emerging from the conducted experiments in terms of final attack rate, spatial spread, epidemic peak day and proportion of infected by age are quite heterogeneous. In particular, epidemics generated by the spatially explicit model spread much more quickly than those simulated by all the other models, regardless of first generation index and seeding region. Also, the epidemic peak occurs systematically earlier, probably because of spatial synchronization effects. Defining unstructured contacts on the basis of commuting data rather than randomly choosing them in the local communities results in a faster epidemic, especially for lower values of the first generation index, in terms of both spatial diffusion and peak day. The effects of occasional long-distance trips are the speeding up of the spatial diffusion and the decreasing of the cumulative attack rate. The proportion of infected by age is also significantly different. Specifically, the proportion of infected in the younger and adult age groups is larger in the models where random contacts are defined on the basis of commuting data while the proportion of infected in the older age groups is lower. No significant differences are observed in the probability of having a large outbreak, especially for small first generation indices.

Wide differences in the models' outputs can result in different evaluations of the effectiveness of the containment/mitigation strategies and they would seriously undermine the usefulness of our models, thus urgently calling for field work aimed at filling this data-gap. In fact, even though the containment strategies are in general based on the structured part of the contacts (social distancing measures, e.g. school and workplaces closure, antiviral prophylaxis on a contact tracing basis), the way we choose to model the unstructured part of the contacts can lead to very different scenarios.

A detailed analysis of the implications in terms of containment strategies evaluation is beyond the aim of this work. A few considerations can be drawn, anyway. Trivially, the cumulative attack rates are quite different, even though the models are initialized in the same way, thus leading to different evaluations of the effectiveness of the same containment measure. More specifically, the difference observed in the peak day can result in different

evaluations on the effectiveness of vaccination campaigns. For instance, in a mass vaccination campaign against a pandemic with $G_0=1.4$, by fixing target population at 60%, vaccine efficacy at 70% and vaccine availability at 4 months after the first national case, the number of avoided cases is, on average, 24.4 millions for the M model, 20.2 millions for the M+T model, 26.8 millions for the L model, 22.7 millions for the L model and 13.4 millions for the S model. Moreover, ignoring the variations in the proportion of infected by age can result in wrong decisions when optimizing the choice of the target population for a vaccination campaign. Furthermore, the variability of the spatial spread can influence the evaluation of strategies based on geographical targeting. We can mention the choice of the dimension of quarantine areas, the effectiveness of antiviral prophylaxis on a geographical basis and the timing for closing schools and workplaces: for instance, close them all simultaneously or wait for a few cases to arise? While the observed differences could not drastically undermine the results in terms of feasibility of the considered interventions (the principal objective of many independent studies), nevertheless they could be relevant in terms of optimality.

Wide differences in the models' outputs can result in different evaluations of the effectiveness of the containment/mitigation strategies. Consequently, all the possible effects of different assumptions should be considered for taking public health decisions: not only sensitivity analysis to various model parameters should be performed, but intervention options should be based on the analysis and comparison of different modeling choices, as it happens in different fields, e.g. global climate change, where uncertainty in the models themselves and in input parameters is a critical factor.

We conclude remarking that unlike what shown in most of the literature [46, 47, 54, 114], no supercomputing techniques have to be employed to perform this kind of simulations on a national scale (57,000,000 of individuals), making them feasible for a standard workstation; our implementation of the five model takes less than 3Gb RAM and a single simulation takes just a few minutes.

5.2 Mitigating an influenza pandemic in Italy

For these reasons, countries have been urged to strengthen their preparedness plans [70], and several countries have considered stockpiling both antiviral drugs and monovalent influenza vaccines containing potentially pandemic strains, such as A/H5N1 (i.e., a pre-pandemic vaccine), for population priming.

However, some control measures can be costly (e.g., stockpiling antiviral drugs, vaccines, and a pre-pandemic vaccine), and others could have limited social acceptance (e.g., closure of schools/workplaces and travel restrictions). For these reasons, several countries have used mathematical models to predict the spread of infection at the national level, which is an important aspect of preparedness, and to evaluate the feasibility of containing the pandemic using different strategies [46, 47, 54, 85, 87, 31, 32, 35].

Individual-based models can provide the most reliable estimates of the spread of influenza [46, 47, 87, 54, 29]. In the present study, we evaluated the diffusion of pandemic influenza in Italy and the impact of various control measures, coupling a global SEIR model with an individual based model. We used actual demographic data, obtained from the 2001 census, which allowed us to simulate the spread of an influenza pandemic and the impact of control measures. In particular, we examined the impact of antiviral prophylaxis of close contacts, social distancing measures, international air travel restrictions, and vaccination (both pandemic and pre-pandemic vaccine), under different R_0 values. Since it has been shown that seasonal influenza vaccine effectiveness is higher in adults than in elderly persons and children [36, 79, 59], we also assumed that both pandemic and pre-pandemic vaccine effectiveness would vary by age.

5.2.1 Control measures

We considered the following control measures:

- vaccination,
- antiviral treatment and/or prophylaxis (AVP),
- social distancing,
- air travel restrictions.

Vaccination. The target population was divided into 4 categories: i) personnel providing essential services (15% of the 25-60-year-old working population); ii) elderly persons

(≥ 65 years of age); iii) children and adolescents from 2 to 18 years of age; and iv) adults from 40 to 64 years of age. Vaccination was modeled by reducing the proportion of susceptible individuals in the target population. This proportion depends on vaccination coverage (VC) and vaccine effectiveness (VE). We assumed that vaccination consists of two vaccine doses administered one month apart and that VC was 60% of the target population. This VC was chosen on the basis of the 2005-2006 seasonal influenza coverage, which was 68% in elderly persons (≥ 64 years) [100]. We assumed that one week is necessary for administering each vaccine dose to all target categories. Vaccination was considered to be effective beginning 15 days after the administration of the second dose. Three different assumptions on VE were considered: i) VE of 70%, for all age-groups; ii) VE of 50%, for all age-groups; and iii) VE of 59 for individuals aged 2-18 years [79], 70% for individuals aged 40-64 years [36], and 40% for individuals aged ≥ 65 years [59]. We assumed that individuals are vaccinated irrespective of whether or not they were infectious or ill.

When considering the impact of single interventions, we assumed that vaccination begins 1, 2, 3, 4, 5 or 6 months after the first world case, targeting three of the target categories (i.e., personnel providing essential services, elderly persons, and 2-18 year-olds), and assuming a VE of 70% for all three categories. When considering multiple interventions, we assumed that vaccination begins at 2, 3 or 4 months after the first world case, and we considered the different assumptions for VE reported above (70% for all, 50% for all, or varying by age).

Given that an estimated 3-6 months would be required to produce pandemic influenza vaccines, the administration of a first dose within 3 months of the first world case would be possible only if this dose contained a precursor of the pandemic strain [104], followed by a dose of pandemic vaccine. The actual VE of this regimen was assumed to be equal to that of two doses of the pandemic vaccine.

Antivirals. We took into consideration the administration of one course of antiviral drugs, providing therapy for the index case and prophylaxis for close contacts. Both therapy and prophylaxis were assumed to start one day after clinical onset in the index case. The treatment of the index case was assumed to reduce infectiousness by 70% [46, 47, 54, 87, 29], whereas AVP was assumed to reduce susceptibility to infection by 30%, infectiousness by 70%, and the occurrence of symptomatic disease by 60% [29, 87].

We assumed that AVP be provided to 90% of the close contacts of clinical cases (50% of all infected individuals), with a treatment course of 10 days [85]. Two different definitions of close contacts were used: i) household contacts only; and ii) household contacts plus

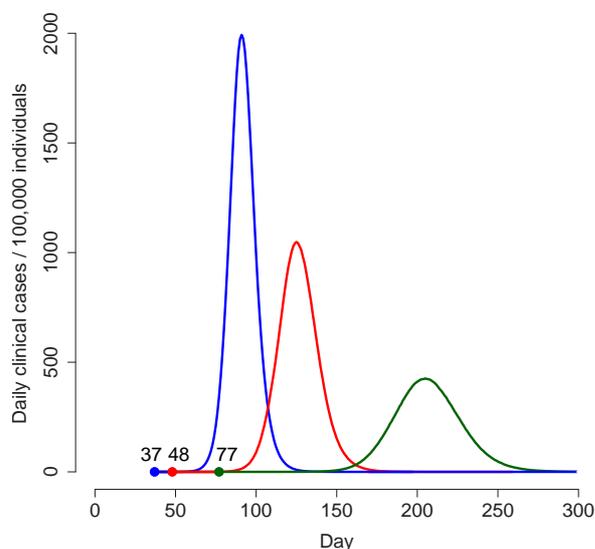


Figure 5.5: Baseline simulations under different R_0 scenarios (blue line: $R_0=2$; red line: $R_0=1.7$; green line: $R_0=1.4$). Bullet points represent the first Italian case and the time elapsed from the first world case.

close contacts in the school or workplace. We considered administering AVP for the entire epidemic period; however, since the feasibility of actually doing this would be limited, we also considered administrating AVP as a policy to be used only for the first 8 weeks after the occurrence of the first Italian case.

Social distancing. We considered the nationwide closing of all schools and some public offices not providing essential services, corresponding to 20% of all employees in these types of offices (from 8^o Censimento dell'Industria e dell'Artigianato, ISTAT, 2004). We assumed that school and office closings begin 4 weeks after the onset of the first 20 symptomatic cases in Italy and that this measure be maintained for 4 weeks.

We also assumed that symptomatic individuals spontaneously limit their school/work attendance. The proportion of symptomatic individuals staying at home from school/workplace would vary by age, from 90% among children ≤ 6 years of age to 50% among the working population.

Air travel restrictions. We considered travel restrictions that would reduce incoming international flights by 90% or 99%, starting from day 30 of the first world case [47] and lasting for the entire duration of the epidemic, or until two months after the introduction of the first case in Italy. The reduction of domestic air travel and the control of land and

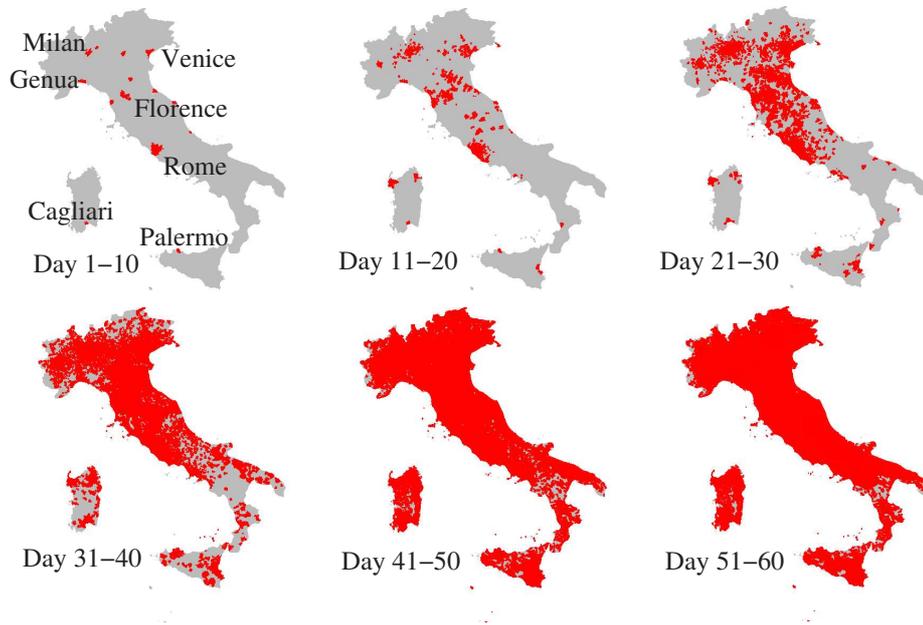


Figure 5.6: Spatial spread of pandemic influenza in Italy, $R_0=1.7$. Red areas represent municipalities where at least one case is present.

sea borders were not considered in the model.

5.2.2 Results

Baseline dynamics

For different R_0 scenarios, the results of the global SEIR model showed that the number of imported symptomatic cases would be 53,000, 72,000, and 83,000, with the first Italian case appearing, respectively, after 77, 48 and 37 days; the epidemic curves for these scenarios are shown in Fig. 5.5.

For $R_0=1.7$, the spatial spread of the epidemic showed that for the first 10 days the new cases would be confined to the municipalities where cases were first imported (Fig. 5.6). At 11-20 days, new cases would begin to occur far from these municipalities, mainly in municipalities with a large population. At 21-40 days (the exponential growth phase), infection would spread simultaneously to nearly the entire country, with no clear spatial pattern.

The epidemic peak is reached after 202, 125 and 91 days, respectively, for the three different scenarios (Fig. 5.5). The pandemic season at the national level would last for a period of 3 to 6 months, with an average of 67,000-243,000 clinical cases per day.

Table 5.6: *Mild scenario: Clinical attack rates and peak day, by control measure.*

Control measure	AR [%]	(95% CI)	Peak day [days]	(95% CI)	Peak daily AR [%]	(95% CI)	Courses [millions]
None	21.2	(21.1-21.3)	202	(200-204)	0.42	(0.41-0.43)	-
Air travel restriction							
90%	21.2	(21.1-21.2)	225	(223-228)	0.42	(0.42-0.43)	-
99%	21.1	(21.1-21.2)	241	(233-245)	0.40	(0.38-0.41)	-
Social distancing							
Closure of schools and workplaces for 4 weeks	21.2	(21.1-21.3)	207	(205-211)	0.43	(0.43-0.43)	-
Antiviral prophylaxis							
Household contacts	4.6	(4.5-4.7)	255	(251-263)	0.03	(0.03-0.03)	7.3
Vaccination (time from the first world case)							
1 month	7.4	(7.4-7.5)	238	(235-239)	0.07	(0.07-0.07)	14.4
2 months	7.5	(7.4-7.6)	237	(234-241)	0.07	(0.07-0.07)	14.4
3 months	7.5	(7.4-7.6)	235	(233-238)	0.07	(0.07-0.07)	14.4
4 months	7.7	(7.6-7.7)	218	(211-222)	0.08	(0.08-0.08)	14.4
5 months	10.5	(9.9-11.2)	182	(178-186)	0.24	(0.20-0.28)	14.4
6 months	17.5	(16.9-18.3)	202	(200-203)	0.43	(0.43-0.44)	14.4

Vaccine efficacy (VE) = 70%.

Vaccination target categories: personnel providing essential services, elderly persons, individuals 2–18 years of age.

The cumulative infected AR would be 42.4%, 61.6% and 77.4%, for the three scenarios, corresponding to a clinical AR of 21.2%, 30.8%, and 38.7%. The clinical daily-peak AR would be 0.4%, 1.0% and 1.9%, respectively. Only the clinical AR is considered below.

Impact of control measures

Single measures. The results of the single control measures for different scenarios are reported in Tab. 5.6, 5.7, 5.8. International air travel restriction would not affect the AR but could delay the importation of cases, increasing the time elapsed from the first world case to importation from a minimum of 7 days to a maximum of 37 days, depending on the R_0 and the level of restriction. The pandemic peak would also be delayed by 6–39 days (Tab. 5.6, 5.7, 5.8; Fig. 5.7). Nationwide closure of schools and workplaces not providing essential services would delay the time of occurrence of the peak by 5–8 days, depending from the scenario considered.

AVP appears to be the most effective single intervention, resulting in a 36%–76% reduction in cumulative ARs. It also contributes to delay the peak day (from 13 to 53 days) and to decrease the peak daily attack rate.

Vaccination impact strongly depends from its timing. In the mild scenario, it would reduce the cumulative AR by approximately 65%, if it is begun within 4 months of the

Table 5.7: *Moderate scenario: Clinical attack rates and peak day, by control measure.*

Control measure	AR [%]	(95% CI)	Peak day [days]	(95% CI)	Peak daily AR [%]	(95% CI)	Courses [millions]
None	30.8	(30.7-30.9)	125	(123-126)	1.03	(1.02-1.05)	-
Air travel restriction							
90%	30.8	(30.7-30.9)	135	(133-138)	1.01	(0.99-1.03)	-
99%	30.8	(30.8-30.9)	150	(146-152)	1.03	(0.97-1.10)	-
Social distancing							
Closure of schools and workplaces for 4 weeks	30.8	(30.7-30.9)	132	(130-135)	1.03	(1.02-1.04)	-
Antiviral prophylaxis							
Household contacts	15.5	(15.4-15.6)	150	(149-151)	0.27	(0.27-0.28)	23.5
Vaccination (time from the first world case)							
1 month	17.8	(17.7-17.9)	144	(141-146)	0.36	(0.36-0.37)	14.4
2 months	18.1	(18.0-18.1)	133	(132-134)	0.36	(0.36-0.37)	14.4
3 months	25.5	(24.8-25.9)	125	(122-126)	1.0	(0.98-1.01)	14.4
4 months	30.6	(30.5-30.7)	125	(124-127)	1.0	(0.98-1.01)	14.4
5 months	30.8	(30.7-30.9)	125	(124-127)	1.0	(0.98-1.01)	14.4
6 months	30.8	(30.7-30.9)	125	(124-127)	1.0	(0.98-1.01)	14.4

Vaccine efficacy (VE) = 70%.

Vaccination target categories: personnel providing essential services, elderly persons, individuals 2–18 years of age.

Table 5.8: *Severe scenario: Clinical attack rates and peak day, by control measure.*

Control measure	AR [%]	(95% CI)	Peak day [days]	(95% CI)	Peak daily AR [%]	(95% CI)	Courses [millions]
None	38.7	(38.6-38.8)	91	(89-92)	1.93	(1.87-1.97)	-
Air travel restriction							
90%	38.7	(38.6-38.8)	97	(95-100)	1.90	(1.86-1.93)	-
99%	38.7	(38.6-38.8)	108	(106-109)	1.91	(1.84-1.94)	-
Social distancing							
Closure of schools and workplaces for 4 weeks	38.7	(38.6-38.8)	99	(97-102)	1.91	(1.89-1.94)	-
Antiviral prophylaxis							
Household contacts	24.9	(24.8-25.0)	104	(100-106)	0.76	(0.75-0.77)	35.4
Vaccination (time from the first world case)							
1 month	27.1	(27.1-27.2)	99	(97-100)	0.94	(0.94-0.96)	14.4
2 months	36.3	(36.1-36.5)	92	(89-93)	1.91	(1.90-1.94)	14.4
3 months	38.7	(38.6-38.8)	92	(89-93)	1.91	(1.90-1.94)	14.4
4 months	38.7	(38.6-38.8)	91	(89-92)	1.93	(1.87-1.97)	-
5 months	38.7	(38.6-38.8)	91	(89-92)	1.93	(1.87-1.97)	-
6 months	38.7	(38.6-38.8)	91	(89-92)	1.93	(1.87-1.97)	-

Vaccine efficacy (VE) = 70%.

Vaccination target categories: personnel providing essential services, elderly persons, individuals 2–18 years of age.

first world case (Tab. 5.6,5.7,5.8). In the moderate and severe scenarios, ARs would be reduced by 42% and 31% respectively, if vaccination starts within 2 months and one month from the pandemic start (Tab. 5.6,5.7,5.8).

Combined measures. Tab. 5.9 shows the impact of combining vaccination with international air travel restrictions. In the mild scenario, there is no clear added value of air travel restriction. In the moderate and severe scenarios, the implementation of 99% of air travel restriction would allow to have one additional month to implement vaccination, since administering first dose within three months instead of two, for the moderate scenario, and within two months instead of one, for the severe scenario, would not modify cumulative AR.

When combining all of the measures, in the mild scenario, the epidemic could be mitigated with moderate efforts. Specifically, performing vaccination for three target categories (i.e., personnel providing essential services, elderly persons, and 2-18 year-olds) and providing AVP to 90% of household contacts for the entire epidemic period would reduce the cumulative AR by 98% (from 21% to 0.3%), independently of the timing of vaccination (2, 3 or 4 months) and the implementation of air-travel restrictions. Limiting AVP to 8 weeks would produce a cumulative AR of 7.7%, which is similar to that observed with vaccination alone.

For the moderate scenario (Tab. 5.9, Fig. 5.7), vaccinating the three above-mentioned target categories and providing AVP to 90% of household contacts for the entire epidemic period, with 90% air travel restriction would reduce the cumulative AR by 77%-87% (from 31% to 4-7%, depending on the timing of the first vaccine dose). The cumulative AR is reduced by 87%, if 99% air-travel restrictions were implemented, and vaccination were begun within 4 months of pandemic start.

If air-travel restrictions were not implemented or were limited to the first two-months after the first national case, the AR decrease would be similar (81-87%), providing that vaccination were started within 3 months of the first world case. The cumulative AR would be even lower (2%, for first dose at 3 months) if AVP were provided to both household contacts and close contacts in schools and workplaces. This would require the administration of 11 millions of AV courses.

For the severe scenario (Tab. 5.10, Fig. 5.7), the cumulative AR would decrease by 64% (from 39% to 14%) if the first vaccine dose were administered within 2 months of the first world case, AVP were provided to household contacts for the entire epidemic period, and 90% air-travel restriction were implemented, independently from its duration. The cumulative AR would further decrease (to 9%) if also vaccinating 40-64-year-old individ-

Table 5.9: *Moderate scenario: Clinical Attack rates, peak day, peak daily attack rate and number of courses by combination of control measures.*

Control measure	Final AR [%] (95% CI)	Peak day [days] (95% CI)	Peak daily AR [%] (95% CI)	AV courses (millions)	Vaccine courses (millions)
90% air travel restriction					
AVP ^u ; fist vaccine dose at 2 months	4.5 (4.4-4.5)	213 (209-215)	0.04 (0.04-0.04)	6.8	14.4
AVP ^u ; first vaccine dose at 3 months	4.6 (4.6-4.7)	186 (177-197)	0.04 (0.04-0.04)	7.1	14.4
AVP ^u ; first vaccine dose at 4 months	6.7 (6.2-7.2)	154 (150-156)	0.15 (0.12-0.18)	10.0	14.4
AVP ^u ; first vaccine dose at 2 months*	5.7 (5.6-5.7)	214 (206-219)	0.05 (0.05-0.06)	8.7	14.4
AVP ^u ; first vaccine dose at 3 months*	5.8 (5.7-5.9)	194 (189-197)	0.06 (0.06-0.06)	8.9	14.4
AVP ^u ; first vaccine dose at 4 months*	7.4 (7.2-7.8)	155 (151-156)	0.14 (0.12-0.16)	11.3	14.4
AVP ^u ; first vaccine dose at 2 months†	7.1 (7.0-7.1)	211 (207-217)	0.07 (0.07-0.07)	10.8	14.4
AVP ^u ; first vaccine dose at 3 months†	7.2 (7.1-7.3)	187 (179-194)	0.08 (0.08-0.08)	11.0	14.4
AVP ^u ; first vaccine dose at 4 months†	8.6 (8.2-9.0)	155 (151-156)	0.15 (0.12-0.18)	13.1	14.4
AVP for 8 weeks; first vaccine dose at 3 months	18.2 (18.1-18.3)	171 (167-176)	0.38 (0.34-0.4)	0.1	14.4
AVP ^u plus school/workplace close contacts; first vaccine dose at 3 months	2.1 (2.0-2.2)	141 (127-166)	0.02 (0.02-0.03)	10.7	14.4
AVP ^u ; first vaccine dose at 3 months, not vaccinating the elderly	5.3 (5.2-5.4)	192 (188-194)	0.05 (0.05-0.05)	8.0	8.5
AVP ^u ; first vaccine dose at 3 months, vaccinating also adults	2.3 (2.2-2.3)	186 (184-188)	0.02 (0.02-0.02)	3.6	24.6
AVP ^u , first vaccine dose at 3 months, time-limited border restrictions**	5.1 (5.0-5.1)	165 (161-170)	0.05 (0.05-0.06)	7.8	14.4
99% air travel restriction					
AVP ^u ; first vaccine dose at 2 months	4.4 (4.4-4.5)	274 (253-280)	0.04 (0.04-0.04)	6.7	14.4
AVP ^u ; first vaccine dose at 3 months	4.4 (4.3-4.5)	251 (246-257)	0.04 (0.04-0.04)	6.7	14.4
AVP ^u ; first vaccine dose at 4 months	4.6 (4.5-4.6)	222 (210-230)	0.04 (0.04-0.04)	7.0	14.4
No air travel restriction					
AVP ^u ; first vaccine dose at 2 months	4.6 (4.6-4.8)	163 (162-165)	0.04 (0.04-0.04)	7.2	14.4
AVP ^u ; first vaccine dose at 3 months	6.2 (6.1-6.3)	126 (123,129)	0.12 (0.12-0.13)	9.5	14.4
AVP ^u ; first vaccine dose at 4 months	11.0 (10.8-11.1)	152 (151-156)	0.28 (0.27-0.28)	10.2	14.4

VE vaccine efficacy, 70% (unless otherwise specified).

Vaccination target categories: Personnel providing essential services, elderly persons, individuals 2-18 years of age (unless otherwise specified).

AVP Antiviral prophylaxis.

* Different vaccine effectiveness for different categories: 59% in individuals 2-18 years of age [79], 70% in individuals 40-64 years of age [36], and 40% in ≥ 65 year-olds [59].

^u Unlimited, household contacts.

† Vaccine effectiveness = 50%.

** Air travel restrictions for 2 months after the first national case.

uals, which would reduce the number of household contacts receiving AVP by 33%. If not vaccinating 40-64-year-olds and providing AVP to both household contacts and close contacts in schools and workplaces, the cumulative AR would decrease to 8%, though this would require an extremely high number of AVP doses (approximately 32 millions).

Finally, with the implementation of 99% air-travel restriction, starting vaccination within three months of pandemic emergence would have the same impact than starting vaccination within two months, with no air-travel restrictions in place (cumulative AR=16%). None of the other combinations of control measures would reduce the cumulative clinical AR to less than 16%.

Assuming a VE of 50% for all age-groups or a different VE by age group (i.e., 59% in individuals aged 2-18 years, 70% in individuals aged 40-64 years, and 40% in individuals ≥ 65 years) would not substantially affect the cumulative AR; in fact, the cumulative AR would be 2 or 3 percentage points higher, respectively, than observed assuming a 70% VE for all age groups (Tab. 5.9 and Tab. 5.10).

Figure 4 shows the cumulative AR by age and vaccination strategy. If no control measures were performed (baseline), the cumulative AR would be highest for individuals ≥ 18 years of age and would decrease with increasing age. None of the considered scenarios included vaccinating 18-25-year-old individuals, who consequently appear to be the age-group with the highest incidence after vaccination. However, if vaccinating personnel providing essential services (15% of the 25-60-year-old working population), elderly persons (≥ 65 -year-olds), and 2-18 year-olds, the AR would also decrease among individuals 19-64 years of age, who are not targeted by vaccination. In particular, the AR would decrease by approximately 75% in unvaccinated 30-50-year-old individuals. Excluding the elderly from vaccination would not affect the cumulative AR in the other age groups.

5.2.3 Conclusions

Recent modeling studies have estimated that the first cases of influenza in a future pandemic would be imported to Europe within 50-90 days of its emergence elsewhere in the world [50, 47]. Our results indicate that the first cases would be imported to Italy within 37-77 days, depending on the R_0 , and that the incidence would peak 54-125 days after importation. When considering separately the three scenarios in our study, the timing of the peak for the severe scenario (i.e., 54 days) was similar to that for the severe scenario in the UK (i.e., 50 days), whereas it differed for the moderate scenario (i.e., 77 days for Italy compared to 65 days for the UK) [47]. The reason for this divergence is likely due to the different R_0 values considered in the global SEIR model, which were scaled in order to be proportional to those considered in the national IBM simulations (i.e., 1.4, 1.7, and 2). Varying the global R_0 , can in fact substantially modify the timing of national first case introduction, and the consequent epidemic peak. The lower number of air travelers coming into Italy per year compared to US and UK (25 millions, versus 73 and 92 millions,

Table 5.10: *Severe scenario: Clinical Attack rates, peak day, peak daily attack rate and number of courses by combination of control measures.*

Control measure	Final AR [%] (95% CI)	Peak day [days] (95% CI)	Peak daily AR [%] (95% CI)	AV courses (millions)	Vaccine courses (millions)
90% air travel restriction					
AVP ^u ; fist vaccine dose at 2 months	14.4 (14.4-14.5)	132 (130-134)	0.28 (0.28-0.29)	21.3	14.4
AVP ^u ; first vaccine dose at 3 months	20.5 (20.1-20.8)	124 (122-125)	0.79 (0.77-0.80)	29.1	14.4
AVP ^u ; first vaccine dose at 4 months	24.6 (24.5-24.7)	124 (122-125)	0.79 (0.77-0.80)	34.9	14.4
AVP ^u ; first vaccine dose at 2 months*	16.2 (16.1-16.3)	130 (129-131)	0.34 (0.33-0.34)	23.7	14.4
AVP ^u ; first vaccine dose at 3 months*	21.1 (21.0-21.3)	124 (123-127)	0.78 (0.77-0.80)	30.0	14.4
AVP ^u ; first vaccine dose at 4 months*	24.7 (24.6-24.7)	124 (122-126)	0.78 (0.77-0.80)	35.0	14.4
AVP ^u ; first vaccine dose at 2 months†	17.6 (17.4-17.7)	130 (127-132)	0.40 (0.39-0.40)	25.8	14.4
AVP ^u ; first vaccine dose at 3 months†	21.6 (21.3-22.0)	124 (122-125)	0.79 (0.77-0.80)	30.8	14.4
AVP ^u ; first vaccine dose at 4 months†	24.7 (24.6-24.8)	124 (122-125)	0.79 (0.77-0.80)	35.0	14.4
AVP for 8 weeks; first vaccine dose at 2 months	27.4 (27.3-27.4)	126(123-130)	0.98 (0.97-0.99)	1.5	14.4
AVP ^u plus school/workplace close contacts; first vaccine dose at 2 months	7.9 (7.7-8.1)	117 (101-127)	0.14 (0.13-0.15)	31.8	14.4
AVP ^u ; first vaccine dose at 2 months, not vaccinating the elderly	16.0 (16.0-16.1)	131 (129-134)	0.32 (0.31-0.33)	23.2	8.5
AVP ^u ; first vaccine dose at 2 months, vaccinating also adults	9.0 (8.8-9.3)	126 (119-132)	0.16 (0.15-0.17)	13.6	24.6
AVP ^u , first vaccine dose at 2 months, time-limited border restrictions**	14.7 (14.6,14.8)	125 (121-128)	0.29 (0.28-0.3)	21.8	14.4
99% air travel restriction					
AVP ^u ; first vaccine dose at 2 months	14.2 (14.1-14.3)	156 (152-158)	0.27 (0.26-0.28)	21.0	14.4
AVP ^u ; first vaccine dose at 3 months	15.8 (15.5-16.3)	129 (127,131)	0.39 (0.35-0.47)	22.9	14.4
AVP ^u ; first vaccine dose at 4 months	23.3 (23.0-23.6)	139 (137-141)	0.77 (0.75-0.79)	33.0	14.4
No air travel restriction					
AVP ^u ; first vaccine dose at 2 months	16.5 (16.1-16.7)	99 (97-101)	0.49 (0.43-0.52)	24.0	14.4
AVP ^u ; first vaccine dose at 3 months	23.7 (23.4-23.8)	109 (107-111)	0.75 (0.75-0.76)	33.6	14.4
AVP ^u ; first vaccine dose at 4 months	24.8 (24.7-24.9)	109 (107-111)	0.75 (0.75-0.76)	35.3	14.4

VE vaccine efficacy, 70% (unless otherwise specified).

Vaccination target categories: Personnel providing essential services, elderly persons, individuals 2-18 years of age (unless otherwise specified).

AVP Antiviral prophylaxis.

* Different vaccine effectiveness for different categories: 59% in individuals 2-18 years of age [79], 70% in individuals 40-64 years of age [36], and 40% in ≥ 65 year-olds [59].

^u Unlimited, household contacts.

† Vaccine effectiveness = 50%.

** Air travel restrictions for 2 months after the first national case.

respectively) [47], could also play a role in explaining this difference.

It is widely accepted that a combination of measures would be necessary to sufficiently control the spread of an influenza pandemic, specifically, vaccination, AVP, social distancing, and air travel restrictions [47, 54]. In our simulations, AVP is confirmed to be

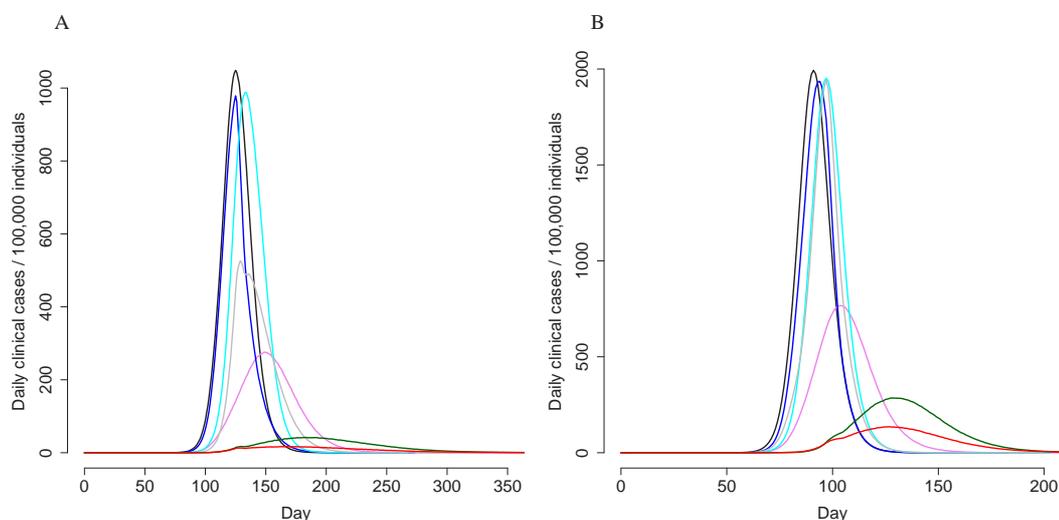


Figure 5.7: Clinical AR, by control measure and scenario (Panel A: $R_0=1.7$; Panel B: $R_0=2$): black = baseline results; light blue = 90% air travel restriction; violet = AVP to household contacts; blue = vaccination, administering first dose within 3 months of the first world case for $R_0=1.7$, or within 2 months for $R_0=2$; grey = 90% air travel restriction + vaccination, as reported for the blue line; green = all control measures combined; red = all control measures combined, extending AVP to school/work close contacts.

the most effective single intervention [54]; however this would require to stockpile a high number of antivirals, to be capable to rapidly identify index cases, to treat a high number of contacts, and to maintain their compliance to a treatment lasting 10 days.

Recent modeling studies have predicted that the use of a pre-pandemic vaccine with a low VE after the first dose (i.e., 30%) would be crucial for pandemic mitigation if the R_0 were 1.7 yet not higher [87]. In our model, we introduced pre-pandemic vaccine for population priming and considered the vaccine to be effective only after the administration of a successive dose of pandemic vaccine, assuming different hypotheses for VE. In particular, we were interested in determining whether variations in VE by age could provide further insight into the impact of control measures. Systematic reviews have shown that the clinical effectiveness of seasonal influenza vaccine varies with age, with a higher VE in adults than in children and the elderly (70% vs. 59% and 40%, respectively) [36, 59].

Our results show that these differences would not substantially affect the cumulative AR. Moreover, vaccinating 2-18 year-olds would reduce by approximately 75% the AR in unvaccinated 30-50-year-old adults, showing a clear herd immunity effect. These results thus support the idea that, during a pandemic, vaccinating children should be a higher

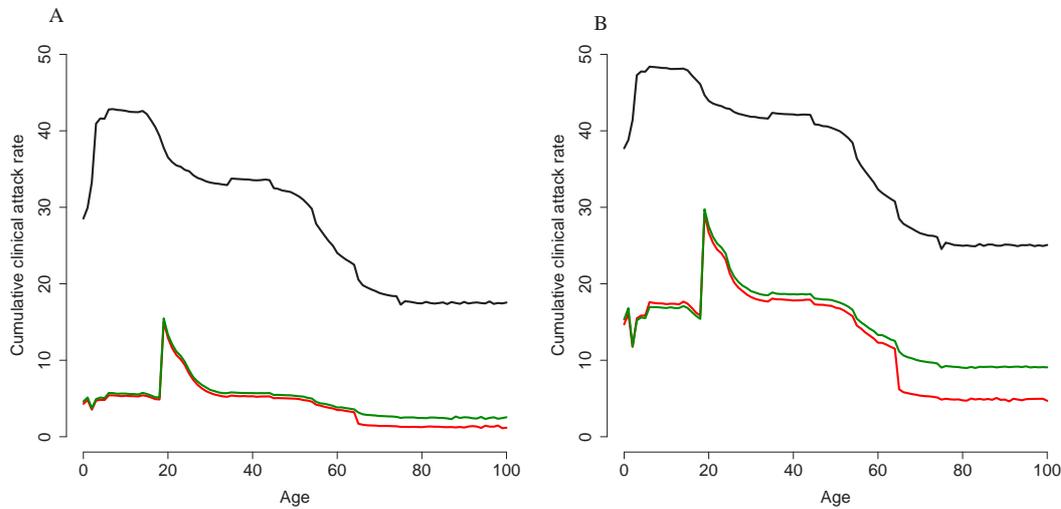


Figure 5.8: Cumulative clinical AR, by age and scenario (Panel A: $R_0=1.7$; Panel B: $R_0=2$). Black line represents baseline results; red line represents the standard vaccination strategy (i.e., personnel providing essential services; elderly persons; and 2-18 year-olds); green line represents the effect of limiting vaccination to essential workers and children.

priority than vaccinating elderly persons [54, 57].

With specific regard to air-travel restrictions, the effectiveness of this measure remains controversial [47, 31, 32, 19, 127, 41, 60, 20]. Our results confirm that international air-travel restrictions can buy about 1 to 3 weeks in delaying the epidemic [47, 32, 20, 41]. In the moderate and severe scenarios, the implementation of 99% air-travel restriction, would allow to gather one-two months of time for administering the vaccine to target population. In detail, if R_0 were 2, starting vaccination within three months of pandemic emergence would have the same impact than starting vaccination within two months of the first world case, with no air-travel restrictions in place.

However, the administration of the first vaccine dose within three months of the first world case would be possible only if vaccines against "high pandemic risk" avian influenza strains (such as A/H5N1) were stockpiled before the pandemic. In any case, because of the antigenic drift of the virus, it is not possible to precisely predict the effectiveness of pre-pandemic vaccines. In this scenario, it is reassuring that a decrease in VE from 70 to 50% would not significantly modify the impact of vaccination.

When using a pre-pandemic vaccine, the maximum reduction in the AR would be achieved by either providing AVP to both household contacts and close contacts in the school/workplace, as shown in a previous work [54] (i.e., 32 million antiviral courses,

covering approximately 56% of the national population), or by vaccinating adults (i.e., 25 million vaccine courses), in addition to the other target categories. In the occurrence of an actual pandemic, the choice of the strategy will be based on several factors which at present are unknown, such as the capacity to produce vaccines, the effectiveness of vaccination and AVP, and logistic constraints in the distribution of vaccines and AVP.

In interpreting the results of this model, some limitations need to be mentioned. The model requires detailed information on the population's characteristics, including age and geographic distribution, the size of households, schools and workplaces, and commuting data. In our study, the source of these data were routinely collected national statistics. The number of students per school and workers per workplace vary in proportion to the resident population in the different geographic areas. However, we assumed that the employment rate was the same throughout Italy, though it is known to vary greatly when comparing northern, central, and southern Italy (4%, 6% and 12%, respectively). Moreover, in modelling the social distancing measures, we only considered the closing of those public workplaces not providing essential services, which could have resulted in an underestimate of the effect of such measures. Furthermore, these workplaces are probably not uniformly distributed throughout Italy.

In the global SEIR model we considered all infected persons to be symptomatic and not traveling; thus we may have overestimated the effect of travel restrictions. By contrast, national data on in-coming flow by land and sea were not easily available, and we therefore did not take into account land and sea importation and control. This could also have overestimated the effect of travel restrictions, since importation via all routes should be considered and eventually reduced. Furthermore, a number of factors, which we did not consider in our analysis, could modify the effects of the delay caused by air-travel restrictions, in particular, seasonality [41], environmental effects, and viral evolution, whereas we assumed that contact, transmission and disease parameters remained constant throughout the pandemic period in Italy. Also, we did not include disease-related mortality, considering that deaths would probably occur at the latter stages of the infectious period and thus would not affect the diffusion of disease.

Despite these limitations, and considering that we cannot predict all aspects of an actual pandemic, this IBM, which is based on country-specific demographic data, could be suitable for the real-time evaluation of measures to be undertaken in the event of the emergence of a new pandemic influenza virus.

5.3 Optimized antivirals administration during a pandemic outbreak

In general, non-pharmaceutical interventions, such as travel restrictions and social distancing measures, might delay the epidemic arrival and peak, while pharmaceutical interventions, such as the use of vaccines and antivirals, might reduce the overall impact of the epidemic. Specifically, antiviral treatment of influenza cases reduces transmissibility and, according to recent results [93, 66], case fatality rates, while post-exposure prophylaxis reduces susceptibility to infection and prevents cases [87].

The World Health Organization suggested that governments stockpile, as part of preparations for the next influenza pandemic, sufficient influenza antiviral drugs to treat approximately 25% of their populations. This recommendation was made with the understanding that the stockpiled drugs would, in the whole, be used for treatment as opposed to significant prophylaxis. Remarkably, however, in many countries the antiviral stockpile is well below the suggested minimum level. For instance, the antivirals stockpiled in Italy are sufficient to treat only 7 million individuals [121], corresponding to the 12% of the population.

Therefore, in this study we face the problem of prioritizing the use of antivirals for treatment of cases as a preventive measure for mitigating the spread of an influenza pandemic as long as a pandemic vaccine is not available. On the other hand, in some countries the antiviral stockpile exceeds the number actually required for the treatment of all cases [121]. Thus, we also search for optimal strategies for prioritizing the use of antivirals for post-exposure prophylactic treatment of close contacts of cases. In this case, however, it should be taken into account, that, once the pandemic is well established, antiviral drugs for prophylaxis should also be provided to high-risk health-care workers and emergency services personnel for the duration of community pandemic outbreaks.

Prioritizing antiviral treatment and prophylaxis

Both treatment and prophylaxis were assumed to start 24 or 48 hours after the clinical onset of symptoms in the index case. Treatment of the index case was assumed to reduce infectiousness by 70% [46, 87, 47, 54, 29], whereas antiviral prophylaxis was assumed to reduce susceptibility to infection by 30%, infectiousness by 70%, and the occurrence of symptomatic disease by 60% [87]. Since it is not realistic that governments will implement prophylaxis without treating index cases first, we consider prophylaxis assuming that antiviral treatment is provided to the index cases. We assumed that 90% of the clinical

cases (corresponding to 45% of infected individuals) are identified and treated and that antiviral prophylaxis is provided to the close contacts, namely household contacts, with a treatment course of 10 days [85]. We assumed that treatment with antivirals is associated with a significant reduction in mortality (70%) [93, 66].

We considered administering antiviral treatment and prophylaxis for the entire epidemic period. Population was divided into three classes, namely children and young adults (2-25 years old, individuals younger than 2 years old are excluded since antivirals can not be administered to them [125]), adults (26-64 years old) and elderly (≥ 65 years old), on the basis of the clinical attack rates by age as resulting from the baseline simulations (Fig. 5.9c), which are consistent with data on attack rates by age classes as reported in [56] for the 1918-19 influenza pandemic. We conducted a systematic simulation study for assessing the effects of targeting the different classes in reducing the number of cases and the excess mortality by minimizing the number of antiviral courses required. To such aim, we consider the number of avoided clinical cases (with respect to the baseline simulations) for each antiviral course as an indicator of efficacy of the different intervention options.

Baseline scenarios

On average, the first Italian case arises 76, 48, 36 and 21 days after the first world case for $R_0 = 1.4$, $R_0 = 1.7$, $R_0 = 2$ and $R_0 = 3$, respectively. Fig. 5.9a shows the stochastic variability in timing of initial case in the baseline scenarios. After the initial highly stochastic phase, the stochasticity decreases over time because of the high number of imported cases over time that, together with long distance travels, contributes to synchronize the local epidemics. Therefore, the simulated epidemics are very stable in terms of parameters as clinical attack rate, peak day and peak daily case incidence. On average, the clinical attack rate is 21.7%, 29.7%, 35.9% and 43.8% for $R_0 = 1.4$, $R_0 = 1.7$, $R_0 = 2$ and $R_0 = 3$ respectively (see first row of Fig. 5.10). The peak day is at 193, 123, 94 and 58 days respectively (see second row of Fig. 5.10) and the peak daily case incidence is 0.44%, 0.96%, 1.59% and 2.85% respectively (see third row of Fig. 5.10). Fig. 5.9b shows the expected pattern of spread for the different transmission scenarios considered.

The time needed from the moment that the vaccine seed virus is available until the first vaccine dose can be used is currently 4 months at best [111]. Other estimates are 6 months at best [128, 112]. Remarkably, according to these estimates the pandemic vaccine will be available in time only in case of a mild epidemic (see Fig. 5.9b). Moreover, the continuous importation of cases make unsuitable all containing strategies based on the isolation and treatment of the first clusters of cases. These findings support the hypothesis that, in

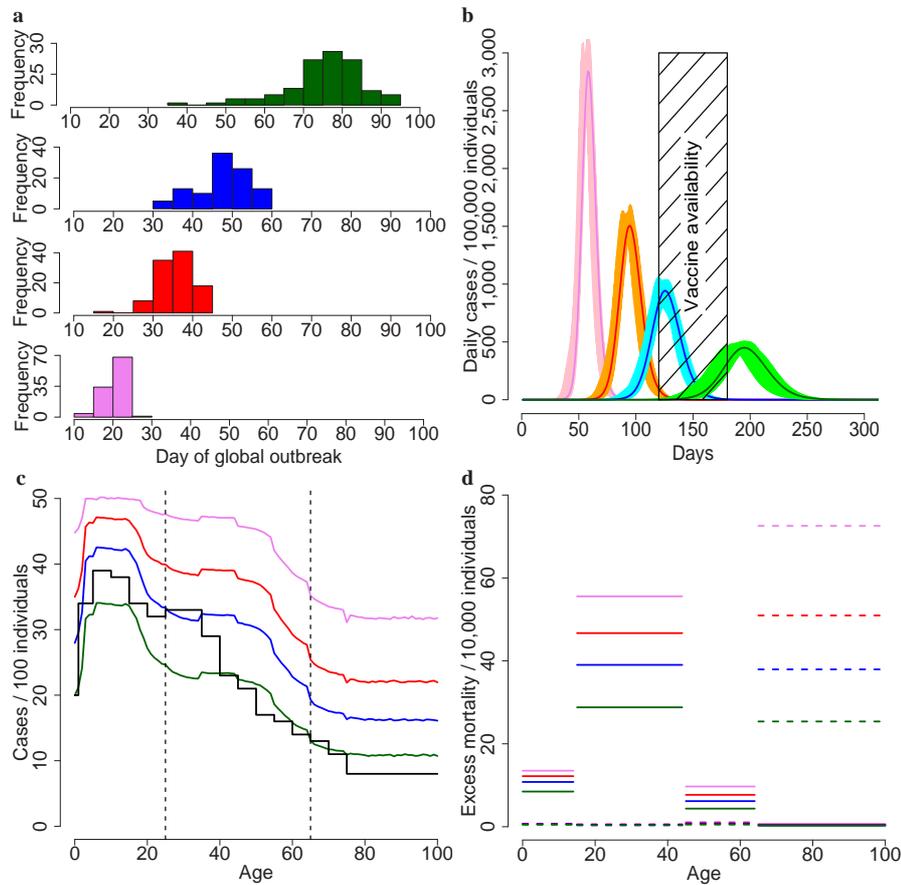


Figure 5.9: Timing for initial case for $R_0 = 1.4$ (green), $R_0 = 1.7$ (blue), $R_0 = 2$ (red) and $R_0 = 3$ (violet) in the baseline scenarios. Histograms are based on 100 simulations each. **b** Expected case incidence over time (solid lines) and 95% confidence intervals (shaded regions) based on 100 simulations for each scenario. Colors as in **a**. The black time window indicate a reasonable time interval for the availability of a pandemic vaccine. **c** Cumulative clinical attack rate by age (colors as in **a**), compared with data on the 1918-19 pandemic [56] (black line). The vertical dashed lines identify the age classes, namely young, adults and elderly, defined for age-prioritization of the use of antivirals. **d** Expected excess mortality by age classes (colors as in **a**) as obtained by assuming two different age-specific case fatality rates, similar to those estimated for the 1918-19 pandemic in Copenhagen (solid lines) and for the 1969-70 pandemic in Italy (dashed lines). Note that in the latter case, the expected excess mortality in the younger age classes (0-64 years old) is very close to 0 for all the R_0 values considered.

large countries, social distancing measures (e.g. school and non essential workplaces closure, case isolation), travel restrictions and pharmaceutical measures based on antiviral treatment of index cases and prophylaxis to close contacts will be key in mitigating and delaying the epidemic as long as the pandemic vaccine is not available. Fig. 5.9d shows the expected age-specific excess mortality in the four considered transmission scenarios and by assuming two different patterns of mortality, namely scenarios *EM1918* and *EM1969*. In the *EM1918* scenario, the excess mortality is estimated to be 14.4/10,000, 19.5/10,000, 23.3/10,000 and 27.8/10,000 for $R_0 = 1.4$, $R_0 = 1.7$, $R_0 = 2$ and $R_0 = 3$ respectively (see first row of Fig. 5.11). In the *EM1969* scenario, it is estimated to be 2.6/10,000, 3.8/10,000, 5.1/10,000 and 7.2/10,000 respectively (see second row of Fig. 5.11).

Age-prioritized use of antivirals: early detection of index cases

We first assume that index cases and close contacts are treated 24 hours after the onset of symptoms in the index cases. If antivirals are used for treatment only, for all age classes, attack rates will decrease to 10.5%, 20.1%, 27.9% and 39.1% (see first row of Fig. 5.10), requiring an antiviral stockpile for treating 5, 10, 14 and 20 million individuals (corresponding to the 9.4%, 17.8%, 24.7% and 34.6% of the population, see first row of Fig. 5.12), for $R_0 = 1.4$, $R_0 = 1.7$, $R_0 = 2$ and $R_0 = 3$ respectively. Moreover, the epidemic peak is slightly delayed (of about 37, 15, 8 and 3 days respectively, see second row of Fig. 5.10) and the peak daily case incidence is greatly reduced (by about 72%, 51%, 39% and 24% respectively, see third row of Fig. 5.10). The number of avoided cases for each antiviral course is 1.2, 0.54, 0.32 and 0.14 respectively (see second row of Fig. 5.12). The excess mortality is greatly reduced by assuming age-specific case fatality rates as those estimated for both the 1918-19 and 1969-70 pandemics. In the *EM1918* scenario, the excess mortality is reduced by 69%, 56%, 50% and 42% for $R_0 = 1.4$, $R_0 = 1.7$, $R_0 = 2$ and $R_0 = 3$ respectively (see first row of Fig. 5.11). In the *EM1969* scenario, the excess mortality is reduced by 75%, 65%, 58% and 50% respectively (see second row of Fig. 5.11).

Treatment of elderly does not lead to any significant reduction of the cumulative number of cases, while the effects of treating younger population and adults are similar to those observed when treatment is considered for all age classes. This means that treatment of elderly has a poor effect in reducing the cumulative attack rate. In fact, when treating only the elderly, the number of avoided cases for each antiviral course decreases by 68%, 56%, 47% and 36% for $R_0 = 1.4$, $R_0 = 1.7$, $R_0 = 2$ and $R_0 = 3$ respectively. The number of avoided cases for each antiviral course is similar when treatment is considered only for

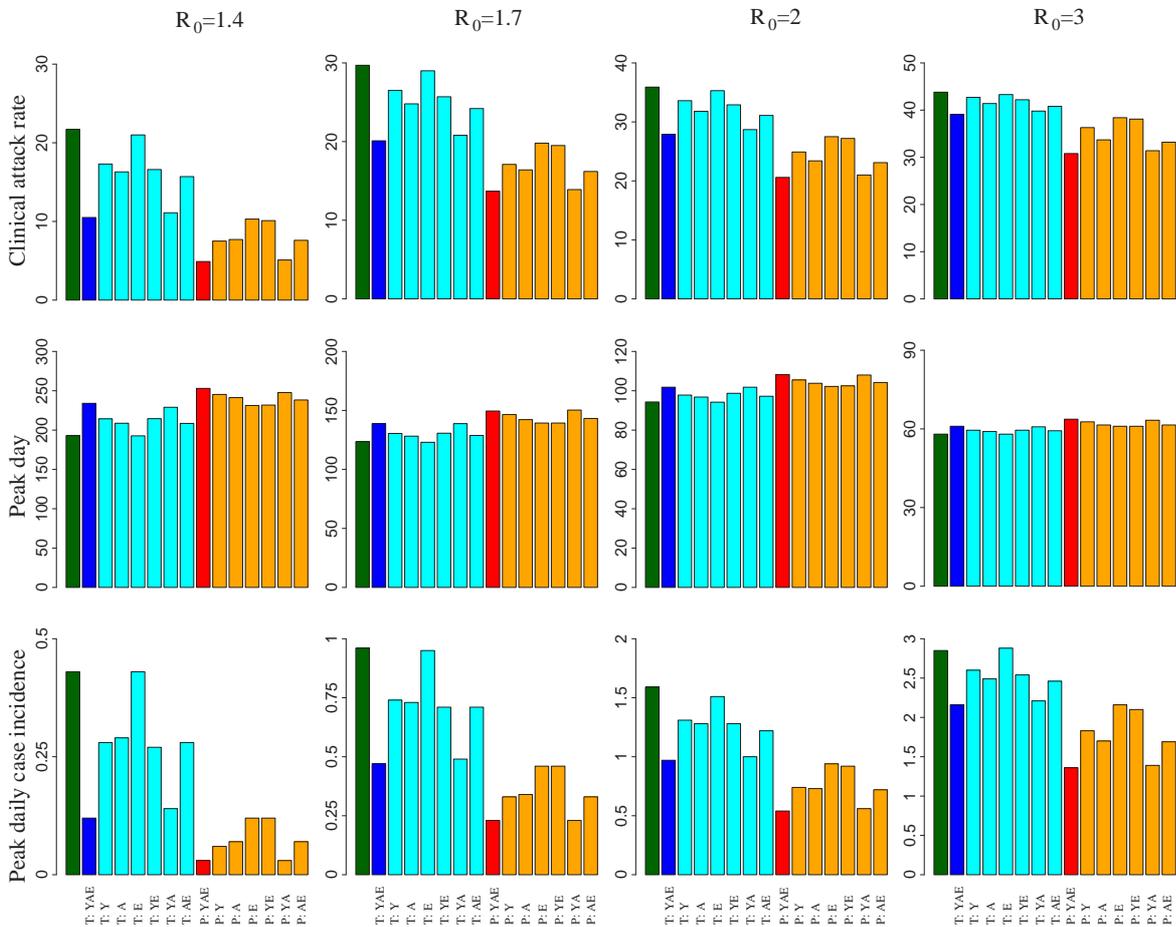


Figure 5.10: Clinical attack rates, peak day and peak daily clinical attack rate for baseline simulations (green), for antiviral treatment provided to index cases of all age classes (blue) or provided only to specific age classes (cyan, Y=young, A=adults, E=elderly), and for post-exposure prophylactic treatment provided to all age classes (red) or only to specific age classes (orange, Y=young, A=adults, E=elderly). When post-exposure prophylactic treatment is considered, we assume that antiviral treatment is also provided to index cases.

young or adult individuals (see last row of Fig. 5.12).

Results can be different when considering the effects on excess mortality (see Fig. 5.11). By assuming age-specific case fatality rates similar to those estimated for the 1918-19 influenza pandemic in Copenhagen, treatment of adults is the much more effective (the excess mortality decreases by 41.2%-28% for R_0 in 1.4-3) than treatment of elderly population (the excess mortality decreases only by 2.8%-0.8% for R_0 in 1.4-3). The opposite

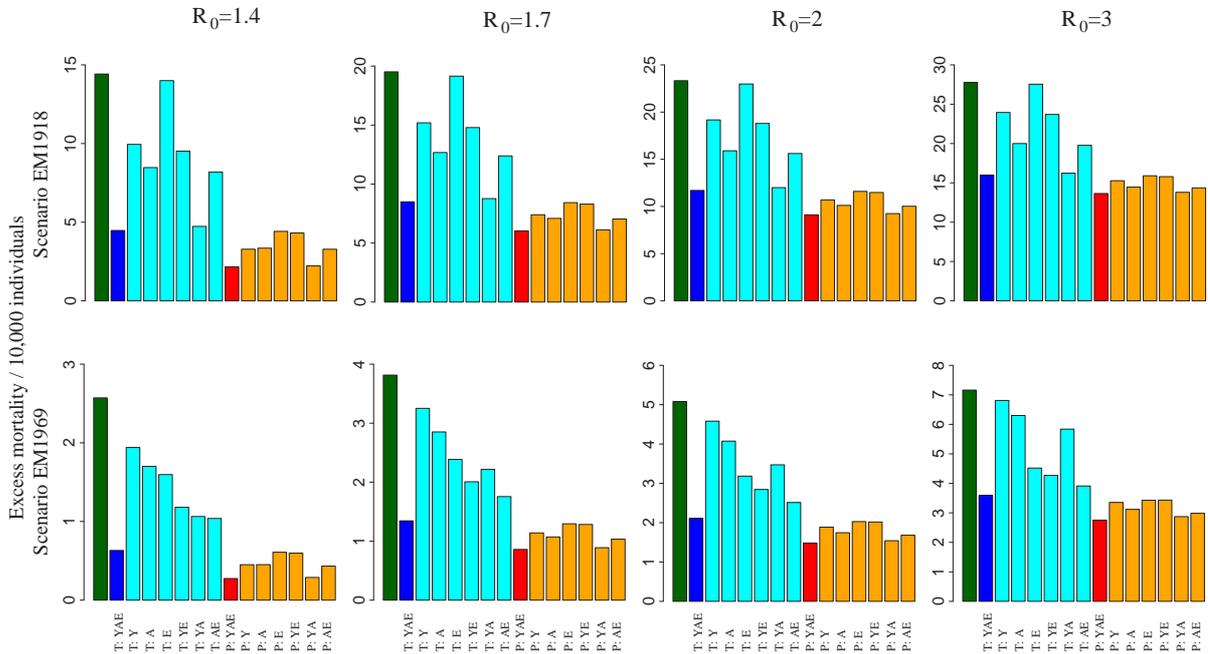


Figure 5.11: Expected excess mortality as obtained by assuming two different age-specific case fatality rates, similar to those estimated for the 1918-19 influenza pandemic in Copenhagen and for the 1969-70 influenza pandemic in Italy respectively, for baseline simulations (green), for antiviral treatment provided to index cases of all age classes (blue) or provided only to specific age classes (cyan, Y=young, A=adults, E=elderly), and for post-exposure prophylactic treatment provided to all age classes (red) or only to specific age classes (orange, Y=young, A=adults, E=elderly).

pattern is observed by assuming age-specific case fatality rates similar to those estimated for the 1969-70 influenza pandemic in Italy: treatment of adults is the much less effective (the excess mortality decreases by 33.8%-12% for R_0 in 1.4-3) than treatment of elderly (the excess mortality decreases by 37.9%-36.9% for R_0 in 1.4-3).

When prophylaxis is provided to close contacts of index cases, the clinical attack rates decrease to 4.9%, 13.7%, 20.6% and 30.8% (see first row of Fig. 5.10), but a larger antiviral stockpile is required (sufficient to treat 8, 21, 31 and 42 million individuals, corresponding to the 13.8%, 37.2%, 53.8% and 73.8% of the population, see first row of Fig. 5.12), for $R_0 = 1.4$, $R_0 = 1.7$, $R_0 = 2$ and $R_0 = 3$ respectively. Moreover, the epidemic peak is significantly delayed (of about 60, 26, 14 and 6 days respectively, see second row of Fig. 5.10) and the peak daily case incidence decreases (approximately by 93%, 76%, 66% and 52% respectively, see third row of Fig. 5.10), with respect to the

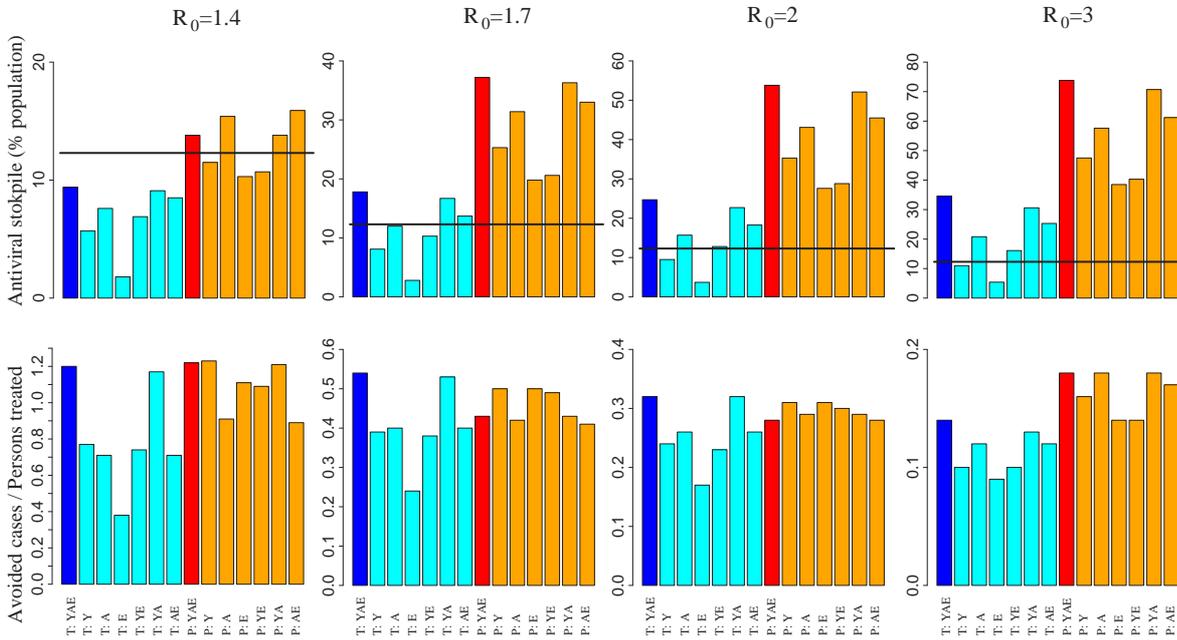


Figure 5.12: Antiviral stockpile required and number of avoided cases divided by the number of persons treated for antiviral treatment provided to index cases of all age classes (blue) or provided only to specific age classes (cyan, Y=young, A=adults, E=elderly), and for post-exposure prophylactic treatment provided to all age classes (red) or only to specific age classes (orange, Y=young, A=adults, E=elderly). When post-exposure prophylactic treatment is considered, we assume that antiviral treatment is also provided to index cases. The horizontal black line represents the Italian antiviral stockpile.

baseline scenarios. The number of avoided cases for each antiviral course is similar to that observed for antiviral treatment, namely 1.22, 0.43, 0.28 and 0.18, respectively for the four transmission scenarios considered (see second row of Fig. 5.12). By assuming age-specific case fatality rates similar to those estimated for the 1918-19 influenza pandemic in Copenhagen, the excess mortality decreases by 51.6%, 29%, 22.1%, and 14.7% with respect to treatment of all cases for $R_0 = 1.4$, $R_0 = 1.7$, $R_0 = 2$ and $R_0 = 3$ respectively (see first row of Fig. 5.11). The excess mortality decreases even more by assuming age-specific case fatality rates similar to those estimated for the 1969-70 influenza pandemic in Italy, namely 56.5%, 36%, 29.7%, and 23.2% (see second row of Fig. 5.11).

Providing prophylaxis only to individuals in some age classes results in the same patterns observed above for the age-prioritized treatment of index cases (see Fig. 5.10 and Fig. 5.12). Age-prioritized prophylaxis does not result in a significant reduction of the excess mortality with respect to the treatment of all cases (see Fig. 5.11). In fact,

when considering prophylaxis to close contacts of cases we are assuming that treatment is first provided to all index cases.

Prophylaxis provided to younger individuals is the only intervention allowing a relevant reduction of the cumulative clinical attack rates (they decrease to 7.5%, 17.1%, 24.9% and 36.3%, respectively for the four transmission scenarios considered) with a significant reduction of the antiviral stockpile required (sufficient to treat 7, 14, 20 and 27 million individuals, corresponding to the 11.5%, 25.3%, 35.3% and 47.5% of the population), at least when R_0 is no much larger than 2 (see Fig. 5.10 and 5.12).

Age-prioritized use of antivirals: late detection of index cases

By assuming that index cases and close contacts are treated 48 hours after the onset of symptoms in the index cases. It is worth noticing that this delay results in a dramatic decrease of the intervention efficacy and, in general, a larger number of antivirals stockpiled is required and a lower decrease of the clinical attack rate is observed. When treatment is considered for all index cases, the clinical attack rate decreases to 14.9%, 23.9% and 31.1% for $R_0 = 1.4$, $R_0 = 1.7$ and $R_0 = 2$ respectively and the number of avoided cases divided by the number of persons treated decreases to 0.51, 0.27 and 0.18. When prophylaxis is also considered, the clinical attack rate decreases to 9.1%, 17.9% and 25.1% for $R_0 = 1.4$, $R_0 = 1.7$ and $R_0 = 2$ respectively and the number of avoided cases divided by the number of persons treated decreases to 0.55, 0.27 and 0.19. Even worst efficacies are observed when $R_0 = 3$.

Realizations and results variability

Results presented in this section were obtained by averaging over 15 simulations for each transmission scenario considered (but for the baseline simulations which were based on 100 simulations). This certainly represents a number large enough to guarantee the stability of the results. Specifically, only the timing of the initial cases is highly variable (however, this is due to the high stochasticity of the epidemic in its initial phase). On the contrary, the epidemiological indicators depending on the whole course of the epidemic are very stable: standard deviations are less than 0.02% of the population for the cumulative attack rates, less than 6 days for the peak day and less than 0.04% of the population for the peak daily case incidence.

5.3.1 Discussion

A recent study conducted in Italy [29] has shown that the use of antivirals, for treatment of index cases and post-exposure prophylactic treatment of household contacts, is the most effective single intervention strategy, resulting in a relevant reduction of the cumulative clinical attack rate, namely of 78%, 50% and 36%, for $R_0 = 1.4$, $R_0 = 1.7$ and $R_0 = 2.1$, respectively. In addition, their use contributes to delay the epidemic peak and to decrease the peak daily case incidence. Similar results have been shown for UK and US [47, 54]. Moreover, school and workplace prophylaxis could dramatically increase the impact, in terms of reduction of the clinical attack rate [47, 29].

However, critically, the antiviral stockpile required is relevant. In Italy [29], an antiviral stockpile large enough to treat 20 to 30 million of individuals (corresponding to the 35% and 53% of the population) is needed for R_0 in $1.7 - 2$. For $R_0 = 1.7$, the antiviral stockpile required decreases to 17% of the population only with the availability of a vaccine within 4 months after the first world case and by considering large scale social distancing measures (e.g., 90% air travel restriction and school closure for 2 months). For $R_0 = 2.1$, the antiviral stockpile required is about 35% of the population. Similar results were obtained in US [47], where the antiviral stockpile required ranges from 25% to 60% of the population, depending on the transmission scenario and the different mitigation measures considered (school/workplace prophylaxis excluded).

We conducted a systematic simulation study of the age-prioritized use of antivirals for mitigating and delaying an influenza pandemic. By assuming R_0 no much larger than 2, our results confirm that the antiviral stockpile required for the treatment of cases ranges from 10% to 25% on the basis of the transmission scenario considered. If $R_0 = 3$, the stockpile required for the treatment of cases increases to 35% of the population. Treatment of index cases is effective in mitigating the epidemic (decrease of cumulative attack rate ranges from 11% to 52% in the four considered transmission scenarios, decrease of peak daily case incidence ranges from 24% to 72%). By assuming that treatment with antivirals is associated with a significant reduction in mortality (70%), a large decrease in the excess mortality is observed in all the transmission scenarios considered (ranging from 42% to 75%).

No suboptimal strategies, based on the treatment of a fraction of cases on an age basis, were found able to remarkably reduce both the clinical attack rate and the antiviral stockpile required. Remarkably, however, a significant reduction of the excess mortality can be achieved by treating only a specific fraction of the population, depending on age-specific case fatality rates: treatment of adults is more effective if age-specific case fatality

rates are similar to those estimated for the 1918-19 influenza pandemic in Copenhagen while treatment of elderly is more effective if age-specific case fatality rates are similar to those estimated for the 1969-70 influenza pandemic in Italy. Therefore, early estimates of age-specific cases fatality rates can be crucial for optimizing the use of antivirals during an influenza pandemic. Moreover, we have shown that treatment of elderly does not lead to any significant reduction of the cumulative attack rate and that the efficacy of treating younger population and adults are similar, but with a different cost in terms of antiviral doses required.

Treatment provided to all cases coupled to prophylaxis for younger individuals is the only intervention allowing a significant reduction of the cumulative clinical attack rate with a significant reduction of antiviral courses required, with respect to provide prophylaxis to the all close contacts of cases. To implement this strategy, the antiviral stockpile should be large enough to treat about the 12%, 25%, 35% and 47.% of the population, for $R_0 = 1.4$, $R_0 = 1.7$, $R_0 = 2$ and $R_0 = 3$ respectively. Since the antivirals stockpiled in Italy are sufficient to treat only about 7 millions individuals, corresponding to the 12% of the population, Italy seems to be able to mitigate an influenza pandemic only at the very beginning of the outbreak. However, the implementation of social distancing measures (e.g. isolation of index cases and school/workplace closure), travel restrictions could slow down the spread of the epidemic. Consequently, the antiviral stockpile required could be significantly lower than that predicted by our model and time could be gained for pandemic vaccine production and distribution, at least under moderate transmission scenarios.

In our study we did not consider treatment and prophylaxis for specific categories, such as patients admitted to hospital, health care workers with direct patient contact and emergency medical service providers, highest risk patients (young children 12-23 months old, elderly ≥ 65 years old), public safety workers (police, fire, corrections), and government decision makers. These policies are consistent with medical practice and ethics to treat those with serious illness and who are most likely to die and those groups which are critical for an effective public health response to a pandemic (preventing absenteeism and maintaining societal functions). Specific work should be conducted for modeling these interventions in order to refine our estimates. However, our strategies of age prioritization could have important ethical impacts that should be taken into account. Recently, the WHO has developed specific guidelines to take into account ethical considerations in developing a public health response to pandemic influenza [131]. As regards age-based prioritization, it is stated that “the goal of reducing overall disease burden might also provide a rationale for favouring younger persons, even if the fair innings argument is not

accepted”. However, “age-based prioritization criteria should be adopted only after wide public consultation”.

Moreover, the potential impact of resistance of the circulating strain to antiviral drugs should also be considered [10, 40, 103, 5]. The extent of such may cause substantial revision to policies regarding the use of such drugs during the next pandemic. In fact, our results should also consider the possibility of the emergence of an antiviral resistant strain as observed in the last two influenza seasons for influenza A(H1N1) strain [51]. The circulation of transmissible oseltamivir-resistant virus may preclude the use of oseltamivir for post-exposure prophylactic treatment of close contacts. However, certain countries have differentiated their stockpile acquiring also zanamivir which is particularly relevant in light of emerging resistance to oseltamivir. This implies that additional antiviral reserve capacity is required and this is likely to come primarily from zanamivir [51].

Our study also highlights the importance of the early detection of cases. In fact, great effort should be made in order to establish a surveillance system able to detect and treat cases as soon as possible since a delay of more than 24 hours could make both antiviral treatment and prophylaxis very inefficient. This means that, to be successful, preparedness to pandemic should not be only stockpiling of antiviral courses. A great effort should also be made in organizing antiviral distribution and implementing specificity and sensitivity of existing surveillance systems for seasonal influenza, in order to detect cases as soon as possible in the occurrence of the emergence of an influenza pandemic. Antiviral drugs must be given early in the course of infection to reduce symptoms (maximum 48 hours) and before any prospect of knowing the sensitivity of the virus [9, 67]. Viral loads begin to decrease 24-48 hours after the onset of symptoms and late antiviral therapy is unhelpful [13]. This critical aspect may have important implications on infrastructure for care delivery. Since health systems may be overwhelmed during a pandemic, new care services for providing the usual health care services (such as drug delivery in hospital or in pharmacies or directly at home) should be considered in order to timely distribute antivirals to cases and close contacts. Also, monitoring systems able to detect adverse events should be considered. However, these aspects are directly related to the organization of the health care system, and should be tailored on the basis of the different resources available.

A characteristic feature of pandemics is to appear in a series of waves. Results presented in this work could be considered fairly unrealistic if waves were determined by virus mutations resulting in the elimination (even partial) of acquired immunity in the population. In fact, a much larger cumulative attack rate would be expected during a series of waves in which acquired immunity is lost at the end of each wave. On the contrary,

no substantial differences, but for the timing of the epidemic spread, would be expected if waves were determined by factors that do not contribute to increase the effective reproductive number (e.g. school closure in the Summer period or spontaneous behavioral changes of the population in response to the epidemic [113]).

5.3.2 Conclusions

Our results strongly suggest that governments stockpile sufficient influenza antiviral drugs to treat approximately 25% of their populations, by assuming that R_0 is not much larger than 2. In fact, no suboptimal strategies, based on the treatment of a fraction of cases on an age basis, were found able to reduce remarkably both the clinical attack rate and the antiviral stockpile required. In countries where the number of antivirals stockpiled is well below 25% of the population, treatment of elderly should be considered as a priority if age-specific case fatality rate were similar to that estimated for the 1969-70 influenza pandemic in Italy, where deaths occurred primarily among elder persons. On the contrary, treatment of adults should be considered as a priority if age-specific case fatality rate were similar to that estimated for the 1918-19 influenza pandemic in Copenhagen, where deaths occurred primarily among adult persons. In countries where the number of antiviral stockpiled exceeds the number required for the treatment of cases, providing prophylaxis only to younger individuals is an option that could be taken into account in the preparedness plans. However, these results are influenced by the timing of cases detection: administration of antivirals 48 hours after the clinical onset of symptoms in the index cases dramatically affects the efficacy of both treatment and prophylaxis.

5.4 Dynamics of an influenza pandemic in Europe

The spread of an infectious disease epidemic is driven by the interplay of two factors: the transmissibility of the virus responsible for the infection and the characteristics of the host population. When the role of host is played by a human population, predicting the spread of an epidemic is a tough problem due to the complexity of modern human societies. It is well established that the spatial structure of the population has an impact on the diffusion of an epidemic: measles waves in England and Wales, spreading from large cities to small towns, are determined by the spatial hierarchy of the host population structure [61], and the spatial distribution of farms influences the regional variability of foot-and-mouth outbreaks in United Kingdom [81]. The heterogeneity of the population itself can play an important role in the spread of an epidemic [38]. It is also well known that human mobility patterns affect the spatiotemporal dynamics of an epidemic: the role played by the airline transportation network has been analyzed in [30], and it has been shown that the high degree of predictability of the worldwide spread of infectious diseases is caused by the strong heterogeneity of the transport network [68].

Large-scale individual-based spatially explicit transmission models of infectious diseases [114] have become a relevant tool to evaluate intervention options for containing [87, 46] or mitigating [85, 47, 54, 29, 64] a new influenza pandemic. Because of their complexity, these models have been developed only at country level, including also some European countries [47, 29].

However, Europe has never been analyzed as a whole and thus it is still uncertain how a new pandemic influenza could spread in Europe. Europe comprises countries characterized by completely different social and economical backgrounds that result in different levels of population heterogeneity, in terms of both sociodemographic structure and mobility. In particular, it is still unclear how differences in the sociodemographic structure, which result in different levels of population heterogeneity, and different patterns of human mobility can affect the spatiotemporal spread of an epidemic. Here we provide quantitative measures of their effects on the impact and the timing of an epidemic at European level.

5.4.1 Population heterogeneity

By analyzing data on the sociodemographic structure of 37 European countries (see Fig. 5.13a) provided by the Statistical Office of the European Communities (Eurostat) and integrated with data provided by the National Statistical Offices for countries not covered by Eurostat, we found that the frequencies of household type and size (Fig. 5.13b-c), the age structure (Fig. 5.13e), the schools size (Fig. 1d), the rates of school attendance and

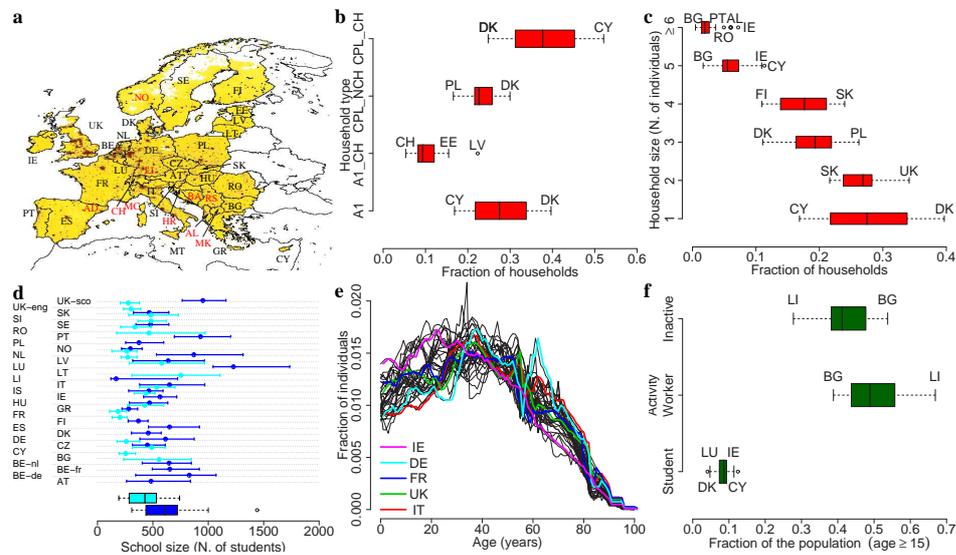


Figure 5.13: Sociodemographic structures. (a) The study area includes 37 countries and accounts for about 515 millions individuals (details are provided in Tab. S1). Colors from yellow to brown indicate increasing density of population. Black labels refer to countries belonging to EU27 while red labels refer to countries which do not belong to EU27. (b) Variability in the frequencies of household type at European level. A1 represents single persons, A1_CH single parents with children, CPL_NCH couples without children, CPL_CH couples with children. More than 95% of European households are structured as one of the four above mentioned types. (c) Variability in the frequencies of household size. (d) Variability in schools size (primary schools in cyan, secondary schools in blue). Horizontal lines identify the percentiles 25 and 75, the points represent the median values. The two boxplots represent the distributions of the average school size in the different countries. (e) Age structure curves in the different countries. (f) Variability in the employment and school attendance rates. Only individuals aged more than 15 are considered. In the model, individuals aged less or equal than 15 are assumed to attend schools.

employment by age (Fig. 5.13f) are highly variable across Europe. The age structure of countries like Ireland, which is one of the youngest European countries (with 31% of the population aged less or equal than 20 years), differs drastically from that of countries like Germany and Italy (where only 22% and 20.5% of the population is aged less or equal than 20 years, respectively), which are characterized by very low fertility rates [123]. This results in largely different frequencies of household type and size. The fraction of households with children ranges from 0.3 in Denmark to 0.6 in Sweden and the average

household size ranges from 2.1 in Denmark to 3 in Cyprus. By restricting our attention to the households with children, a large variability in the number of children per household is also observable (see Fig. S1), with countries as Ireland and Cyprus, where households have several children, opposite to countries like Germany and Bulgaria. We have also observed a large difference in terms of employment rates in the population aged more than 15 years (the legal working age in Europe is 15 or 16, with some exceptions): the fraction of workers ranges from 0.39 in Bulgaria to 0.67 in Lichtenstein. The fraction of students in the population aged more than 15 years ranges from 0.04 in Denmark to 0.12 in Cyprus. According to the PIRLS 2001 and PISA 2000 and 2003 international surveys, as elaborated in [44], the average size of primary schools ranges from 200 to 750 and the average size of secondary schools ranges from 270 to 1000. We used an independent data set providing information on all the Italian schools to validate the surveys data (see Fig. S2). Data concerning workplaces in Italy and United Kingdom do not highlight significant differences in the size of workplaces (see Fig. S3). We used the above described sociodemographic data to generate a highly realistic synthetic population of individuals, explicitly grouped in households, schools and workplaces, for simulating the populations in the different countries of the study area. Since it is reasonable to assume that the epidemiological characteristics of the virus do not vary among the European countries, we should expect that the high variability in the sociodemographic structure of the European countries results in a high variability in the impact of a new influenza pandemic in the different European countries. This is the first key issue we want to address. Details on the analysis of the European sociodemographic structure can be found in [95].

5.4.2 Human mobility

We analyzed air and railway transportation data as provided by Eurostat. We found that in the 2007 more than 360 millions passengers have taken international trips across EU27 (see Fig. 5.13a), 323 millions of whom by airplane and 37 millions by train. In the same year, more than 135 millions passengers have entered EU27 from countries outside EU27. The great majority of these travels are from and to the western part of Europe (see Fig. 5.14a-b), namely United Kingdom, Germany, France, Italy and Spain (about the 85% of the travels are from and to these countries). The probability density function of travel distances is shown in Fig. 5.14c. As shown in [91, 62], international travel flows are related to economic factors. By considering only the travels across EU27, we found that the flow from country i to country j can be explained by a gravity model depending on the GDP (Gross Domestic Product: it is an economic index measuring the national income

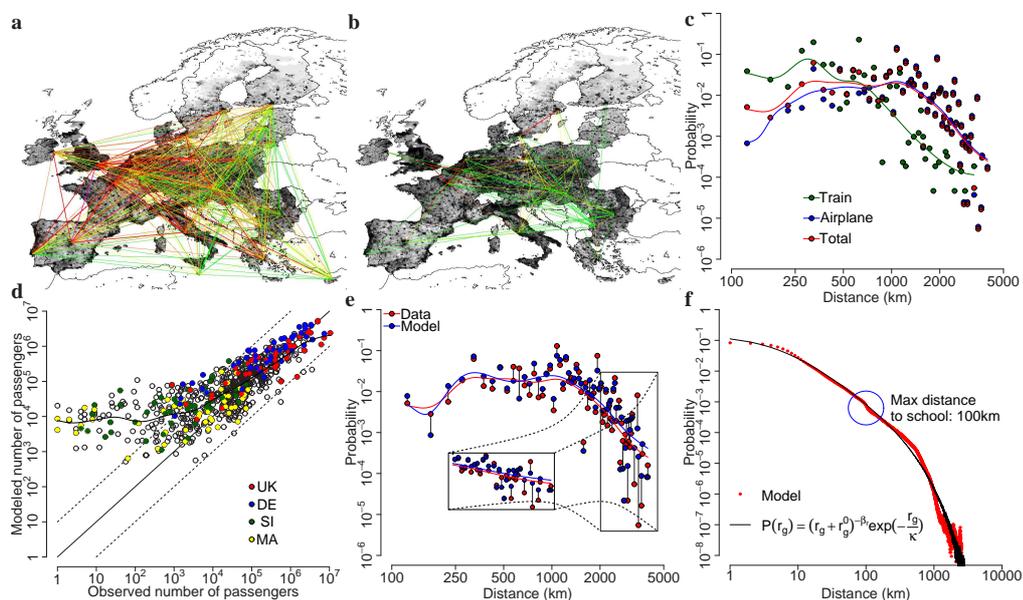


Figure 5.14: Population movement patterns. **(a)** Network of yearly airplane travelers across Europe (colors are defined as follows. Green: less than 10,000 travelers, yellow: 10,000 to 100,000 travelers, orange: 100,000 to 1,000,000 travelers, red: 1,000,000 to 10,000,000 travelers, purple: more than 10,000,000 travelers). Each link between two countries is identified by an arc connecting the two capitals. **(b)** as **(a)** but for train travelers. **(c)** Probability density function of travel distances by train (green points), by airplane (blue points) and total (red points). Solid lines represent smooth interpolations of data. **(d)** Model A) (described in the main text; parameters: $\tau_f = 0.57$, $\tau_t = 0.99$ and $\rho = 0.39$): comparison between the observed and the modeled origin-destination matrix. Points compare generic entries of the two matrix and the solid black line represents a smooth interpolation. The model tends to overestimate the number of travelers when the actual yearly number of travelers is less than 1,000; it is in good agreement with the data on the most important links. **(e)** Model A): resulting probability density function of travel distances compared with that resulting from the analysis of the observed data (shown in **c**) (red points). **(f)** Internal commuting: probability density function of travel distances to school/workplace (in the model, red points), compared with that proposed in [58] for the radius of gyration of mobile phone users (black points). In the model, students are assumed to attend schools no more than 100Km from home. This results in a change in the slope of the probability curve (blue circle).

and output for a country) per capita, the population and the distance: $F_{ij} = \theta \frac{g_j^{\tau_t} g_i^{\tau_f}}{d_{ij}^\rho}$, where g_i is a normalized GDP of country i ($g_i = \frac{G_i}{G^*} p_i$ where G_i is the GDP per capita of country i , G^* is the average GDP per capita of EU27 and p_i is the population of country i) and d_{ij} is the distance between the two countries. τ_f and τ_t tune the dependence of dispersal on donor and recipient sizes and ρ tunes the dependence on the distance, θ is a proportionality constant. To show this, we generated a synthetic population of travelers taking travels according to a gravity model whose masses are given by the normalized GDPs (model A), by the population sizes, as in [126] (model B) and, finally, taking travels by choosing a random destination (model C). We found that model A explains the origin-destination matrix (Fig. 5.14d) and the distance distribution (Fig. 5.14e) better than models B and C (see Fig. S5). This considered, we used model A for simulating long-distance travels across the study area. As for the internal commuting, i.e. daily trips to school and workplace, we adopted the following procedure. First, schools and workplaces of the proper size were spatially-distributed proportionally to the population (see Fig. S2). Afterward, students and workers were randomly assigned to a school or workplace, in such a way that the resulting distance to school/work distribution complies with a truncated power-law distribution (see Fig. 5.14f), as proposed in [58] for the radius of gyration of mobile phone users, extending the precursor work presented in [18] on the circulation of bank notes in the United States of America. Fig. S6 shows how well the proposed model of internal commuting compares with a gravity model previously developed on Italian commuting data. To what extent the observed mobility patterns and their inhomogeneity across the European countries affect the spread of a new influenza pandemic in Europe is the second key issue we want to address. Details on the analysis of human mobility patterns can be found in [95].

5.4.3 Spatiotemporal spread of a pandemic influenza in Europe

The transmission rates, defined as the product of the contacts rate times the probability of transmitting the infection, of a new influenza pandemic are unknown. By looking at past pandemics, we can only make assumptions on its transmissibility potential, which can be summarized by the reproductive number R_0 (essentially, the number of secondary infections that result from a single infectious individual in a fully susceptible population [6]). Therefore, according to recent estimates of the reproductive number for last influenza pandemics [46, 99, 27], plausible transmissibility scenarios on R_0 are drawn: the investigated values range from 1.6 to 2.4. Moreover, according to [47], the model is parametrized so that in the United Kingdom 30% of transmission occurs in households, 37% in schools

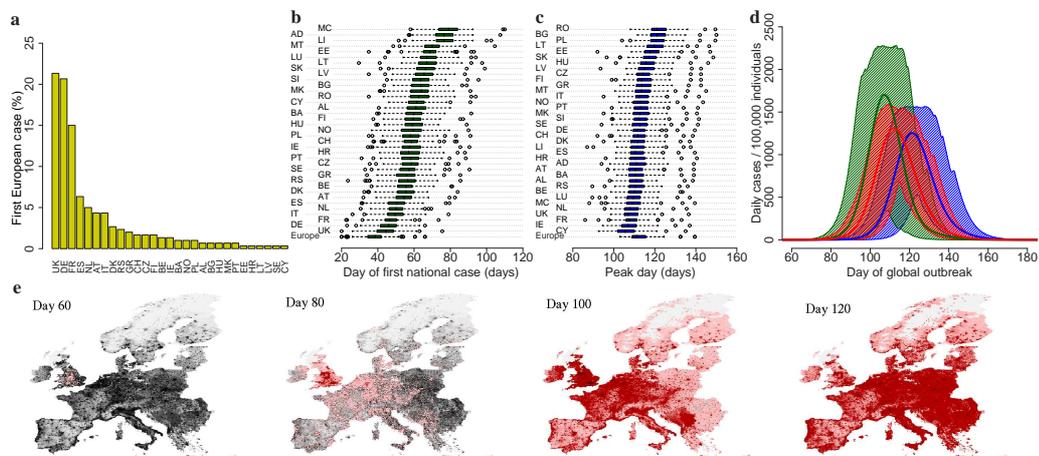


Figure 5.15: Spatiotemporal dynamics of a new pandemic influenza ($R_0 = 2$). (a) Probable destination of the first case imported in Europe. (b) Distributions of the day of the first national case (days since the first world case) in the different countries. (c) Distributions of the peak day (days since the first world case) in the different countries. (d) Expected number of daily cases per 100,000 individuals in time in Europe (red line) and 95% confidence intervals (shaded area). Green and blue lines (and shaded areas) refer to the expected number of daily cases per 100,000 individuals in Ireland and Bulgaria respectively. These two countries are among those where the impact of the epidemic is expected to be the highest and the lowest respectively. (e) Time sequence (in days) of a simulated epidemic. A single simulation with first European case in United Kingdom is shown. Colors from pink to dark red indicate an increasing number of daily cases (dark red indicates more than 10,000 daily cases).

and workplaces and 33% in the general community. Since the contact rates and, consequently, the reproductive number are determined by the sociodemographic structure of the population we are somehow setting the probability of transmitting the infection in the different social contexts. After having parametrized the model in the United Kingdom, the same transmission rates are assumed in the rest of the study area. In what follows, when not differently stated, we are assuming $R_0 = 2$ (as discussed, it means $R_0 = 2$ in the United Kingdom). We assume that the latent period is 1.5 days and the infectious period is 2 days (however we provide a sensitivity analysis for values of this epidemiological parameter in the range 1.5 - 3 days). Infected individuals are assumed to have a probability 0.5 of developing clinical symptoms.

We found that the probability of importing the first case is higher in the western

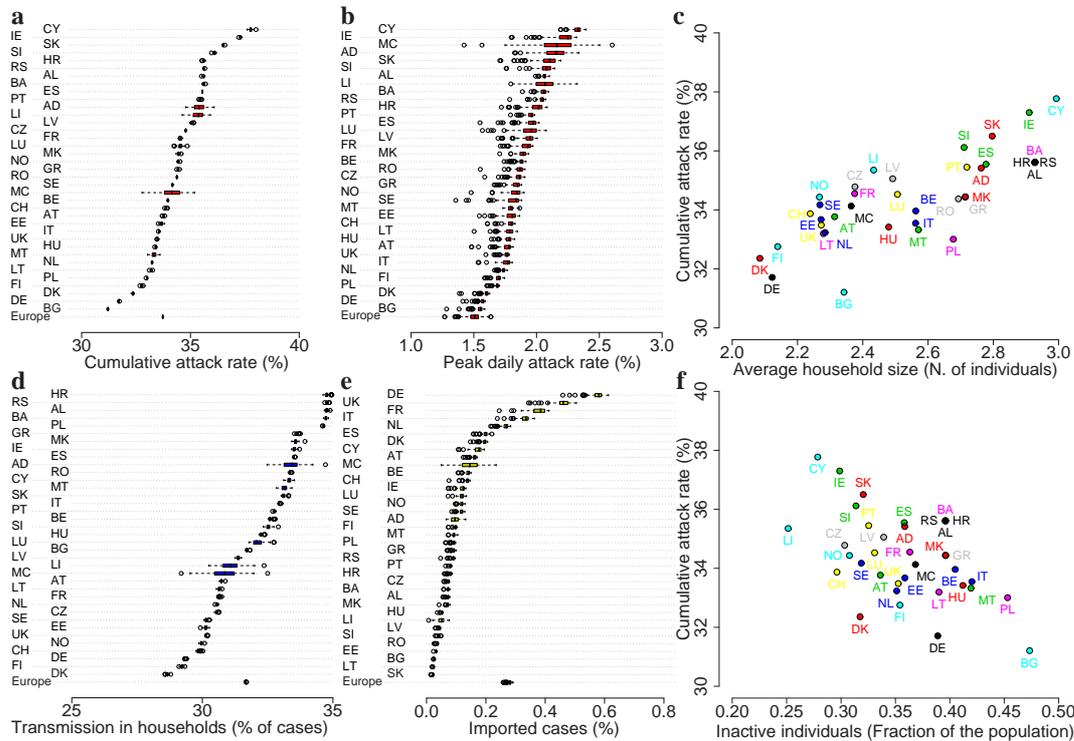


Figure 5.16: Impact of a new pandemic influenza ($R_0 = 2$). (a) Distributions of the cumulative attack rate in the different countries. (b) Distributions of the peak daily attack rate in the different countries. (c) Average cumulative attack rate as a function of the average household size in the different countries. (d) Distributions of the percentage of cases due to transmission among household members in the different countries. (e) Distributions of the percentage of the population infected during long-distance travels across or outside Europe. (f) Average cumulative attack rate as a function of the fraction of inactive (neither students nor workers) individuals in the different countries.

countries (the first case is imported in United Kingdom or Germany in almost 50% of simulations, see Fig. 5.15a). The distributions of the timing of the first case differ largely from country to country (see Fig. 5.15b): in average, the first case occurs 44 and 79 days after the first world case in the United Kingdom and Principality of Monaco respectively. By ignoring the less populous countries, a west-east gradient is clearly observable. The variability in the peak day in the different countries (see Fig. 5.15c) is less remarkable (in average it ranges from 106 days in Cyprus to 122 in Romania) since long-distance travels tend to synchronize the national epidemics and these are much faster in the less populous countries. In general, we have observed that the high mobility inside the countries (in-

ternal commuting) and the long-distance travels tend to synchronize, thus fastening, the epidemic. The average peak day in a country is positively correlated with the longitude of the country (Spearman test, $\rho = 0.55$, $p = 0.0003$), as confirmed by the clear spatial trend observable in the time sequence of the simulated epidemic shown in Fig. 5.15e. This finding is supported by the results presented in [43], where a spatial analysis revealed a significant west-east pattern in the timing of peak influenza activity across Europe for the eight winters since 1999-2000. The average peak day in a country is negatively correlated with the yearly number of passengers entering the country from other countries in the study area (Spearman test, $\rho = -0.59$, $p = 0.001$), supporting the hypothesis that the observed pattern of epidemic spread is related to patterns of human movement. The expected pattern of spread in Europe is shown in Fig. 5.15d. The epidemic peaks some 110 days after the first world case and the epidemic lasts about 3 months. We remark that Fig. 5.15d reports the expected number of new cases in time. Since the epidemics are not synchronized in time, the actual peak incidence will be much higher than the value corresponding to the peak day reported in the figure (it will be closer to the upper 95% confidence limit, see also Fig. 5.16b).

In average, the cumulative attack rate in the different countries ranges from 31.2% in Bulgaria to 37.8% in Cyprus (see Fig. 5.16a). By looking at the study area as a whole, the average cumulative attack rate is 33.7%. Among the most populous countries, the cumulative attack rate is expected to be 31.7% in Germany, 33.4% in the United Kingdom, 33.5% in Italy, 34.5% in France and 35.5% in Spain. It is worth noticing that the value obtained for the United Kingdom is very similar to that obtained in [47]. The standard deviations of the distributions of the national cumulative attack rates are very small, but for the less populous countries. The average cumulative attack rate in a country is positively correlated with the average household size (Spearman test, $\rho = 0.77$, $p < 0.0001$) (see Fig. 5.16c) and with the fraction of students in the population (Spearman test, $\rho = 0.77$, $p < 0.0001$), and negatively correlated with the fraction of inactive individuals in the population (Spearman test, $\rho = -0.38$, $p = 0.02$) (see Fig. 5.16f). It is worth noticing that a simple linear regression model whose independent variables are the average household size, the fraction of students and the fraction of inactive individuals in the population predicts very well the average cumulative attack rate in the different countries (coefficient of determination $R^2 = 0.985$, root mean square error RMSE=0.17). The peak daily attack rate in the different countries is also highly variable. It ranges from 1.5% in Bulgaria to 2.3% in Cyprus (see Fig. 5.16b). Since the national epidemics are not synchronized, the average peak daily attack rate of the whole study is similar to the value observed in Bulgaria, namely 1.5%. In the United Kingdom we obtained a lower peak daily

attack rate (1.8%) with respect to the 2.1% as reported in [47]. This is due to different modeling choices for the infective period. We assume an exponential distribution for both latent and infectious period (as in classical mathematical models of infectious diseases) and infectiousness is assumed to be constant during the infectious period (2 days). In [47] individuals transmit more at the very beginning of the infectious period, giving rise to faster simulated epidemics and to higher peak daily attack rates. These different modeling choices can affect the evaluation of some containment strategies (e.g. antiviral treatment) and can lead to differences in the timing of the simulated epidemics but do not affect the results presented in this work. Fig. S7 shows the dependence of the timing of the epidemic and of the peak daily incidence on the infectious period: given R_0 , the shorter the infectious period is, the faster the epidemic spreads and the higher the peak daily attack rate is. The cumulative attack rate does not depend on the infectious period. Values of peak day and peak daily attack rate in the United Kingdom similar to that reported in [47] were obtained by assuming an infectious period of 1.5 days. We found that the peak daily attack rate in a country is positively correlated with the average household size (Spearman test, $\rho = 0.72$, $p < 0.0001$) and with the fraction of students in the total population (Spearman test, $\rho = 0.79$, $p < 0.0001$), and slightly negatively correlated with the fraction of inactive individuals in the total population (Spearman test, $\rho = -0.31$, $p = 0.06$). Moreover, it is negatively correlated with the number of inhabitants (Spearman test, $\rho = -0.51$, $p = 0.001$) since the national epidemics tend to be less spatially synchronized in the larger countries. It is also relevant to analyze where transmission occurs. We found that the transmission in households in the different countries ranges from 28.6% in Denmark to 34.8% in Croatia (see Fig. 5.16d) and that the fraction of the population contracting the infection in foreign countries ranges from 0.02% in Slovakia to 0.57% in Germany (see Fig. 5.16e). These last results suggest that the efficacy of some targeted interventions, e.g. post-exposure prophylactic antiviral treatment [47, 29] and travel restrictions [31], could be largely different from country to country. By examining the results for values of R_0 in the range 1.6-2.4 we did not find significant qualitative differences (of course, the timing and the impact of the epidemics are drastically different).

5.4.4 Conclusions

The transmissibility of a new influenza virus is uncertain. At the time of writing, preliminary estimates of R_0 for the ongoing A(H1N1) outbreak in Mexico are available [52] but no estimates are available for Europe (the low number of secondary cases in Europe, due to active surveillance of cases, is not sufficient to obtain reliable estimates). As predicted

by our model, however, the number of imported and secondary cases in the initial phase of the epidemic is much larger in the Western side of Europe [130]. Reliable data on the proportion of transmission in social contexts, crucial to the disease transmission, as prisons, leisure places, public transportation systems, hospitals are not available, though some research works are contributing to fill the gap [105, 133]. Spontaneous behavioral changes in the population, as a protective response to a (possibly lethal) epidemic, could affect the spread of the epidemic [113], as well as the imminent school closure in the Summer period. Despite these limitations, our results clearly show that, once the infection will be well established in Europe, European countries have to be prepared to face a fast spread of the epidemic because of the high mobility of the population, resulting in an early importation of the first cases from abroad and in a high synchronization of the local epidemics. The impact of the epidemic is different among the European countries. Specifically, countries as Ireland would have to face more severe epidemic than countries as Germany and Bulgaria because of their sociodemographic structure, characterized by large household groups and by a large fraction of students in the population. Our results are supported by the findings presented in [105], where the authors analyzed social contacts and mixing patterns in eight European countries. Specifically, they found that living in a larger household size was associated with higher number of reported contacts. Moreover, they found that the dominant feature of the contact matrix data is the strong diagonal element: individuals in all age groups tend to mix assortatively (i.e., preferentially with others of similar age) and this pattern is most pronounced in those aged 5-24 years, i.e. the scholar age. They also found that 58% of all reported contacts occur at home, at work, or at school. This results supports our assumption on the proportion of transmission in the different social contexts (in the model, 67% of transmission occurs in households, schools and workplaces, at least in the United Kingdom). These results should have to be take into account for planning strategies for mitigating future pandemics and for controlling the ongoing A(H1N1) influenza outbreak.

5.5 Controlling hepatitis A in Italy

Hepatitis A virus (HAV) is the cause of viral hepatitis A infection, which results in an acute form of hepatitis. Patients recover completely and develop full immunity against future HAV infections [122]. Nevertheless, viral hepatitis A represents an important public health issue, imposing a remarkable economic burden worldwide [15, 34] and in Italy as well [90].

HAV is transmitted via the fecal–oral route by person–to–person contact (direct transmission) or by ingestion of contaminated food or water (indirect transmission). Indirect transmission represents the most important source of infection in countries with high living standards [49, 89]. Specifically, raw mussels and shellfish consumption represents the main source of infection in Italy [94], especially in the most affected regions: Puglia and Campania [24, 25, 118]. Another significant source of infection is represented by travels to high endemicity areas [94], where both direct and indirect transmission can occur.

Nowadays, even in the absence of vaccination, hepatitis A is in a decaying phase, mainly determined by improved hygienic conditions derived from economic development and higher standards of living [77, 78]. Although this is true also for Italy as a whole, as documented by both notification [94, 72] and serological data [33], Southern Italy shows a different pattern. HAV infections are still common in Puglia and Campania, two regions in Southern Italy, where a very large outbreak was observed in 1996–1997 [90], despite the improvement of socio–economic conditions.

Thanks to a new generation of mathematical modeling tools, the effectiveness of both pharmaceutical (e.g., treatment or prophylaxis on a contact tracing basis) and non-pharmaceutical (e.g., social distancing) individually targeted intervention measures can be thoroughly investigated. Recently, highly detailed individual–based models have been developed for evaluating the effectiveness of control measures for diseases such as pandemic influenza [54, 46, 47, 87, 85, 29] or fighting back a bioterroristic attack (e.g., by employing smallpox virus) [42, 86, 115, 65]. As highlighted by Riley [114], the need arises “to develop a simple model of household demographics, so that these large–scale models can be extended to the investigation of long–time scale human pathogens”. In fact, endemic diseases such as tuberculosis, measles, malaria, hepatitis and HIV represent important public health issues worldwide [106].

In this work we introduce an individual–based model for investigating the dynamics of viral hepatitis A in the most affected Italian areas, and for evaluating the effectiveness of intervention options. The model is based on the real sociodemographic composition of the population and it takes into account the vital dynamics of the population and the

changes in the network of contacts (e.g., due to the birth and death of individuals, the generation of new households, the educational career of individuals).

Since hepatitis A is a vaccine-preventable disease, we have evaluated the effectiveness of different vaccination programmes. Moreover, we have analyzed the effects of improvements in standards of living and hygiene and of social distancing measures, such as isolation of symptomatic cases and closure of day care centers and kindergartens.

Results show that a very low vaccination coverage is sufficient to control hepatitis A in Italy. The above described social distancing measures do not have any positive effect in decreasing the seroprevalence. Moreover, an undesired effect of the considered social distancing measures is to decrease the fraction of children and adolescents contracting the infection while increasing that of adults and elderly people. Improvements in hygienic conditions can remarkably decrease the seroprevalence. Finally, the elimination of hepatitis A appears unfeasible since new cases are continuously imported from high endemicity areas outside the country.

Intervention measures

Since hepatitis A is a vaccine-preventable disease, we will evaluate the effectiveness of different vaccination programmes. Moreover, we will evaluate the effectiveness of social distancing measures. In fact, this is possible mostly thanks to the explicit representation of the individuals and of the places where transmission can occur. Finally, since hepatitis A is undergoing a global decline, due to the worldwide improvements in standards of living and hygiene [77, 78], we will draw different scenarios on possible hygienic improvements.

Vaccination For hepatitis A, the target population consists of children and adolescents. We consider newborns and 12 years old adolescents as the target population. Vaccination is modeled by reducing the proportion of susceptible individuals in the target population. This proportion depends on vaccination coverage (i.e., the proportion of target population that is covered in the vaccination campaign) and vaccine effectiveness (i.e., the probability of developing immunity after the administration of a vaccine dose). The latter is kept fixed at 98% according to [12]. Since vaccination coverage depends on the vaccination programme implemented by public health agencies and on the collaboration of the population, many scenarios are evaluated.

Social distancing The aim of the considered social distancing measures is to interrupt the chains of cases observed in day care centers and kindergartens, which are typical of many childhood diseases [53], and viral hepatitis A as well [26]. This is possible

thanks to the explicit representation of the individuals and of the places where person-to-person transmission can occur. Specifically, we investigate the effectiveness of the following strategies:

- d1* symptomatic individuals are assumed to be isolated for two weeks. In this period they do not transmit the infection neither in their household nor in their school or workplace, if any;
- d2* as *d1*. Moreover, day care centers and kindergartens attended by symptomatic individuals are closed for 2 weeks;
- d3* as *d2*. Moreover, the policy is extended to day care centers and kindergartens attended by other household members of symptomatic individuals.

The aim of strategy *d2* is to limit as much as possible the transmission within day care centers and kindergartens, which are the most important routes of person-to-person HAV transmission [83]. Isolation of symptomatic cases is not sufficient to interrupt person-to-person transmission within day care centers and kindergartens, since the probability of being symptomatic in individuals aged 0–6 is very low (about 4% [122, 49]). In this respect, the closure of day care centers and kindergartens is required.

The aim of strategy *d3* is to limit the transmission within day care centers and kindergartens by suspected infectious individuals, i.e., individuals living in the same household of a symptomatic individual.

Since isolation of symptomatic individuals will hardly occur immediately after the onset of symptoms, we assume a one-week delay between the onset of symptoms and the application of such strategies. We investigate the effectiveness of such strategies when applied to a fraction of the symptomatic individuals or only to notified cases.

Let us note, however, that symptomatic individuals continue to excrete the virus in the environment even if social distancing measures are applied.

Modeling hygienic improvements In Italy, the main source of contagion is represented by the consumption of infected raw mussels and shellfish [94]. HAV infection of mussels and shellfish can occur both in the marine environment and in fish market stands where infected water is often employed to wash the shellfish [120]. Thus, we assume that hygienic improvements would be mainly related to the indirect contacts component. Such improvements can be reasonably expected to occur in the considered areas.

We model the improvement of hygienic conditions simply as a reduction of the quantity of infected seafood. In particular, after solving Equation 4.10, we decrease the value of

Table 5.11: *Epidemiological parameters.*

Parameter	Unit	Value	Reference
Reproductive number			
for Campania region	dimensionless	2.9	[1]
for Puglia region	dimensionless	3.8	[1]
Latent period	weeks	2	[23, 122]
Infectious period	weeks	3	[23, 122]
Decay of survival of HAV in the environment	weeks ⁻¹	0.0833	[1, 16, 92]
Probability of becoming symptomatic			
for individuals aged 0–6	percentage	4	[49, 122]
for individuals aged 7–16	percentage	16	[49, 122]
for individuals older than 16	percentage	80	[49, 122]
Vaccine efficacy	percentage	98	[12]

$U(\tau)$ by a reduction factor accounting for the improved conditions (e.g., better hygiene in the fish markets or better awareness in the choice of fishing areas).

Model parametrization

The basic reproductive number R_0 is the average number of secondary cases generated by an infectious individual in a completely susceptible population [6]. In the absence of any pharmacological interventions, for epidemic outbreaks R_0 can be estimated from notification data during the initial exponential growth phase of the epidemic [28], while for endemic diseases it can be defined as the inverse of the fraction of susceptible individuals at the endemic equilibrium [6]. The latter approach was employed for determining the basic reproductive number of the simulated epidemics. Note that the method can not be applied when pharmacological interventions (e.g. vaccination) are implemented.

As regards hepatitis A in Italy, R_0 was estimated to be 2.9 in Campania and 3.8 in Puglia¹.

The free parameters of the model are the transmission rates associated to the different sources of infection (i.e. β_h , β_p^θ , $\beta_s(a)$ and $\beta_t(a)$), while all the other epidemiological parameters (e.g., the latent period and the infectious period) are kept fixed according to the literature (see Table 5.11). We set the transmission rates in such a way that the fraction of cases generated by each source of infection complies with the risk factors by age as reported in [94]. In particular, as regards the direct contacts, only transmission within household, day care, kindergarten and primary school is relevant.

Finally, actual hepatitis A notification data [72] are used to estimate the reporting

¹In light of a recent study on the HAV seroprevalence in Italy [8], the estimation of R_0 given by [1] were slightly revised.

Table 5.12: *Effectiveness of vaccination and improvements in hygienic conditions in Campania**.

Vaccination coverage of newborns	Vaccination coverage of 12 aged individuals	Duration of the vaccination campaign (in years)	Reduction of seafood contamination (%)	Estimated average R_0	Number of notified cases per year per 100,000	95% Confidence Interval	Recovered individuals* (% of population)	95% Confidence Interval	Symptomatic recovered individuals (% of recovered)	95% Confidence Interval	Vaccinated (% of population)	95% Confidence Interval	Number of vaccine doses per year per 100,000
-	-	-	-	2.9	11.7	11.1/12.3	65.8	62.7/68.9	59.2	58.5/60.3	-	-	-
10	50	†	-	-	8.3	8.0/8.8	47.2	44.9/49.6	61.0	60.2/62.2	23.9	22.4/25.0	573.5
20	80	†	-	-	7.0	6.5/7.3	39.9	37.3/41.8	62.5	61.6/63.5	35.1	33.0/36.7	818.9
80	90	†	-	-	6.0	5.5/6.4	32.7	29.9/34.7	67.6	66.7/68.3	51.0	49.1/53.3	1,121.7
20	80	5	-	-	11.1	10.5/11.5	62.3	59.8/64.8	59.8	59.1/60.9	3.5	2.9/4.1	90.5
20	80	10	-	-	10.3	9.9/10.7	58.6	56.8/60.5	60.3	59.6/61.4	7.1	5.9/8.4	181.8
20	80	20	-	-	9.0	8.8/9.2	51.4	50.5/52.4	61.3	60.6/62.3	13.9	11.4/16.0	343.7
-	-	-	20	2.1	9.7	8.9/10.4	51.2	48.7/54.7	66.0	65.7/66.8	-	-	-
-	-	-	50	1.9	8.7	8.0/9.3	46.2	43.7/49.2	67.5	66.9/68.3	-	-	-
-	-	-	80	1.8	8.3	7.7/9.0	44.6	42.1/47.4	67.9	67.4/68.8	-	-	-
20	80	†	20	-	5.8	5.4/6.1	33.3	30.9/35.0	67.9	67.4/68.8	36.7	34.4/38.3	818.5
20	80	†	50	-	5.5	5.1/5.8	31.7	29.3/33.3	69.0	68.6/69.8	37.0	34.7/38.6	818.7
20	80	†	80	-	5.4	5.0/5.7	31.1	28.8/32.7	69.3	69.0/70.2	37.1	34.8/38.7	818.7

* the duration of the simulation is 50 years.

* it represents the age-adjusted percentage of recovered individuals in the population.

† for the entire duration of the simulation.

rate.

5.5.1 Results

Baseline scenarios

The risk of infection by age for the three sources of infection estimated by the model agrees well with national survey data as reported in [94]. A comparison is shown in Fig. 5.17a.

At the endemic equilibrium, the fraction of recovered individuals² estimated by the model is 65.8% in Campania, and 73.8% in Puglia (see Table 5.12 and 5.13). These estimates comply with the seroprevalence data reported in [8], where an average value of 67.7% is reported for Southern Italy. Fig. 5.17b shows the fraction of recovered individuals per age cohort, compared with national seroprevalence data [8]. Age at infection of symptomatic individuals is shown in Fig. 5.17c. Since the age at infection in Puglia is lower than in Campania and the probability of developing symptoms increases with age, the fraction of symptomatic cases in the two regions is not significantly different (38.9% in Campania and 38.3% in Puglia).

Hepatitis A cases are seriously under-reported [11]. The estimated reporting rate is

²Individuals who contracted the infection and recovered; vaccinated individuals are thus excluded. Basically, it corresponds to the age-adjusted percentage of recovered individuals in the population.

Table 5.13: *Effectiveness of vaccination and improvements in hygienic conditions in Puglia**.

Vaccination coverage of newborns	Vaccination coverage of 12 aged individuals	Duration of the vaccination campaign (in years)	Reduction of seafood contamination (%)	Estimated average R_0	Number of notified cases per year per 100,000 cases	95% Confidence Interval	Recovered individuals* (% of population)	95% Confidence Interval	Symptomatic recovered individuals (% of recovered)	95% Confidence Interval	Vaccinated (% of population)	95% Confidence Interval	Number of vaccine doses per year per 100,000
-	-	†	-	3.8	53.3	50.9/54.0	73.8	71.1/76.1	51.9	50.9/53.9	-	-	-
10	50	†	-	-	37.6	35.8/39.6	54.1	52.3/56.2	54.3	53.3/56.3	22.5	21.4/23.6	573.4
20	80	†	-	-	30.6	29.2/32.5	45.3	43.0/47.2	56.1	55.2/58.0	33.7	32.0/35.2	818.7
80	90	†	-	-	24.5	22.7/26.2	35.9	33.0/37.9	62.9	61.6/64.2	50.2	48.4/52.6	1,122.1
20	80	5	-	-	50.6	48.4/52.6	70.6	68.4/72.3	52.6	51.5/54.4	3.2	2.7/3.8	90.5
20	80	10	-	-	47.9	46.0/49.4	67.2	65.7/68.4	53.2	52.3/54.9	6.7	5.5/7.7	181.9
20	80	20	-	-	40.7	38.8/42.2	57.0	56.1/58.0	56.0	55.1/57.3	13.4	11.0/15.3	343.7
-	-	-	20	2.5	50.1	44.8/54.5	60.3	57.1/64.0	62.4	62.1/63.1	-	-	-
-	-	-	50	2.2	44.4	39.5/48.6	53.5	51.0/57.1	64.7	64.1/65.6	-	-	-
-	-	-	80	2.1	42.2	37.7/46.0	51.1	48.5/54.4	65.5	64.7/66.4	-	-	-
20	80	†	20	-	26.6	24.8/28.6	37.8	35.1/39.6	64.4	64.0/65.2	35.9	33.8/37.5	818.6
20	80	†	50	-	24.7	23.0/26.5	35.7	33.1/37.5	66.1	65.7/67.0	36.4	34.2/37.9	818.7
20	80	†	80	-	24.1	22.4/25.8	35.0	32.4/36.7	66.7	66.2/67.5	36.5	34.3/38.1	818.7

* the duration of the simulation is 50 years.

* it represents the age-adjusted percentage of recovered individuals in the population.

† for the entire duration of the simulation.

about 2% in Campania and 8% in Puglia. These estimates are quite consistent with those reported by the Centers for Disease Control and Prevention for viral hepatitis A, namely less than 10% in the US [22]. These values are partially justified by the low perceived risk of hepatitis A among the population, especially in Campania where only 39% of the interviewed people would visit a medical doctor in case of icteric sclerae [118].

The estimated average number of notified cases per year per 100,000 inhabitants is 11.7 in Campania and 53.3 in Puglia, which are in good agreement with the values reported in [72], namely 14.3 in Campania and 52.6 in Puglia.

The model exhibits a “cluster-like pattern” (i.e., a large number of cases in a “short” period of time) within day care centers and kindergartens (see Fig. 5.17d). As discussed before, this is a typical pattern of most childhood diseases [53], and hepatitis A as well [26].

Effectiveness of vaccination programmes

In presence of a vaccination programme of unlimited duration involving both newborns and 12-years-old adolescents, hepatitis A can be controlled quite well. By performing a very mild vaccination programme (vaccination coverage of 50% among 12-years-old adolescents and of 10% among newborns), the fraction of symptomatic individuals reduces of 26% in Campania and of 23% in Puglia in 50 years. In both regions, the cost of this

Table 5.14: *Effectiveness of social distancing measures in Campania**.

Strategy	Strategy applied to (% of symptomatic individuals)	Estimated average R_0	Number of notified cases per year per 100,000	95% Confidence Interval	Recovered individuals* (% of population)	95% Confidence Interval	Symptomatic recovered individuals (% of recovered)	95% Confidence Interval
d1	100	2.8	11.7	11.1/12.3	64.3	61.0/67.5	60.8	60.0/62.4
d2	100	2.8	11.7	11.1/12.3	64.1	60.9/67.3	61.0	60.2/62.6
d3	100	2.8	11.7	11.1/12.4	64.1	60.9/67.3	61.1	60.3/62.7
d1	50	2.9	11.7	11.2/12.3	64.9	61.7/68.0	60.2	59.4/61.6
d2	50	2.9	11.7	11.2/12.3	64.7	61.6/67.9	60.3	59.6/61.7
d3	50	2.9	11.8	11.2/12.3	64.8	61.6/67.9	60.4	59.7/61.8
d1	10	2.9	11.7	11.0/12.3	65.6	62.4/68.7	59.4	58.6/60.6
d2	10	2.9	11.7	11.1/12.3	65.6	62.4/68.7	59.4	58.7/60.6
d3	10	2.9	11.7	11.1/12.3	65.6	62.4/68.7	59.5	58.8/60.7
d1	†	2.9	11.7	11.1/12.3	65.8	62.7/68.8	59.2	58.5/60.3
d2	†	2.9	11.7	11.0/12.2	65.8	62.7/68.8	59.2	58.5/60.3
d3	†	2.9	11.7	11.1/12.2	65.7	62.7/68.8	59.2	58.5/60.4

* it represents the age-adjusted percentage of recovered individuals in the population.
 * the duration of the simulation is 50 years.
 † strategy applied to only to notified cases.

vaccination campaign is about 570 vaccine doses per 100,000 individuals per year. Better results are obtained for higher vaccination coverage. In particular, for a very efficient vaccination strategy (vaccination coverage of 90% among 12-years-old adolescents and of 80% among newborns), the fraction of symptomatic individuals declines to 22.1% in Campania and to 22.6% in Puglia, with a cost of about 1,100 vaccine doses per 100,000 individuals per year. Whatever vaccination programme is applied, hepatitis A can not be eliminated because some sporadic cases can be caused by travels to high endemicity areas.

A vaccination coverage of 20% among newborns and of 80% among 12-years-old adolescents mimics the intervention implemented in Puglia which is the only Italian region that has introduced a vaccination programme started in 1997 [88], though vaccination is only recommended (and not mandatory): this is the reason of such a low vaccination coverage. In this setting, a vaccination campaign of limited duration (e.g., 10 years) does not have a significant effect in mitigating the epidemic. In fact, the fraction of recovered individuals decreases of only 5% in Campania and 4% in Puglia in 50 years. Fig. 5.18a, 5.18d, 5.18g, 5.19a, 5.19d, 5.19g, Table 5.12 and 5.13 show how better results can be obtained by extending the vaccination campaign over a longer period.

In both the considered regions, when vaccination is performed for at least 20 years, the fraction of recovered individuals drastically decreases in almost all age classes (but for the class of > 59 years old individuals) as an effect of the vaccination (see Fig. 5.18

Table 5.15: *Effectiveness of social distancing measures in Puglia**.

Strategy	Strategy applied to (% of symptomatic individuals)	Estimated average R_0	Number of notified cases per year per 100,000	95% Confidence Interval	Recovered individuals* (% of population)	95% Confidence Interval	Symptomatic recovered individuals (% of recovered)	95% Confidence Interval
$d1$	100	3.7	54.0	51.5/56.4	72.8	69.9/75.2	53.5	52.4/55.9
$d2$	100	3.7	54.2	51.6/56.8	72.6	69.9/75.0	53.7	52.6/56.1
$d3$	100	3.7	54.1	51.7/56.7	72.6	69.8/75.0	53.8	52.6/56.2
$d1$	50	3.8	53.7	51.1/56.0	73.2	70.4/75.6	52.9	51.8/55.1
$d2$	50	3.7	53.8	51.2/56.2	73.1	70.3/75.5	53.0	51.9/55.3
$d3$	50	3.7	53.8	51.3/56.1	73.1	70.3/75.4	53.1	52.0/55.3
$d1$	10	3.8	53.4	50.8/55.7	73.6	70.9/76.0	52.1	51.0/54.2
$d2$	10	3.8	53.3	50.9/55.4	73.6	70.9/75.9	52.1	51.1/54.2
$d3$	10	3.8	53.4	50.9/55.7	73.6	70.8/75.9	52.2	51.1/54.3
$d1$	†	3.8	53.3	50.8/55.5	73.7	71.0/76.1	52.0	50.9/54.1
$d2$	†	3.8	53.3	50.7/55.5	73.7	71.0/76.0	52.0	51.0/54.1
$d3$	†	3.8	53.3	50.8/55.5	73.7	71.0/75.9	52.0	50.9/54.1

* it represents the age-adjusted percentage of recovered individuals in the population.

* the duration of the simulation is 50 years.

† strategy applied only to notified cases.

5.18b, 5.18e, 5.18h, 5.19b, 5.19e and 5.19h). This reduction is not observed in the class of elderly individuals since they lived most of their life before the start of the vaccination programme. When vaccination is performed for at least 20 years, age at infection increases with respect to the baseline simulations as a consequence of the higher protection in the younger (and vaccinated) individuals and of the overall reduction in the number of cases in the population (see Fig. 5.18c, 5.18f, 5.18i, 5.19c, 5.19f and 5.19i). However, the increased age at infection is well compensated by the drastic reduction of the overall impact of the epidemic.

Effectiveness of social distancing measures

Social distancing measures $d1$, $d2$ and $d3$ do not have any positive effect: neither the fraction of recovered individuals nor the number of notified cases, nor the R_0 decline (see Table 5.14 and 5.15). Specifically, only a slight decrease in the fraction of recovered individuals (of the order of 1–2%, not statistically significant) can be observed when considering the more restrictive measures $d2$ and $d3$ applied to the 100% of symptomatic individuals, which is fairly unrealistic. In the more realistic case of strategies applied only to notified cases, no significant differences are observable with respect to the baseline scenarios.

Moreover, an undesired observed effect is to decrease the fraction of children and adolescents who contract the infection while increasing that of adults and elderly people

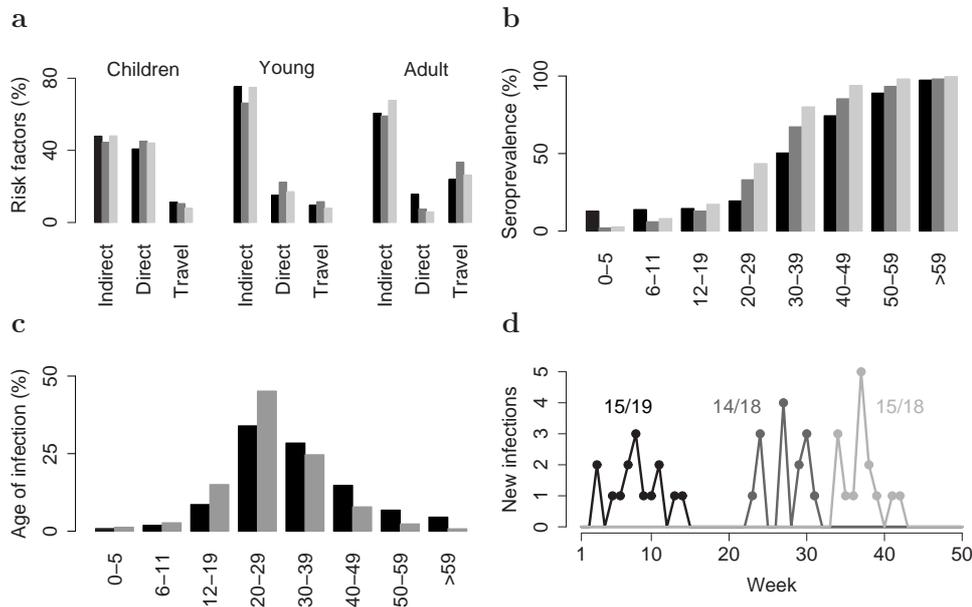


Figure 5.17: **a** Risk factors by age group. Comparison between the National data reported in [94] (black bars) and the baseline simulations in Campania (dark gray bars) and Puglia (light gray bars). Children account for individuals aged 0–14, Young for individuals aged 15–24 and Adults for individuals older than 24 years old. **b** HAV seroprevalence by age cohort. Comparison between National data as reported in [8] (black bars) and baseline simulations in Campania (dark gray bars) and Puglia (light gray bars). **c** Age at infection of symptomatic individuals in Campania (black bars) and Puglia (gray bars). **d** Number of cases over time in three randomly chosen day care centers in Puglia ($R_0 = 3.8$), highlighting the classical “clusters-like pattern”. The numbers refer to the fraction of cases over the number of children enrolled.

(see Fig. 5.20). Since the probability of developing acute symptoms increases with age, these interventions can be considered counterproductive. This effect is more relevant in Puglia and it is due to the higher value of R_0 which results in a lower age at infection.

Effects of improvements in hygienic conditions

Improvements in hygienic conditions have a wide effect on the hepatitis A dynamics. Although the decrease in the number of notified cases is not so relevant (see Table 5.12 and 5.13), the fraction of recovered and symptomatic individuals noticeably declines in 50 years, leading to estimated values of R_0 of 1.8–2.1 in Campania and 2.1–2.5 in Puglia: this values are typical of medium–low endemicity areas.

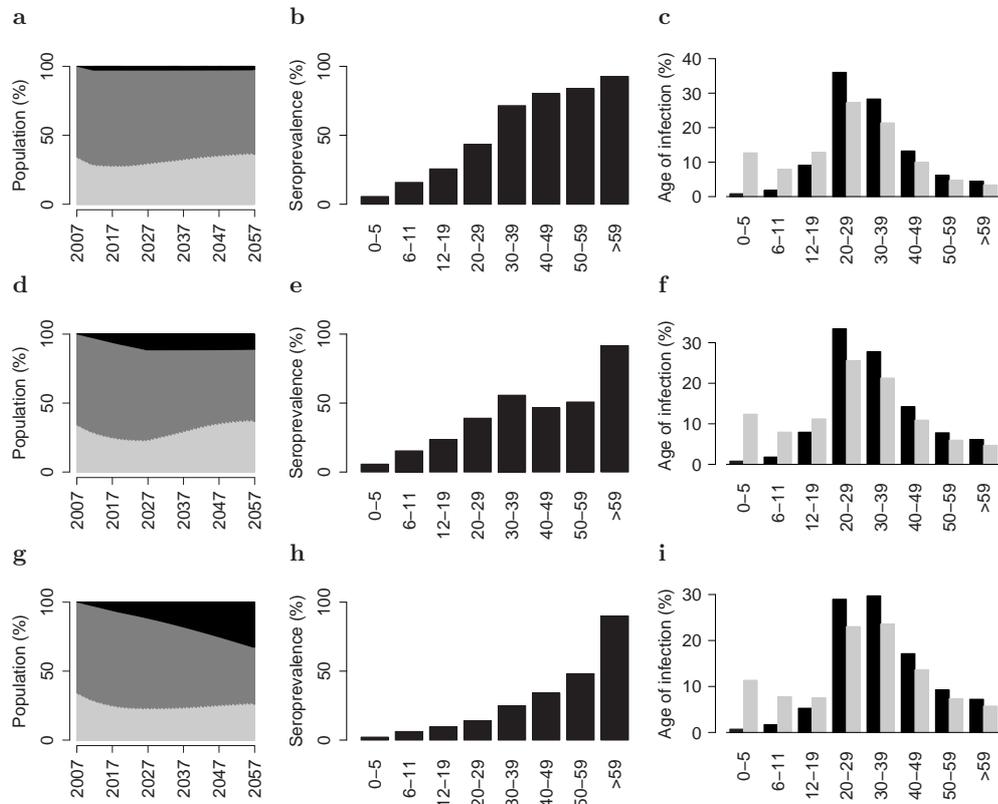


Figure 5.18: Effects of a vaccination programme in Campania. Vaccination coverage is assumed to be 20% among newborns and 80% among 12 years old adolescents. **a** percentage of susceptible individuals over time (light gray area), recovered individuals (dark gray area), vaccinated individuals (black area) and infected individuals (“small” white area between susceptible and recovered individuals) for a vaccination programme lasting 5 years. **b** HAV seroprevalence by age cohort for a vaccination programme lasting 5 years. **c** Age at infection of symptomatic individuals (black bars) and asymptomatic individuals (gray bars) for a vaccination programme lasting 5 years. **d**, **e**, **f** as **a**, **b**, **c**, but for a vaccination programme lasting 20 years. **g**, **h**, **i** as **a**, **b**, **c**, but for a vaccination programme lasting 50 years.

Finally, combining a vaccination programme with natural hygienic improvement leads, not surprisingly, to a very low fraction of recovered individuals values (see Table 5.12 and 5.13).

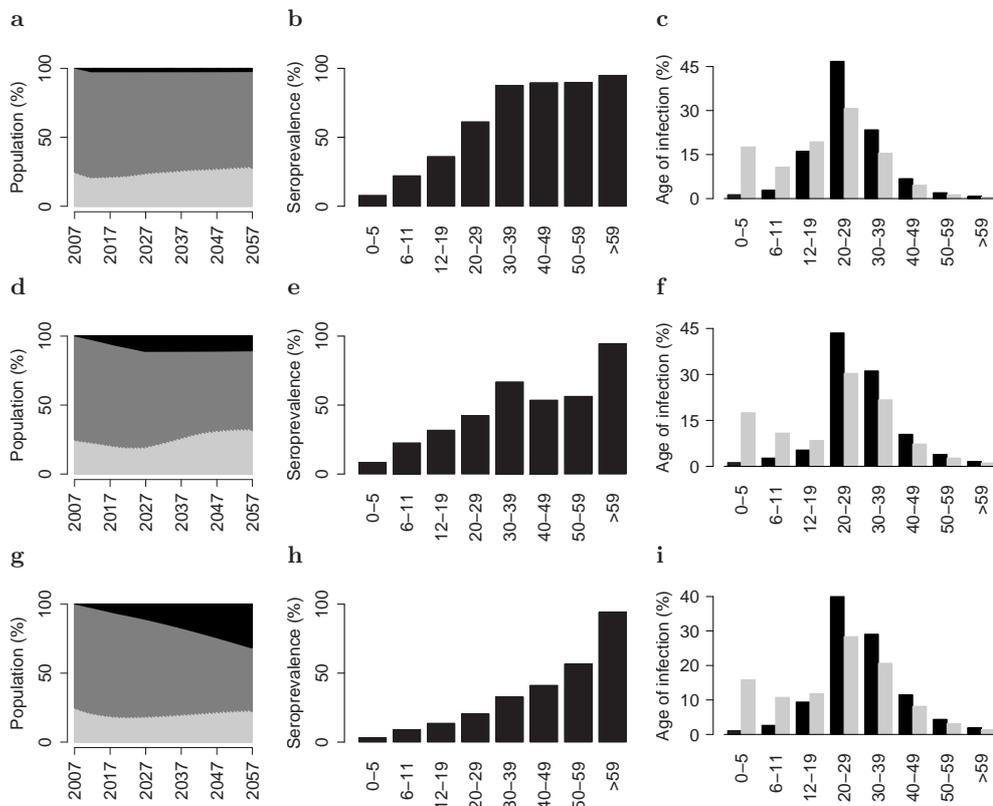


Figure 5.19: Effects of a vaccination programme in Puglia. Vaccination coverage is assumed to be 20% among newborns and 80% among 12 years old adolescents. **a** percentage of susceptible individuals over time (light gray area), recovered individuals (dark gray area), vaccinated individuals (black area) and infected individuals (“small” white area between susceptible and recovered individuals) for a vaccination programme lasting 5 years. **b** HAV seroprevalence by age cohort for a vaccination programme lasting 5 years. **c** Age at infection of symptomatic individuals (black bars) and asymptomatic individuals (gray bars) for a vaccination programme lasting 5 years. **d**, **e**, **f** as **a**, **b**, **c**, but for a vaccination programme lasting 20 years. **g**, **h**, **i** as **a**, **b**, **c**, but for a vaccination programme lasting 50 years.

5.5.2 Conclusions

The development of an algorithm for simulating the temporal evolution of the sociodemographic structure of the population, where individuals come to life, grow, attend school, go to work, can generate new households, can procreate and die, allowed us to develop the first large scale, spatially-explicit individual-based model, applied to the analysis

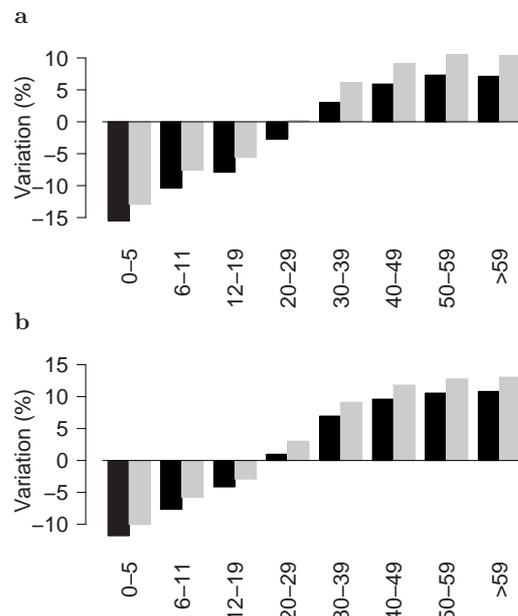


Figure 5.20: **a** Social distancing measure $d\beta$ applied to all symptomatic individuals: variation of the number of cases (gray bars) and number of symptomatic cases (black bars) in Campania by age cohorts. **b** As **a** but for Puglia region.

of hepatitis A in Italy. The dynamic sociodemographic model proposed in this work can be easily adapted to the study of other endemic diseases in which person-to-person transmission can play a relevant role.

The model complies with the available epidemiological data (e.g., the seroprevalence or the fraction of cases generated in the different sources of infection) and it allows analyzing the effectiveness of different intervention measures. A low vaccination coverage of unlimited duration is able to reduce hepatitis A, at a limited cost in terms of vaccine stockpile. However, a vaccination programme of limited duration does not result in a substantial decrease of the proportion of susceptible individuals in the population. Consequently, HAV can spread in the population and new outbreaks will arise as soon as the vaccination campaign is interrupted. Therefore, in the light of the results presented in this section, the mass vaccination adopted in the region of Puglia in 1996–1997 can be effective only if the vaccination programme is protracted. Unfortunately, serological data for the pre-vaccination period and for the vaccination period in Puglia are not available to us. Therefore, it is hard to validate this result empirically.

Expected and desirable improvements in hygienic conditions are sufficient to lead to a decline of the epidemic, even in the absence of vaccination (as observed in [77,

78]). The obtained estimates of R_0 suggest that the two high endemicity Italian regions could naturally evolve towards a medium–low endemicity level. The considered social distancing measures (isolation of symptomatic cases and closure of day care centers and kindergartens) do not have any positive effect on the containment of hepatitis A. On the contrary, the effect of these interventions is to decrease the fraction of children and adolescents who contract the infection while increasing that of adults and elderly people. Consequently, a side effect is to increase the risk of developing acute symptoms. In fact, blocking transmission in day care centers and kindergartens may not be an effective measure for decreasing the overall prevalence of infection in the population if more than 50% of cases are due to ingestion of infected seafood, as in the two considered regions [94]. This kind of measures could be effective only when coupled to drastic improvements in hygienic conditions. Elimination of the disease is not possible since new cases are continuously imported from high endemicity areas outside the country.

The assumption of a constant population (neither immigration nor emigration are considered in the model and the number of newborns is assumed to be equal to the number of deaths) represents a critical issue, though this is a reasonable modeling choice. Since the demographic projections are not completely reliable, the longer the temporal interval of the simulations, the less accurate the predictions are expected to be. Individual–based models allow considering highly heterogeneous populations, improving the characterization of contacts and representing the age–structure in a natural way. They offer the advantage of avoiding (arbitrary) assumptions on the transmission rates among age classes. However, in order to estimate the transmission rates, they require specific data on (*i*) the type of contact and (*ii*) the frequency of contact in the different places where transmission can occur.

We focused the study on the two most affected Italian regions. Since the risk factors greatly vary from region to region (e.g., the consumption of raw seafood is not common in Northern Italy), it would be interesting to develop a nationwide model, to analyze, for instance, the mechanisms behind the spatio–temporal spread of the epidemic. In some more detail, it would be interesting to analyze the effects of a vaccination campaign conducted only in the most affected regions on the temporal evolution of viral hepatitis A in the rest of the country.

We conclude by remarking the main findings of this work, potentially related to the implementation of control measures by public health authorities. A low vaccination coverage is sufficient to reduce the impact of hepatitis A, at a limited cost in terms of vaccine stockpile. However, if the vaccination programme is not protracted in time new outbreaks can arise as soon as the vaccination campaign is interrupted. Since vaccine is offered on

a voluntary basis, it is thus crucial to carry out stabilization measures. Improvements in hygienic conditions are sufficient to lead to a decline of the epidemic. Though they are difficult to achieve in a short period of time, monitoring of fish market stands and mass stabilization campaigns could be implemented to raise consumer and industry awareness. Social distancing measures are discouraged from both clinical and social points of view. They could be reassessed in the light of high vaccination coverage and drastic improvements in hygienic conditions.

Chapter 6

Related work

In the last 5 years, a lot of large-scale individual-based models have been proposed in literature. They are mainly used for evaluating the effectiveness for containing/mitigation measures in response to an influenza pandemic [85, 46, 87, 47, 54, 64, 29, 35, 96] or to a bioterroristic smallpox attack [42, 115, 86], but they have been allied also for the testing the efficacy of control policies for an endemic diseases, as the case of viral hepatitis A [3]. The scale at which such models work varies a lot among the studies, ranging from a community level (e.g., see [85]), a city level (e.g., see [42]) or a country level (e.g., see [47]), but none of them is focused on a continental scale as it is the case of this study.

For the topic, the scale, the considered level of details and the type of used datasets, the models much comparable to the one proposed here are the ones developed by Ferguson et al. [47] and by Germann et al. [54], which are both focused on the dynamics of an influenza pandemic in the US. Recently such models have been already compared in [64].

In [47], individuals are explicitly represented and co-located in households, schools and workplaces, which have a specific geographic locations. Within such groups an homogeneous mixing is assumed, while for all the other kind of contacts, the risk of infection depends on the geographic distance between individuals, which is weighted by a kernel function. Also in [54] individuals are explicitly represented and the epidemic can spread on a sophisticated network. This is obtained by considering the contacts that an individual can have within the household, school, workplace, extended family, neighborhood, local community and by long distance travels. Moreover, the force of infection explicitly depends on the age of the individual.

The structure of the model described in [47] is quite similar to the one adopted here. However, since it was focused on a single country, the population considered is, in some sense, homogeneous and the authors could not take into account differences in the so-

ciodemographic structure and in the level of human mobility. This holds also for the model described in [54], while it is one of the main novel feature of the model analyzed in this thesis. The network of contacts considered in [54], is much sophisticated than the one adopted here (and the one in [47]), but, as highlighted in [114], this means that the authors have to made more “arbitrary” assumptions on the transmission rates.

The explicit dependence of the force of infection on the distance between individuals used in [47] is similar to the one considered in this study. However, we consider also cross border diffusion and long distance travels. Finally, let us remark that, in both the previous models and in the one proposed here, the seeding of the infection is based on the arrival of infected individuals from abroad.

Since our European model (see Sec. 4.2) covers all the European countries, including United Kingdom, we are able to make a specific comparison on our predictions and the ones given by Ferguson et al. in [47]. This can be seen as a sort of validation of our model predictions against the results given in one of the milestones in this area. In order to perform such numerical comparison, we assume $R_0 = 2$. Remarkably, by assuming an infectious period of 1.5days in our model, cumulative attack rates as predicted by our and [47] models are 33.5% and 34% respectively; peak daily attack rate are 2% and 2.1% respectively; 95% confidence intervals for the peak day are (89,104) days and roughly (90,105) days respectively; 95% confidence intervals for the timing of initial case are (30,51) days and roughly (30,55) days respectively (see Fig. 1d in [47] and supplementary material of [47]). This excellent agreement in the predictions proves the robustness of the two models, since they are both based on different assumptions on the mobility of the population (e.g., the law for assigning commuting destinations) and on the epidemiological parameters (e.g., the distribution of the viral load over time in infectious individuals).

Chapter 7

Conclusion

Predicting the spread of epidemic outbreaks and evaluating the effectiveness of control strategies represent a priority for public health policy makers worldwide. In this context, mathematical models represent the scientific basis on which decisions can be taken. Recently, IBM have become the much relevant tool for the evaluation of individually targeted public health measures, which play a central role for controlling epidemic outbreaks. Probably, the much important applications are focused on the spread of an influenza pandemic (e.g., see [46, 54, 64]) or on a deliberate reintroduction of smallpox virus (e.g., see [42, 115, 86]).

In this thesis project, the developed models have been applied mainly to the investigation of an influenza pandemic. Our investigation have been focused on the theoretical aspects of the IBM, the description of the spatiotemporal dynamics of an influenza pandemic and on the evaluation of the effectiveness of different intervention strategies. In particular, at Italian scale, the diffusion scenarios drawn by our Italian model (described in Sec. 4.1) and our evaluation of the effectiveness of control strategies for the ongoing A/H1N1 influenza pandemic are currently used by the Italian Ministry of Health. Moreover, also at European scale, the predictions of our model (described in Sec. 4.2) are currently taken into account by the European Centre for Disease Prevention and Control for preparing their recommendation on the vaccine distribution.

As it has been anticipated in [114], a new frontier for individual-based simulations would be the application to long-time scale human diseases, such as tuberculosis, measles, hepatitis or HIV. In this thesis (and in the paper derived from it [3]), it has been shown the first application of an individual-based model to an endemic diseases: the viral hepatitis A. Our results show the effectiveness of individually targeted interventions which are not possible to test with other kind of models (e.g., classical ODE models). This would be

very useful for public health agencies.

Other plausible applications, not yet discussed in the scientific literature (nor in this project), could be to the analysis of the dynamics of a possible future severe acute respiratory syndrome (SARS) or marburg hemorrhagic fever (MHF) epidemics. Therefore, dealing with such kind of diseases could represent natural future applications of the models proposed in this thesis. Specifically, it would be easy to extend the transmission model for all diseases where household and school/workplaces contacts are relevant sources of infections. Moreover, the detailed models of the sociodemographic structure and of human mobility patterns of the Italian and European populations can be used for describing also other kinds of diffusion processes (e.g., the spread of other diseases, informations, mobile phone viruses).

A crucial point for a further more intensive employment of IBM is represented by the development of reliable methods for validating model projections against empirical epidemiological data. A lot of work was already done for comparing classical model predictions to empirical data, but as regards IBM, the theory is currently being developed. A first crucial step in this way has recently been made in [21], but a lot of effort should be done in this direction.

However, IBM does not represent the only suitable approach for simulating epidemic outbreaks. Therefore, one major issue thus concerns to which extent the epidemic evolution found by different modeling approaches may differ and depend on the different approximations and assumptions used. On this context, we are currently providing for the first time a side by side comparison of the results obtained with an IBM and a structured metapopulation stochastic model for the evolution of a baseline pandemic event in Italy. Preliminary results show a good agreement between the two modeling approaches. This thus defines the possibility of hybrid models combining IBM and metapopulation approaches according to the available data and computational resources.

Bibliography

- [1] M. Ajelli, M. Iannelli, P. Manfredi, and M. L. Ciolfi degli Atti. Basic mathematical models for the temporal dynamics of HAV in medium-endemicity Italian areas. *Vaccine*, 26(13):1697–1707, 2008.
- [2] M. Ajelli and S. Merler. The Impact of the Unstructured Contacts Component in Influenza Pandemic Modeling. *PLoS ONE*, 3(1):e1519, 2008.
- [3] M. Ajelli and S. Merler. An individual-based model of hepatitis A transmission. *Journal of Theoretical Biology*, 259(3):478–488, 2009.
- [4] M. Akazawa, J. L. Sindelar, and A. D. Paltiel. Economic Costs of Influenza-Related Work Absenteeism. *Value in Health*, 6(2):107–115, 2003.
- [5] M.E. Alexander, C.S. Bowman, Z. Feng, M. Gardam, S.M. Moghadas, G. Rost, J. Wu, and P. Yan. Emergence of drug resistance: implications for antiviral control of pandemic influenza. *Proceedings of the Royal Society B: Biological Sciences*, 274(1619):1675–1684, 2007.
- [6] R. M. Anderson and R. M. May. *Infectious diseases of humans: dynamics and control*. Oxford, UK: Oxford University Press, 1992.
- [7] V. Andreasen, C. Viboud, and L. Simonsen. Epidemiologic characterization of the 1918 influenza pandemic summer wave in Copenhagen: implications for pandemic control strategies. *Journal of Infectious Diseases*, 197(2):270–278, 2008.
- [8] F. Ansaldi, B. Bruzzone, M. C. Rota, A. Bella, M. L. Ciolfi degli Atti, P. Durando, R. Gasparini, and G. Icradi. Hepatitis A incidence and hospital-based seroprevalence in Italy: a nation-wide study. *European Journal of Epidemiology*, 23:45–53, 2008.
- [9] F. Y. Aoki, M. D. Macleod, P. Paggiaro, O. Carewicz, A. El Sawy, C. Wat, M. Griffiths, E. Waalberg, and P. Ward. Early administration of oral oseltamivir in-

- creases the benefits of influenza treatment. *Journal of Antimicrobial Chemotherapy*, 51(1):123–129, 2003.
- [10] J. Arino, C. Bowman, and S. Moghadas. Antiviral resistance during pandemic influenza: implications for stockpiling and drug use. *BMC Infectious Diseases*, 9(1):8, 2009.
- [11] G. L. Armstrong and B. P. Bell. Hepatitis A virus infection in the United States: model-based estimates and implications for childhood immunization. *Pediatrics*, 109:839–845, 2002.
- [12] F. Averhoff, C. N. Shapiro, B. P. Bell, I. Hyams, L. Burd, A. Deladisma, E. P. Simard, D. Nalin, B. Kuter, C. Ward, M. Lundberg, N. Smith, and H. S. Margolis. Control of Hepatitis A Through Routine Vaccination of Children. *Journal of the American Medical Association*, 286(23):2968–2973, 2001.
- [13] P. Baccam, C. Beauchemin, C. A. Macken, F. G. Hayden, and A. S. Perelson. Kinetics of Influenza A Virus Infection in Humans. *Journal of Virology*, 80(15):7590–7599, 2006.
- [14] D. Balk and G. Yetman. The global distribution of population: evaluating the gains in resolution refinement. Technical report, Center for International Earth Science Information Network, 2004.
- [15] J. J. Berge, D. P. Drennan, R. J. Jacobs, A. Jakins, A. S. Meyerhoff, W. Stubblefield, and M. Weinberg. The cost of hepatitis A infections in American adolescents and adults in 1997. *Hepatology*, 31(2):469–473, 2000.
- [16] E. Biziagos, J. Passagot, J. M. Crance, and Deloince R. Long-term survival of Hepatitis A virus and poliovirus Type 1 in mineral water. *Applied and Environmental Microbiology*, 54(11):2705–2710, 1988.
- [17] R. Breban, R. Vardavas, and S. Blower. Theory versus Data: How to Calculate R0? *PLoS ONE*, 2(3):e282, 2007.
- [18] D. Brockmann, L. Hufnagel, and T. Geisel. The scaling laws of human travel. *Nature*, 439:462–465, 2006.
- [19] J.S. Brownstein, C.J. Wolfe, and K.D. Mandl. Empirical evidence for the effect of airline travel on inter-regional influenza spread in the United States. *PLoS Med*, 3(10):e401, 2006.

- [20] P. Caley, NG Becker, and DJ Philp. The waiting time for inter-country spread of pandemic influenza. *PLoS ONE*, 2(1):e143, 2007.
- [21] S. Cauchemez, A. J. Valleron, P. Y. Boelle, A. Flahault, and N. M. Ferguson. Estimating the impact of school closure on influenza transmission from sentinel data. *Nature*, 452(7188):750–754, 2008.
- [22] CDC. Disease Burden from Hepatitis A, B, and C in the United States. http://www.cdc.gov/ncidod/diseases/hepatitis/resource/dz_burden.htm, 2006. Centers for Disease Control and Prevention.
- [23] CDC. *Health Information for International Travel 2008*. Atlanta, US, Department of Health and Human Services, Public Health Service, 2007. Centers for Disease Control and Prevention.
- [24] M. Chironna, C. Germinario, D. De Medici, A. Fiore, S. Di Pasquale, M. Quarto, and S. Barbuti. Detection of hepatitis A virus in mussels from different sources marketed in Puglia region (South Italy). *International Journal of Food Microbiology*, 75:11–18, 2002.
- [25] M. Chironna, A. Grottola, C. Lanave, E. Villa, S. Barbuti, and M. Quarto. Genetic analysis of HAV strains recovered from patients with acute hepatitis from Southern Italy. *Journal of Medical Virology*, 70(3):343–349, 2003.
- [26] S. D. Chitambar, M. S. Chadha, L. R. Yeolekar, and V. A. Arankalle. Hepatitis A in day care centre. *Indian Journal of Pediatrics*, 63(6):781–783, 1996.
- [27] G. Chowell, C. E. Ammon, N. W. Hengartner, and J. M. Hyman. Transmission dynamics of the great influenza pandemic of 1918 in Geneva, Switzerland: assessing the effects of hypothetical interventions. *Journal of Theoretical Biology*, 241(2):193–204, 2006.
- [28] G. Chowell, N.W. Hengartner, C. Castillo-Chavez, P.W. Fenimore, and J.M. Hyman. The basic reproductive number of Ebola and the effects of public health measures: the cases of Congo and Uganda. *J Theor Biol*, 229(1):119–126, 2004.
- [29] M. L. Ciofi degli Atti, S. Merler, C. Rizzo, M. Ajelli, M. Massari, P. Manfredi, C. Furlanello, G. Scalia Tomba, and M. Iannelli. Mitigation measures for pandemic influenza in Italy: an individual based model considering different scenarios. *PLoS ONE*, 3(3):e1790, 2008.

- [30] V. Colizza, A. Barrat, Barthélemy, M., and A. Vespignani. The role of the airline transportation network in the prediction and predictability of global epidemics. *Proceedings of the National Academy of Sciences*, 103(7):2015–2020, 2006.
- [31] V. Colizza, A. Barrat, M. Barthélemy, A.-J. Valleron, and A. Vespignani. Modeling the worldwide spread of pandemic influenza: baseline case and containment interventions. *PLoS Med*, 4(1):e13, 2007.
- [32] B.S. Cooper, R.J. Pitman, W.J. Edmunds, and N.J. Gay. Delaying the international spread of pandemic influenza. *PLoS Med*, 3(6):e212, 2006.
- [33] R. D’Amelio, A. Mele, A. Mariano, L. Roman, R. Biselli, F. Lista, A. Zanetti, and T. Stroffolini. Hepatitis A, Italy. *Infectious Diseases*, 11(7), 2005.
- [34] A. Das. An economic analysis of different strategies of immunization against hepatitis A virus in developed countries. *Hepatology*, 29(2):548–552, 2003.
- [35] V.J. Davey, R.J. Glass, H.J. Min, W.E. Beyeler, and L.M. Glass. Effective, robust design of community mitigation for pandemic influenza: a systematic examination of proposed US guidance. *PLoS ONE*, 3(7):e2606, 2008.
- [36] V. Demicheli, D. Rivetti, JJ Deeks, and TO Jefferson. Vaccines for preventing influenza in healthy adults. *Cochrane database of systematic reviews (Online)*, (3), 2004.
- [37] O. Diekmann and J. A. P Heesterbeek. *Mathematical epidemiology of infectious diseases: model building, analysis and interpretation*. John Wiley & Son, 2000.
- [38] J. Dushoff and S. Levin. The effects of population heterogeneity on disease invasion. *Math Biosci*, 128(1-2):25–40, 1995.
- [39] ECDC. ECDC Situation Report. Pandemic influenza (H1N1) 2009, update 6 August 2009., 2009. http://ecdc.europa.eu/en/files/pdf/Health_topics/Situation_Report_09080%6_1700hrs.pdf.
- [40] M. Eichner, M. Schwehm, H.-P. Duerr, M. Witschi, D. Koch, S. Brockmann, and B. Vidondo. Antiviral prophylaxis during pandemic influenza may increase drug resistance. *BMC Infectious Diseases*, 9(1):4, 2009.
- [41] J.M. Epstein, D.M. Goedecke, F. Yu, R.J. Morrison, D. Wagener, and G.V. Bobashev. Controlling pandemic flu: the value of international air travel restrictions. *PLoS ONE*, 2(5):e401, 2007.

- [42] S. Eubank, H. Guclu, A. V. S. Kumar, M. V. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang. Modelling disease outbreaks in realistic urban social networks. *Nature*, 429:180–184, 2004.
- [43] European Centre for Disease Prevention and Control. Annual Epidemiological Report on Communicable Diseases in Europe 2008, 2008.
- [44] European Commission. Key Data on Education in Europe 2005, 2005.
- [45] N. M. Ferguson. Capturing human behaviour. *Nature*, 446:733, April 2007.
- [46] N. M. Ferguson, D. A. Cummings, S. Cauchemez, C. Fraser, S. Riley, A. Meeyai, S. Iamsirithaworn, and D. S. Burke. Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature*, 437(7056):209–214, 2005.
- [47] N. M. Ferguson, D. A. Cummings, C. Fraser, J. C. Cajka, and P. C. Cooley. Strategies for mitigating an influenza pandemic. *Nature*, 442(7101):448–452, 2006.
- [48] N. M. Ferguson, C. A. Donnelly, and R. M. Anderson. The Foot–and–Mouth Epidemic in Great Britain: Pattern of Spread and Impact of Interventions. *Science*, 292:1155–1160, 2001.
- [49] A. E. Fiore. Hepatitis A Transmitted by Food. *Clinical Infectious Diseases*, 38:705–715, 2004.
- [50] A. Flahault, E. Vergu, L. Coudeville, and R. F. Grais. Strategies for containing a global influenza pandemic. *Vaccine*, 24:6751–6755, 2006.
- [51] D. M. Fleming, A. J. Elliot, A. Meijer, and W. J. Paget. Influenza virus resistance to oseltamivir: what are the implications? *The European Journal of Public Health*, 19(3):238–239, 2009.
- [52] C. Fraser, C.A. Donnelly, S. Cauchemez, W.P. Hanage, M.D. Van Kerkhove, T.D. Hollingsworth, J. Griffin, R.F. Baggaley, H.E. Jenkins, E.J. Lyons, et al. Pandemic potential of a strain of influenza A (H1N1): early findings. *Science*, 324(5934):1557, 2009.
- [53] K. Galil, B. Lee, T. Strine, C. Carraher, A. AL. Baughman, M. Eaton, J. Montero, and J. Seward. Outbreak of Varicella at a Day-Care Center despite Vaccination. *The New England Journal of Medicine*, 347(24):1909–1915, 2002.

- [54] T. C. Germann, K. Kadau, I. M. J. Longini, and C. A. Macken. Mitigation strategies for pandemic influenza in the United States. *Proceedings of the National Academy of Sciences*, 103(15):5935–5940, 2006.
- [55] R.J. Glass, L.M. Glass, W.E. Beyeler, and H.J. Min. Targeted social distancing design for pandemic influenza. *Emerg Infect Dis*, 12(11):1671–1681, 2006.
- [56] W.P. Glezen. Emerging Infections: Pandemic Influenza. *Epidemiologic Reviews*, 18(1):64–76, 1996.
- [57] W.P. Glezen and L. Simonsen. Commentary: Benefits of influenza vaccine in US elderly—new studies raise questions. *International Journal of Epidemiology*, 35(2):352–353, 2006.
- [58] M.C. Gonzalez, C.A. Hidalgo, and A.-L. Barabasi. Understanding individual human mobility patterns. *Nature*, 453(5):779–782, 2008.
- [59] K. Goodwin, C. Viboud, and L. Simonsen. Antibody response to influenza vaccination in the elderly: a quantitative review. *Vaccine*, 24(8):1159–1169, 2006.
- [60] R.F. Grais, J. Hugh Ellis, and G.E. Glass. Assessing the impact of airline travel on the geographic spread of pandemic influenza. *European Journal of Epidemiology*, 18(11):1065–1072, 2003.
- [61] B. T. Grenfell, O. N. Bjornstad, and J. Kappey. Travelling waves and spatial hierarchies in measles epidemics. *Nature*, 414(6865):716–723, 2001.
- [62] T. Grosche, F. Rothlauf, and A. Heinzla. Gravity models for airline passenger volume estimation. *J Air Transp Manag*, 13(4):175–183, 2007.
- [63] I. M. Hall, R. Gani, H. E. Hughes, and S. Leach. Real-time epidemic forecasting for pandemic influenza. *Epidemiol Infect*, 135(3):372–385, 2007.
- [64] M. E. Halloran, N. M. Ferguson, S. Eubank, I. M. Longini Jr, D. A. T. Cummings, B. Lewis, S. Xu, C. Fraser, A. Vullikanti, T. C. Germann, D. Wagener, R. Beckman, K. Kadau, C. Barrett, C. A. Macken, D. S. Burke, and P. Cooley. Modeling targeted layered containment of an influenza pandemic in the United States. *Proceedings of the National Academy of Sciences*, 105(12):4639, 2008.
- [65] M. E. Halloran, I. M. Jr Longini, A. Nizam, and Y. Yang. Containing Bioterrorist Smallpox. *Science*, 298:1428–1432, 2002.

- [66] W. Hanshaoworakul, J.M. Simmerman, U. Narueponjirakul, W. Sanasuttipun, V. Shinde, S. Kaewchana, D. Areechokechai, J. Levy, and K Ungchusak. Severe Human Influenza Infections in Thailand: Oseltamivir Treatment and Risk Factors for Fatal Outcome. *PLoS ONE*, 4(6):e6051, 2009.
- [67] F. G. Hayden. Developing New Antiviral Agents for Influenza Treatment: What Does the Future Hold? *Clinical Infectious Diseases*, 48(S1):S3–S13, 2009.
- [68] L. Hufnagel, D. Brockmann, and T. Geisel. Forecast and control of epidemics in a globalized world. *Proc Natl Acad Sci USA*, 101:15124–15129, 2004.
- [69] M. Iannelli and P. Manfredi. Demographic Change and Immigration in Age-structured Epidemic Models. *Mathematical Population Studies*, 14(3):169–191, 2007.
- [70] Influenza team (ECDC). Pandemic preparedness in the European Union - multi-sectoral planning needed. *Eurosurveillance*, 12(8):3143, 2007.
- [71] ISTAT. Life tables. <http://demo.istat.it/unitav/index.html?lingua=eng>, 1974–2004. Italian Institute of Statistics.
- [72] ISTAT. Annuario statistico Italiano, 1980–2006. Italian Institute of Statistics. (In Italian).
- [73] ISTAT. VIII Censimento generale dell’industria e dei servizi. <http://dwcis.istat.it/cis/index.htm>, 2001. Italian Institute of Statistics. (In Italian).
- [74] ISTAT. Strutture familiari e opinioni su famiglia e figli. http://www.istat.it/dati/catalogo/20060621_03/, 2003. Italian Institute of Statistics. (In Italian).
- [75] ISTAT. Resident population. http://demo.istat.it/pop2007/index_e.html, 2007. Italian Institute of Statistics.
- [76] ISTAT. Matrimoni, separazioni e divorzi. <http://www.istat.it/altridati/matrimoni/>, 2008. Italian Institute of Statistics. (In Italian).
- [77] K. H. Jacobsen and J. S. Koopman. Declining hepatitis A seroprevalence: a global review and analysis. *Epidemiology and Infection*, 132:1005–1022, 2004.
- [78] K. H. Jacobsen and J. S. Koopman. The effects of socioeconomic development on worldwide hepatitis A virus seroprevalence patterns. *International Journal of Epidemiology*, 34:600–609, 2005.

- [79] T. Jefferson, S. Smith, V. Demicheli, A. Harnden, A. Rivetti, and C. Di Pietrantonj. Assessment of the efficacy and effectiveness of influenza vaccines in healthy children: systematic review. *The Lancet*, 365(9461):773–780, 2005.
- [80] N. Johnson and J. Mueller. Updating the accounts: global mortality of the 1918–1920 “Spanish” influenza pandemic. *Bulletin of the History of Medicine*, 76(1):105–115, 2002.
- [81] M.J. Keeling, M. E. J. Woolhouse, D. J. Shaw, L. Matthews, M. Chase-Topping, D. T. Haydon, S. J. Cornell, J. Kappey, J. Wilesmith, and B. T. Grenfell. Dynamics of the 2001 uk foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape. *Science*, 294:813–817, 2001.
- [82] W. O. Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 115:700–721, 1927.
- [83] Y. Lerman, G. Chodik, H. Aloni, J. Ribak, and S. Ashkenazi. Occupations at Increased Risk of Hepatitis A: A 2-Year Nationwide Historical Prospective Study. *American Journal of Epidemiology*, 150(3):312–320, 1999.
- [84] M. Lipsitch and C. Viboud. Influenza seasonality: Lifting the fog. *Proc Natl Acad Sci USA*, 106(10):3645, 2009.
- [85] I. M. Jr Longini, M. E. Halloran, A. Nizam, and Y. Yang. Containing Pandemic Influenza with Antiviral Agents. *American Journal of Epidemiology*, 159(7):623–633, 2004.
- [86] I. M. Jr Longini, M. E. Halloran, A. Nizam, Y. Yang, S. Xu, D. Burke, D. Cummings, and J. Epstein. Containing a Large Bioterrorist Smallpox Attack: A Computer Simulation. *International Journal of Infectious Diseases*, 11(2):98–108, 2006.
- [87] I. M. Jr Longini, A. Nizam, S. Xu, K. Ungchusak, W. Hanshaoworakul, D. A. Cummings, and M. E. Halloran. Containing pandemic influenza at the source. *Science*, 309(5737):1083–1087, 2005.
- [88] P. L. Lopalco, P. Malfait, F. Menniti-Ippolito, R. Prato, C. Germinario, M. Chironna, M. Quarto, and S. Salmaso. Determinants of acquiring hepatitis A virus disease in a large Italian region in endemic and epidemic periods. *Journal of Viral Hepatitis*, 12(3):315–321, 2005.

- [89] P. L. Lopalco, P. Malfait, S. Salmaso, C. Germinario, G. Salamina, M. Quarto, S. Barbuti, R. Cipriani, A. Mundo, and G. Pesole. A persisting outbreak of hepatitis A in Puglia, Italy, 1996: epidemiological follow-up. *Euro Surveillance*, 2(4):31–32, 1997.
- [90] C. Lucioni, V. Cipriani, S. Mazzi, and Panunzio M. Cost of an Outbreak of Hepatitis A in Puglia, Italy. *Pharmaco Economics*, 13(2):257–266, 1998.
- [91] H. Matsumoto. International urban systems and air passenger and cargo flows: some calculations. *J Air Transp Manag*, 10(4):239–247, 2004.
- [92] J. N. Mbithi, V. S. Springthorpe, and S. A. Sattar. Effect of relative humidity and air temperature on survival of Hepatitis A virus on environmental surfaces. *Applied and Environmental Microbiology*, 57(5):1394–1399, 1991.
- [93] A. McGeer, K. A. Green, A. Plevneshi, A. Shigayeva, N. Siddiqi, J. Raboud, and D. E. Low. Antiviral therapy and outcomes of influenza requiring hospitalization in Ontario, Canada. *Clinical Infectious Diseases*, 45(12):1568–1575, 2007.
- [94] A. Mele, M.E. Tosti, E. Spada, A. Mariano, E. Bianco, and SEIEVA Collaborative Group. Epidemiology of acute viral hepatitis: twenty years of surveillance through SEIEVA in Italy and a review of the literature. *Technical Report ISTISAN*, 12(6):1–39, 2006.
- [95] S. Merler and M. Ajelli. The role of population heterogeneity and human mobility in the spread of pandemic influenza. *Proc Royal Soc B*, 2009. In press.
- [96] S. Merler, M. Ajelli, and C. Rizzo. Age-prioritized use of antivirals during an influenza pandemic. *BMC Infectious Diseases*, 9:117, 2009.
- [97] S. Merler, P. Poletti, M. Ajelli, B. Caprile, and P. Manfredi. Coinfection can trigger multiple pandemic waves. *J Theor Biol*, 254(2):499–507, 2008.
- [98] S. Merler, C. Rizzo, M. Ajelli, M. Massari, G. Scalia Tomba, C. Furlanello, and M.L. Ciofi degli Atti. Modelling preventive measures during an influenza pandemic in Italy: a real time simulation strategy. In *Options for the Control of Influenza VI: Proceedings of the Sixth International Conference on Options for the Control of Influenza*, pages 169–172, 2007.
- [99] C.E. Mills, J.M. Robins, and M. Lipsitch. Transmissibility of 1918 pandemic influenza. *Nature*, 432(7019):904–906, 2004.

- [100] Ministero della Salute. Influenza Vaccination Coverage, 2005. <http://www.ministerosalute.it/promozione/malattie/malattie.jsp>.
- [101] MIUR. La scuola in cifre. <http://www.miur.it/ustat/documenti/pub2005/>, 2005. Italian Ministry of University and Research. (In Italian).
- [102] MIUR. L'università in cifre. <http://www.miur.it/ustat/documenti/pub2005/>, 2005. Italian Ministry of University and Research. (In Italian).
- [103] S.M. Moghadas, C.S. Bowman, G. Rost, and J. Wu. Population-wide emergence of antiviral resistance during pandemic influenza. *PLoS ONE*, 3(3):e1839, 2008.
- [104] A.S. Monto. Vaccines and antiviral drugs in pandemic preparedness. *Emerging infectious diseases*, 12(1):55, 2006.
- [105] J. Mossong, N. Hens, M. Jit, P. Beutels, K. Auranen, R. Mikolajczyk, M. Massari, S. Salmaso, G.S. Tomba, J. Wallinga, J. Heijne, M. Sadkowska-Todys, M. Rosinska11, and J.W. Edmunds. Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases. *PLoS Med*, 5(3):e74, 2008.
- [106] C. J. L. Murray and A. D. Lopez. Mortality by cause for eight regions of the world: Global Burden of Disease Study. *The Lancet*, 349:1269–1276, 1997.
- [107] C.J.L. Murray, A.D. Lopez, B. Chin, D. Feehan, and K.H. Hill. Estimation of potential global pandemic influenza mortality on the basis of vital registry data from the 1918–20 pandemic: a quantitative analysis. *Lancet*, 368(9554):2211–2218, 2007.
- [108] K. M. Neuzil, C. Hohlbein, and Y. Zhu. Illness Among Schoolchildren During Influenza Season. Effect on School Absenteeism, Parental Absenteeism From Work, and Secondary Illness in Families. *Arch Pediatr Adolesc Med.*, 156:986–991, 2002.
- [109] J. S. Nguyen-Van-Tam and A. W. Hampson. The epidemiology and clinical impact of pandemic influenza. *Vaccine*, 21:1762–1768, 2003.
- [110] Italian Institute of Statistics (ISTAT). XIV Censimento generale della popolazione e delle abitazioni. Availibile on-line at <http://dawinci.istat.it/MD/>, 2001.
- [111] A. D. M. E. Osterhaus. Pre- or post-pandemic influenza vaccine? *Vaccine*, 25(27):4983–84, 2007.

- [112] M. T. Osterholm. Preparing for the Next Pandemic. *The New England Journal of Medicine*, 352(18):1839–42, 2004.
- [113] P. Poletti, B. Caprile, M. Ajelli, A. Pugliese, and S. Merler. Spontaneous behavioural changes in response to epidemics. *J Theor Biol*, 260(1):31–40, 2009.
- [114] S. Riley. Large-Scale Spatial-Transmission Models of Infectious Disease. *Science*, 316(5829):1298–1301, 2007.
- [115] S. Riley and N. M. Ferguson. Smallpox transmission and control: Spatial dynamics in Great Britain. *Proceedings of the National Academy of Sciences*, 103(33):12637–12642, 2006.
- [116] S. Riley, C. Fraser, C. A. Donnelly, A. C. Ghani, L. J. Abu-Raddad, A. J. Hedley, G. M. Leung, L.-M. Ho, T.-H. Lam, T. Q. Thach, P. Chau, K.-P. Chan, S.-V. Lo, P.-Y. Leung, T. Tsang, W. Ho, K.-H. Lee, E. M. C. Lau, N. M. Ferguson, and R. M. Anderson. Transmission Dynamics of the Etiological Agent of SARS in Hong Kong: Impact of Public Health Interventions. *Science*, 300:1961–1966, 2003.
- [117] C. Rizzo, A. Bella, C. Viboud, L. Simonsen, M. A. Miller, M. C. Rota, S. Salmaso, and M. L. Ciofi degli Atti. Trends for influenza-related deaths during pandemic and epidemic seasons, Italy, 1969–2001. *Emerg Infect Dis*, 13(5):694–699, 2007.
- [118] G. Salamina and P. D’Argenio. Shellfish consumption and awareness of risk of acquiring hepatitis A among Neapolitan families - Italy 1997. *Euro Surveillance*, 3(10), 1998.
- [119] J. Shaman and M. Kohn. Absolute humidity modulates influenza survival, transmission, and seasonality. *Proc Natl Acad Sci USA*, 106(9):3243–3248, 2009.
- [120] H. Shuval. Estimating the global burden of thalassogenic diseases: human infectious diseases caused by wastewater pollution of the marine environment. *Journal of Water and Health*, 1:53–64, 2003.
- [121] A. C. Singer, B. M. Howard, A. C. Johnson, C. J. Knowles, S. Jackman, C. Accinelli, A. Barra Caracciolo, I. Bernard, S. Bird, T. Boucard, A. Boxall, J. V. Brian, E. Cartmell, C. Chubb, J. Churchley, S. Costigan, M. Crane, M. J. Dempsey, B. Dorrington, B. Ellor, J. Fick, J. Holmes, T. Hutchinson, F. Karcher, S. L. Kelleher, P. Marsden, G. Noone, M. A. Nunn, J. Oxford, T. Rachwal, N. Roberts, M. Roberts, M. L. Sacca, M. Sanders, J. O. Straub, A. Terry, D. Thomas, S. Toovey, R. Townsend,

- N. Voulvoulis, and C. Watts. Meeting Report: Risk Assessment of Tamiflu Use Under Pandemic Conditions. *Environmental Health Perspectives*, 116(11):1563–1567, 2008.
- [122] J. T. Stapleton and S. M. Lemon. *Infectious Diseases*, chapter Hepatitis A and hepatitis E, pages 790–797. Philadelphia, US, Lippincott Co, 1994.
- [123] United Nations. World Population Prospects - The 2006 Revision, 2007.
- [124] L. Uscher-Pines, S.B. Omer, D.J. Barnett, T.A. Burke, and R.D. Balicer. Priority setting for pandemic influenza: an analysis of national preparedness plans. *PLoS Med*, 3(10):e436, 2006.
- [125] T.M. Uyeki. Influenza diagnosis and treatment in children: a review of studies on clinically useful tests and antiviral treatment for influenza. *The Pediatric Infectious Disease Journal*, 22(2):164–177, 2003.
- [126] C. Viboud, O. N. Bjornstad, D. L. Smith, L. Simonsen, M. A. Miller, and B. T. Grenfell. Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science*, 312(5772):447–451, 2006.
- [127] C. Viboud, M. A. Miller, B. T. Grenfell, O. N. Bjornstad, and L. Simonsen. Air travel and the spread of influenza: important caveats. *PLoS Medicine*, 3(11):e503, 2006.
- [128] R. J. Webby and R. G. Webster. Are We Ready for Pandemic Influenza? *Science*, 302(5650):1519–1522, 2003.
- [129] WHO. WHO global influenza preparedness plan, 2005. http://www.who.int/csr/resources/publications/influenza/WHO_CDS_CSR_GIP%_2005_5/en/.
- [130] WHO. Influenza A(H1N1) - update 50, 2009. http://www.who.int/csr/don/2009_06_17/en/index.html.
- [131] World Health Organization. *Ethical considerations in developing a public health response to pandemic influenza*. WHO, Geneva, 2007.
- [132] J. T. Wu, S. Riley, C. Fraser, and G. M. Leung. Reducing the impact of the next influenza pandemic using household-based public health interventions. *PLoS Medicine*, 3:e361, 2006.

-
- [133] E. Zagheni, F.C. Billari, P. Manfredi, A. Melegaro, J. Mossong, and W.J. Edmunds. Using Time-Use Data to Parameterize Models for the Spread of Close-Contact Infectious Diseases. *Am J Epidemiol*, 168(9):1082–1090, 2008.