



UNIVERSITÀ DEGLI STUDI DI TRENTO

Facoltà di Scienze Matematiche, Fisiche e Naturali

A thesis submitted for the degree of Doctor of Philosophy

Ilaria Dorigatti

**Mathematical modelling of emerging and
re-emerging infectious diseases in human and
animal populations**

Advisor:

Prof. Andrea Pugliese

February 2011

Acknowledgments

First, I would like to thank my advisor Prof. Andrea Pugliese, who supervised on my education and research since my bachelor. Thank you Andrea for your teaching, all the time you spent for me during these years in clarifying my doubts, reading my drafts and for having always given to me useful suggestions to improve my work. Thank you for having been a brilliant reference and for having given to me the opportunity to travel, meet and collaborate with other people in the field. Briefly said, thank you for having led me to this stage.

I feel to owe a special thank to Prof. Neil Ferguson, who first allowed me to visit his research group and then to be part of it. Thanks Neil for the opportunities you gave and are giving to me, for your patient and stimulating supervision, for the thousands of suggestions you offered when things seemed not to work and for the honesty with which, upon my request, you said your opinion to me.

Another big thank goes to Dr. Simon Cauchemez, who was, is and will be my referee for a variety of statistical issues I cannot deal by myself. Many thanks Simon for being a careful advisor, for your patience, fundamental help and for making statistics sweet through doughnut bets!

Dr. Stefano Merler and Dr. Marco Ajelli deserve another special acknowledgement for the useful discussions I had with them, for having put in my hands their new individual-based model and having allowed me to work and learn from it; thanks also to Dr. Roberto Rosá and Dr. Marco Pautasso for having involved me in interesting works. Thanks also to Dr. Antonella

Lunelli for the clarifying discussions we had on each-other doubts, for her generosity in sharing the computational tricks she discovered by herself but foremost, for her friendship and invaluable support.

A huge, massive thank is for my parents Grazia and Mariano. Thanks for having always supported me, even in my most stressful days when I could barely stand myself. Thanks for the freedom you have always given to me, for having shared with me my difficult periods, my doubts on the future and for having always been a fixed point and a reference in my life.

It's now difficult and dangerous to name the friends who shared with me the experience started three years ago, without taking the risk of forgetting someone. Thanks to all the colleagues (Valentina, Laura, Ester, Philipp, Andrea, Davide, Giovanni) with whom I shared the most of my time during these years. Thanks for the uncountable funny lunches and coffees (discussions included) we had together!

Christian deserves a huge thank for his invaluable moral and technical support during these last months, for his patience with me and for having been able to make me smile, even when I was mostly stressed. Loads of thank also to Matteo for always having such a positive attitude towards life and, together with Iacopo and Claudio, for having being present when I mostly needed some help.

Thanks to the whole UFO ultimate frisbee team for having shared with me amazing days of sport and agonism and to my cousin Diego and friend Domenico for having been the perfect company of many hikings and mountaineering ski trips.

Finally, thanks to all the friends I met in London who made my experience so unique and exciting: thanks to the DIDE (especially Dr. Lorenzo Pellis, Dr. Paul Parham, Ms. Roalind Eggo, Miss Ide Cremin, Dr. Emily Lyons and Miss Francesca Tracey), to Mrs. Jenny and Mr. Peter Littlehales for having taken care of me like if I were a daughter, to my flat-mate Cecilia, for the interesting chats on the most different subjects and to the Flump (the Imperial College ultimate frisbee team) for the beautiful time spent together on challenging practices and tough matches.

Contents

Introduction	1
1 Emerging and re-emerging infections	5
1.1 Factors driving the emergence of infectious diseases	5
1.2 An overview on the mathematical models used in epidemiology	7
1.2.1 Deterministic models	9
1.2.2 Stochastic models	10
1.2.3 Beyond the homogeneous mixing assumption	11
1.3 Statistical Inference	15
2 Analysis of a vaccine model with cross-immunity: when can two competing infectious strains coexist?	19
2.1 Introduction	19
2.2 Model Formulation	21
2.3 Existence and Stability of Equilibria	24
2.3.1 Disease Free Equilibrium	24
2.3.2 Subtype-i-Only Equilibrium	25
2.3.3 Coexistence Equilibrium	34
2.4 Examples	44
2.5 Discussion	47
2.6 Appendix	50
2.6.1 Computation of the Routh-Hurwitz coefficients	50
2.6.2 Proof of Lemma 2	51
2.6.3 Proof of Lemma 7	52

3	Modelling the Spatial Spread of H7N1 Avian Influenza Virus among Poultry Farms in Italy	55
3.1	Introduction	55
3.2	Data	56
3.3	Models Analysed	58
3.4	Parameter Estimates	60
3.5	Simulations	63
3.5.1	How data are reproduced by the model	63
3.5.2	Assessment of the effectiveness of the intervention measures	65
3.6	Results and Discussion	68
4	A new approach to estimate the spread and transmission of infectious diseases from Sentinel surveillance: application to the 2009-2010 A/H1N1 influenza pandemic in Italy	71
4.1	Introduction	71
4.2	Data	72
4.3	Model Formulation	73
4.3.1	Mathematical model	73
4.3.2	Statistical model	74
4.3.3	Fixed Z_t^i	75
4.3.4	Random Z_t^i	77
4.4	Models definition and parametrisation	78
4.5	Parameter estimation	80
4.6	Results	81
4.7	Discussion	87
4.8	Supplementary Information	90
4.8.1	Data	90
4.8.2	Model formulation	91
4.8.3	Models definition and parametrization	96
4.8.4	Parameter estimation	99
4.8.5	How data are reproduced by the model	100
5	Estimation of R_0 from real and simulated school outbreaks	103
5.1	Introduction	103

5.2	Estimation of exponential growth rate	104
5.3	The individual-based model	108
5.3.1	Parameterization of the individual-based model	110
5.3.2	Computation of the within school reproduction number	110
5.3.3	Analysis of simulated school epidemics	113
5.4	Real school outbreaks	117
5.4.1	The survey	117
5.4.2	Analysis of real school epidemics	117
5.5	First results & discussion	125
	Bibliography	128

CONTENTS

List of Figures

2.1	The flow chart of the model	22
2.2	The function $\gamma_2 = \Psi_2(\gamma_1)$ and the corresponding regions in the plane (γ_1, γ_2) where cases (a) or (b) of Lemma 6 hold. Parameter values are $R_0^1 = 4$, $R_0^2 = 2$, $\mu = 1$, $\xi_1 = 0.9$, $\xi_2 = 0.5$	42
2.3	Equilibrium fractions y_1 and y_2 as function of pe for fixed $\gamma_1 = 0.015$, $\gamma_2 = 0.517$, $R_0^1 = 8.363$, $R_0^2 = 3.790$, $\mu = 0.423$, $\xi_1 = 0.990$, $\xi_2 = 0.020$, $\pi = 1$. Coexistence of the strains occurs for $0.153 < pe < 0.552$	45
2.4	Equilibrium fractions y_1 and y_2 as function of pe for fixed $\gamma_1 = 0.026$, $\gamma_2 = 0.966$, $R_0^1 = 4.723$, $R_0^2 = 2.293$, $\mu = 0.235$, $\xi_1 = 0.923$, $\xi_2 = 0.650$, $\pi = 1$. A bi-stability region occurs for $0.822 < pe < 0.829$; in this region unstable coexistence of the strains occurs.	46
2.5	Trajectories of the fractions y_1 (left panel) and y_2 (right panel) as functions of time; parameter values are $\gamma_1 = 0.026$, $\gamma_2 = 0.966$, $R_0^1 = 4.723$, $R_0^2 = 2.293$, $\mu = 0.235$, $\xi_1 = 0.923$, $\xi_2 = 0.650$, $\pi = 1$ and $pe = 0.8234$. Both trajectories start close to the Coexistence Equilibrium $x = 0.148$, $y_1 = 0.007$, $y_2 = 0.022$, $N = 3.896$; the starting point of the red one, converging to the Subtype-1-Only Equilibrium, is $(0.148, 0.010, 0.029, 3.896)$; the starting point of the blue one, converging to the Subtype-2-Only Equilibrium, is $(0.148, 0.007, 0.029, 3.896)$	47

LIST OF FIGURES

3.1	Infected farms (red dots), not infected farms (green dots), farms banned from restocking (yellow dots) and pre-emptive culled farms (blue dots) in the HPAI epidemic of years 1999-2000 in Italy (left panel) and in the study area (Veneto and Lombardia) (right panel)	58
3.2	Comparison of the number of new cases between the 3-day running average of the observed epidemic and of 100 replicates of the stochastic 2-Phases Susceptibility Model	65
3.3	Comparison of the number of new cases between the 3-day running average of the observed epidemic and the 20th, 40th, 60th and 80th of 100 replicates of the stochastic 2-Phases Susceptibility Model	66
3.4	Status of farms in the study area at time $t = 1$ (top left), $t = 50$ (top right), $t = 100$ (bottom left), $t = 150$ (bottom right) in one simulation of the 2-Phases Susceptibility Model. Yellow dots represent empty farms, green dots represent susceptible units, red dots represent infectious units, blue dots represent (either pre-emptive or previously infected) culled farms.	67
3.5	Distance (km) reached by infection in time (days) in the 1999-2000 epidemic in Italy	70
4.1	Graphical representation of the populations taken into account and notation adopted in the work. The Italian population is considered constant over the whole study period while the monitored patients population changes every week, due to the voluntary nature of the surveillance system. Index i denotes the age-class ($i = 1, \dots, 4$) and index t denotes the week, ranging from week 38 of year 2009 to week 7 of year 2010.	74
4.2	Incidence (per thousand) of the total number of reported ILI cases (black dots) and of the number of reported H1N1-attributable ILI-cases (red dots), obtained by multiplication of the weekly ILI datum times the proportion of positive samples on the corresponding week.	83

4.3	Susceptibility model (in the Basic, Age-Dependent Reporting and Time-Varying Reporting versions without overdispersion and in the Age-Dependent Reporting version with overdispersion parameter estimated from the data) : plot of the simulated weekly reported incidence (per thousand) of H1N1 cases in the 0–4 years age-class (blue), 5–14 years age-class (green), 15 – 64 years age-class (orange), 65+ years age-class (purple) and in the population as a whole (black) in comparison to the respective observed data (red).	86
4.4	Estimated incidence (per thousand) of H1N1 cases in the Italian population using the ADR Susceptibility model with overdispersion parameter r estimated from the data: 0 – 4 years age-class (blue), 5 – 14 years age-class (green), 15 – 64 years age-class (orange), 65+ years age-class (purple), in the population as a whole (black). Predictions have been obtained from the numerical resolution of the SEIR model having fixed the parameters as resulted from 500 draws from the estimated joint posterior distribution.	87
4.5	Plot of the fit of the infectivity function since infection $A(\tau)$ (defined in the SI by equation (4.23)) to the data reported by Baccam et al. (2006)	99
4.6	Immunity model without overdispersion: plot of the simulated weekly reported incidence (per thousand) of the new H1N1 cases in the 0 – 4 years age-class (blue), 5 – 14 years age-class (green), 15 – 64 years age-class (orange), 65+ years age-class (purple) and in the population as a whole (black) in comparison to the respective observed data (red).	102
5.1	Plot of the logarithm of the cumulative function $c(t)$ defined in (5.2) versus time t	106
5.2	Plot of R_s given in (5.13) as a function of the school size N_s having fixed $p_s = \bar{p}_s = 0.65$ and $\Delta_t = 0.5$	113

LIST OF FIGURES

5.3 Plot of the number of new cases (left panel) and of the cumulative number of observed new cases (right panel) in the school of Povo in time, starting from the day of detection of the index case. 118

5.4 Plot of: (a) the incidence data collected in the Povo school (lin-log scale) and the best linear approximation obtained by linear least square fitting to the filled dots, (b) the cumulative data collected in the Povo school (lin-log scale) and the best linear approximation obtained by linear least square fitting to the filled dots, (c) the incidence as a function of the cumulative data collected in the Povo school and the best linear approximation obtained by linear least square fitting to the filled dots. 120

5.5 Plot of the number of new cases (left panel) and of the cumulative number of observed new cases (right panel) in the school of Villazzano in time, starting from the day of detection of the index case. 121

5.6 Plot of: (a) the incidence data collected in the Villazzano school (lin-log scale) and the best linear approximation obtained by linear least square fitting to the filled dots, (b) the cumulative data collected in the Villazzano school (lin-log scale) and the best linear approximation obtained by linear least square fitting to the filled dots, (c) the incidence as a function of the cumulative data collected in the Villazzano school and the best linear approximation obtained by linear least square fitting to the filled dots. 124

List of Tables

3.1	MLE and 95% Confidence Intervals of the Basic Model’s parameters	60
3.2	MLE and 95% Confidence Intervals of the Susceptibility Model’s parameters	61
3.3	MLE of the Basic SEIR Model’s parameters	62
3.4	MLE and 95% Confidence Intervals of the 2-Phases Susceptibility Model’s parameters	63
3.5	Values of the transmissibility constants for the 2-Phases Susceptibility Model	63
3.6	Mean numbers and 5–95 percentile intervals computed on 100 realizations that generated at least 10 cases, using the Basic Model and the 2-Phases Basic Model	64
3.7	Mean numbers and 5 – 95 percentile intervals computed on 100 realizations that generated at least 10 cases, using the Susceptibility Model and the 2-Phases Susceptibility Model . .	64
3.8	Mean numbers and 5 – 95 percentile intervals computed on 100 realizations that generated at least 10 cases, using the 2-Phases Susceptibility Model with and without BR or PEC .	66
3.9	Mean numbers and 5 – 95 percentile intervals computed on 100 realizations that generated at least 10 cases, using the 2-Phases Susceptibility Model with different intervention strategies	68
4.1	Summary of the parameter values fixed and estimated in the models. With the expression “ind.comp.” we mean “indirectly computed” from R_0 , as explained in the main text.	80

LIST OF TABLES

4.2	Susceptibility model without overdispersion: DIC score, mean and equal-tailed 95% credible interval of the marginal posterior distribution of the parameters for each specified model. . .	82
4.3	ADR Susceptibility model with overdispersion: DIC score, mean and equal-tailed 95% credible interval of the marginal posterior distribution of the parameters having fixed the dispersion parameter r to the specified value and having estimated r from the data.	83
4.4	Estimated age-specific peak-incidence (per thousand) and attack rate of H1N1 cases caused by the A/H1N1 virus in the Italian population during the 2009-2010 pandemic as resulted from simulations of the ADR Susceptibility model with estimated overdispersion parameter r having fixed the parameters at the values obtained by 500 draws from the joint estimated posterior distribution. Mean and, in brackets, 5-95 percentile interval.	84
4.5	Symmetrized contact matrix of all reported contacts (physical and non-physical) in Italy, consisting of the average number of contact persons recorded per working day per survey participant (Polymod 2008). Row index represents the age class of the participant, column index represents the age class of the contact.	91
4.6	Symmetrized contact matrix of all reported contacts (physical and non-physical) in Italy, consisting of the average number of contact persons recorded per holiday day per survey participant (Polymod 2008). Row index represents the age class of the participant, column index represents the age class of the contact.	91
4.7	Immunity model without overdispersion: mean and equal-tailed 95% credible interval of the marginal posterior distribution of the parameters for each specified model.	97

4.8 Fixed reporting model with overdispersion ($r = 10$) in the “Basic” and “Age-Dependent Reporting” versions (i.e. having fixed the reporting rates as resulted respectively from the “Basic” and “Age-Dependent Reporting” versions of the “Susceptibility” model): mean and equal-tailed 95% credible interval of the marginal posterior distribution of the parameters; the susceptibility estimated have also been rescaled to the values fixed on Table 1 in the main text for the purpose of comparison. 98

4.9 Sensitivity analysis on the distribution of the initial cases I_0 : mean and equal-tailed 95% credible interval of the marginal posterior distribution of the parameters for the “Age-Dependent Reporting Susceptibility” model without overdispersion. 100

4.10 Susceptibility model: mean and equal-tailed 95% credible interval of the marginal posterior distribution of the parameters for the “Susceptibility” model without overdispersion assuming a mean length of the latent period of 1.3 days and a mean generation time T_g of 2.6 days. 100

5.1 Some basic statistics on the simulated school epidemics. 115

5.2 Simulated school epidemics: threshold values for R^2 , for the different methods L.1, L.2 and L.3. The values reported on Table 5.1 satisfy the constrains here defined. 115

5.3 Ranges (i.e. maximum and minimum value) of R_s estimated using methods L.1, L.2 and L.3 for the relative school epidemics and the theoretical value of R_s given by formula (5.13) on the basis of the data provided on Table 5.1. The selected estimates satisfy the constrains reported on Table 5.2 and the choice of the temporal intervals used to perform linear regression (in the three variants L.1, L.2 and L.3 and for each school epidemic) has been discussed in the text. 116

5.4 Some basic statistics on the survey led in the primary schools of Povo and Villazzano. 117

5.5 Threshold values for R^2 for the estimates obtained for the school of Povo. 118

LIST OF TABLES

5.6 School of Povo: summary of the estimated values of the exponential growth rate r obtained through the fit of $\log_e(i(t))$ vs t and the corresponding R_0 . Linear regression has been performed on the time intervals given by $[t_i, t_i + \delta_t]$. The marked (*) linear fit is plotted on Figure 5.4(a). 119

5.7 School of Povo: summary of the estimated values of the exponential growth rate r obtained through the fit of $\log_e(c(t))$ vs t and the corresponding R_0 . Linear regression has been performed on the time intervals given by $[t_i, t_i + \delta_t]$. The marked (*) linear fit is plotted on Figure 5.4(b). 119

5.8 School of Povo: summary of the estimated values of the exponential growth rate r obtained through the fit of $i(t)$ vs $c(t)$ and the corresponding R_0 . Linear regression has been performed on the time intervals given by $[t_i, t_i + \delta_t]$. The marked (*) linear fit is plotted on Figure 5.4(c). 120

5.9 Threshold values for R^2 for the estimates obtained for the school of Villazzano. 121

5.10 School of Vilazzano: summary of the estimated values of the exponential growth rate r obtained through the fit of $\log_e(i(t))$ vs t and the corresponding R_0 . Linear regression has been performed on the time intervals given by $[t_i, t_i + \delta_t]$. The marked (*) linear fit is plotted on Figure 5.6(a). 122

5.11 School of Vilazzano: summary of the estimated values of the exponential growth rate r obtained through the fit of $\log_e(c(t))$ vs t and the corresponding R_0 . Linear regression has been performed on the time intervals given by $[t_i, t_i + \delta_t]$. The marked (*) linear fit is plotted on Figure 5.6(b). 123

5.12 School of Vilazzano: summary of the estimated values of the exponential growth rate r obtained through the fit of $i(t)$ vs $c(t)$ and the corresponding R_0 . Linear regression has been performed on the time intervals given by $[t_i, t_i + \delta_t]$. The marked (*) linear fit is plotted on Figure 5.6(c). 124

Introduction

Since the early 20th century the study of the spread of infectious diseases has been a theme of deep interest, great importance and often a challenge for human and veterinary medicine. The main objective of epidemiology is the understanding of the leading factors and the complex mechanisms that produce the observed outbreaks in order to provide tools for disease control and prevention, in the interest of public health.

In this context, the use of mathematical models is particularly significant. Indeed, mathematical models can give insight into the understanding of the mechanisms behind the spread of infectious diseases, they are a tool for assessing the effectiveness of control measures and therefore selecting the best strategy to be adopted for the containment of an outbreak, they can be used to assess the efficacy of vaccine treatments, to explore what-if scenarios and to inform policy decisions.

On Chapter 1 I present the actual introduction to this thesis, which consists in a review of the main mathematical tools traditionally used in epidemiology.

On the same chapter I also place my original contributions into the field.

Below I am going to present a brief description of my research work and the specific outline of this thesis.

Research description

The works presented in this thesis are very different one from the other but they all deal with the mathematical modelling of emerging infectious diseases which, beyond being the leitmotiv of this thesis, is an important research area in the field of epidemiology and public health.

A minor but significant part of the thesis has a theoretical flavour. This part

is dedicated to the mathematical analysis of the competition model between two HIV subtypes in presence of vaccination and cross-immunity proposed by Porco and Blower (1998). We find the sharp conditions under which vaccination leads to the coexistence of the strains and using arguments from bifurcation theory, draw conclusions on the equilibria stability and find that a rather unusual behaviour of hysteresis-type might emerge after repeated variations of the vaccination rate within a certain range.

The most of this thesis has been inspired by real outbreaks occurred in Italy over the last 10 years and is about the modelling of the 1999-2000 H7N1 avian influenza outbreak and of the 2009-2010 H1N1 pandemic influenza.

From an applied perspective, parameter estimation is a key part of the modelling process and in this thesis statistical inference has been performed within both a classical framework (i.e. by maximum likelihood and least square methods) and a Bayesian setting (i.e. by Markov Chain Monte Carlo techniques).

However, my contribution goes beyond the application of inferential techniques to specific case studies. The stochastic, spatially explicit, between-farm transmission model developed for the transmission of the H7N1 virus has indeed been used to simulate different control strategies and asses their relative effectiveness. The modelling framework presented here for the H1N1 pandemic in Italy constitutes a novel approach that can be applied to a variety of different infections detected by surveillance system in many countries. We have coupled a deterministic compartmental model with a statistical description of the reporting process and have taken into account for the presence of stochasticity in the surveillance system. We thus tackled some statistical challenging issues (such as the estimation of the fraction of H1N1 cases reporting influenza-like-illness symptoms) that had not been addressed before. Last, we apply different estimation methods usually adopted in epidemiology to real and simulated school outbreaks, in the attempt to explore the suitability of a specific individual-based model at reproducing empirically observed epidemics in specific social contexts.

Structure of the thesis

In the first Chapter of this thesis I present a brief review of the mathematical models adopted in epidemiology for the modelling of emerging infectious diseases and place the works presented in this thesis within the field.

On Chapter 2 we analyse the Vaccine Model with Cross-Immunity proposed by Porco and Blower (1998). Porco and Blower (1998) show that vaccination can shift the competitive balance in favour of a strain that, without vaccination, would be out-competed and that vaccination can also promote coexistence of different strains, something that normally is not expected (Bremermann and Thieme, 1989). Their results have been mainly obtained through numerical simulations, so that the conditions under which a shift in competitive balance or coexistence occurs have not been fully established. We give a rather complete description of its behavior, at least in terms of equilibria. We find the exact conditions under which vaccination may lead to a shift in competitive balance and show that, under these conditions, there always exist a range of vaccination rates under which a coexistence equilibrium exists. We also find that a coexistence equilibrium exists (and is unstable) in a ‘bi-stability’ region, where both monomorphic equilibria are stable. This fact has been rarely observed in models of competition between pathogen strains. The work presented in this chapter has been submitted for publication and is currently under review.

Chapter 3 is about the analysis of the between-farm transmission of the H7N1 highly pathogenic avian influenza virus that disrupted the Italian poultry production in the 1999-2000 epidemic. We define a SEIR model with a spatial transmission kernel, accounting for the containment measures actually undertaken, find significant differences in susceptibility between species and a reduction in transmissibility after the first phase. We performed simulations to assess the effectiveness of the implemented and new control measures. The most effective measure was the ban on restocking. An earlier start of pre-emptive culling promotes eradication; restricted pre-emptive culling delays eradication but causes lower losses. This work has been published on the

Epidemics Journal, 2 (2010): 29-35 (doi:10.1016/j.epidem.2010.01.002).

On Chapter 4 we propose a novel and general modelling framework which allowed us to tackle some statistical challenges that were usually bypassed through the introduction of assumption and has been applied to the 2009-2010 H1N1 pandemic in Italy. The analysis of surveillance data, often the only information available in real time, poses many statistical challenges that have not been addressed yet. For instance, the fraction of cases that report infection is unknown. We propose here a general modelling framework that explicitly takes into account the way the surveillance data was generated. Our approach couples a deterministic mathematical model with a statistical description of the reporting process and has been applied to surveillance data collected in Italy during the 2009-2010 A/H1N1 influenza pandemic. We estimate that the reproduction number R_0 has been into the range 1.3–1.4, that the youngest age-classes reported the symptoms caused by the H1N1 virus infection significantly more than the adults and that, in the Italian population, school-age children were effectively the most affected by the A/H1N1 virus. In terms of both estimated peak-incidence and attack rate of A/H1N1 cases, the 5 – 14 years age-class was about 5 times more affected than the 65+ years old age-group and about twice more than the the other age-classes; the overall case attack rate was about 30%.

The fifth and final Chapter is about the first results of a topic I started working on only very recently. The chapter deals with the estimation of the reproduction number from real and simulated school outbreaks data. In this context, we explore whether an individual-based model recently developed to model the spatio-temporal spread of the pandemic H1N1 virus in Europe (Merler and Ajelli, 2010), used here as a tool for generating within school outbreaks, gives compatible results (in terms of estimated within school reproduction number) with real school epidemics observed in Italy over the past 2009-2010 pandemic influenza season. The real school outbreaks in question have been retrospectively reconstructed through a survey and this topic is presented in the last part of the chapter.

Chapter 1

Emerging and re-emerging infections

1.1 Factors driving the emergence of infectious diseases

An emerging pathogen can be defined as an infectious agent whose incidence or geographic range is increasing following its first introduction into a new host population; a re-emerging pathogen is one whose incidence or geographic range is increasing in an existing host population as a result of long-term changes in its underlying epidemiology (Morse, 1995; Woolhouse, 2002).

Pathogen emergence can be based on subjective criteria, which can reflect increased awareness, improved diagnosis, discovery of previously unrecognized infectious agents as much as any objective epidemiological data. This is why Woolhouse (2002) suggests that reporting bias must be considered as a possible explanation for any apparent pattern. Indeed, despite dozens of pathogen species are regarded as emerging or re-emerging in livestock, domestic animal and wildlife, data for non-human hosts are likely to be far less comprehensive than those for humans (Woolhouse, 2002).

Infectious disease emergence can be viewed operationally as a two-step process consisting in the introduction of the infectious agent into a new host population followed by the establishment and further dissemination of the in-

fectious agent within the new host population (also called “adoption”) (Morse, 1995).

Broadly, there are three sources of emerging and re-emerging pathogens: from within the host population itself (like *Mycobacterium tuberculosis*, whose re-emergence in the 1980s was fuelled by the immune-deficiencies of people with AIDS), from the external environment (like *Legionella pneumophila*, whose emergence as a human pathogen might not have occurred were it not for the environmental niche provided by air-conditioning systems) and from populations of other host species (like the Human Immunodeficiency Virus (HIV) in humans). Many wildlife species are reservoirs of pathogens that threaten domestic animal, human health and the conservation of the global biodiversity as well (Daszak et al., 2000). Using the WHO definition, zoonotic pathogens are defined as those producing diseases or infections which are naturally transmitted between vertebrate animals and humans. Bats, carnivores, primates, rodents, ungulates and other mammals and non-mammals (birds, reptiles, amphibians and fish) constitute the broad categories into which we can split the “zoonotic pool”. Three-quarters of emerging and re-emerging human pathogens originate as zoonose (Woolhouse, 2002) and are disproportionately viruses (Woolhouse and Gowtage-Sequeria, 2005; Woolhouse et al., 2005).

Several (not mutually exclusive) factors drive the emergence of infectious diseases: genetic changes in the pathogen (for example the evolution of HIV from the simian immunodeficiency virus), immunocompromised hosts (for example *M. tuberculosis* in AIDS patients) and changes in host-pathogen ecology. This last category includes changes in host demography, movement or behaviour; climate, agricultural changes or changes in the land use; changes in industry and technology (e.g. food production); international travels and commerce or the breakdown of public health measures (Morse, 1995; Morens et al., 2004; Racaniello, 2004). It’s worth noticing that suprisingly often disease emergence is caused by human actions.

In the following chapters we propose, analyse and apply different modelling approaches to the spread of HIV, of the Highly Pathogenic Avian Influenza (HPAI) H7N1 virus and to the recent H1N1 virus that caused the 2009-2010 influenza pandemic: these viruses are recognized as the cause of emerging or

re-emerging infectious diseases. Despite the precise ancestry of HIV is still uncertain, it appears to have had zoonotic origins. Genetical changes are the leading factors of many influenza pandemics too. Influenza A viruses, which endemically live in the gastrointestinal apparatus of wild waterfowl, have indeed evolved elaborate mechanisms to jump species into domestic fowl, farm animals and humans. It is widely known that antigenic drift (point mutations, primarily in the gene for the surface protein, hemagglutinin) causes annual or biennial influenza epidemics and antigenic shift (genetic reassortment generally between avian and mammalian influenza strains) caused the emergence of pandemic influenza strains as in 1888, 1918, 1957, 1968 and in the recent swine-origin H1N1 2009-2010 pandemic (Webster, 2001; Shortridge et al., 2003; Neumann et al., 2009).

The enormous global burden in terms of human and animal disease and deaths posed by emerging and re-emerging pathogens makes the study of emerging and re-emerging infections a challenge for human and veterinary medicine.

1.2 An overview on the mathematical models used in epidemiology

An impressive amount of works flourished recently, given the emergence of dramatic disease outbreaks such as the foot-and-mouth-disease (FMD) outbreak of 2001 in the British cattle farms, the severe acute respiratory syndrome (SARS) outbreaks of 2003 in Asia and Canada and the recent 2009-2010 influenza pandemic caused by the A/H1N1 virus. Mathematical models are a useful tool that can give insight into the understanding of the leading factors and mechanisms behind the spread of infectious diseases (Anderson and May, 1992; Diekmann and Heesterbeek, 2000; Fraser et al., 2004) and have been used in the past to design efficient observational studies and to plan mass vaccination campaigns (Grassly et al., 2006; Yang et al., 2006). It is widely recognized that they are a valuable tool to investigate the effectiveness of control measures, to assess the efficacy of vaccine and prophylactic treatments and to explore what-if scenarios so that they have also been employed

1.2. An overview on the mathematical models used in epidemiology

to inform policy decisions (Ferguson et al., 2001; Halloran et al., 2007).

The definition and construction of a useful model usually depends on the issues the modeller wants to tackle and requires knowledge of a variety of different aspects, from the biological to the epidemiological and demographic ones. Whether the interest of the modeller is focused on the theoretical properties of a model or the mathematical framework is used to make inference on unknown quantities, model validation is an always desirable but often unfeasible stage, since it requires quantitative data, usually collected by surveillance systems. The successful application of statistical, mathematical and computational techniques for the analysis of outbreak data and the choice of a suitable model framework strongly depends on the nature and availability of information at all the levels (biological, demographic, epidemiological) outlined above. As a matter of fact, limited data and inconclusive epidemiological information place severe restrictions on the efforts the modeller can make to model the spread of the etiological agent if his/her objectives go beyond the intrinsic interests and the theoretical exploration of the model's behaviour.

In this section I would like to propose a brief and far from exhaustive review of the most common mathematical modelling approaches that have been undertaken in infectious disease epidemiology in presence of outbreaks caused by emerging or re-emerging infectious entities both in the human and in some animal populations.

The mathematical models we will deal with in this section and, at a broader extent, in this thesis, are compartmental models at a population level. It means that our interest is focussed on the dynamics of disease spread among the individuals of a population rather than on the processes occurring within the host after infection by the infectious agent. Epidemiological models of disease spread at a population level split the population into compartments that usually characterize the infectious state (e.g. susceptible, infectious, recovered individuals) and can also include other forms of partitioning (e.g. vaccinated, treated, hospitalized, quarantined individuals).

Within the wide variety of existing models, a first distinction can be made between the deterministic and the stochastic approach.

Another distinction can be made in terms of the level of mixing so that we

can distinguish models assuming homogenous mixing from models adopting more heterogeneous contact patterns between the individuals of a population.

A comprehensive introduction and an outline of the development of mathematical modelling of infectious diseases can be found in the texts by Anderson and May (1991), Bailey (1975) and the more recent works by Hethcote (2000) and Keeling and Rohani (2008).

1.2.1 Deterministic models

The history of epidemic modelling can be traced back to the early 20th century, when the deterministic approach first appeared in the literature (Hamer, 1906; Ross, 1916; Ross and Hudson, 1917a,b; Bailey, 1975) and culminated with the milestone and still relevant work by Kermack and McKendrick (1927).

The description of a phenomenon is often translated, in mathematical terms, into a set of differential equations. The theory of differential equations is a well established branch of mathematics in which both theoretical results and numerical methods have been readily available since the early 20th century. A deterministic model is characterized by the fact that, once that the initial conditions and the parameter values have been fixed, its evolution is uniquely determined. The successful application of the deterministic approach in the field of epidemiology lies in the relative flexibility and contemporary mathematical tractability of the modelling framework. Deterministic models can indeed be enriched to account for realistic features such as, for example, the presence of different stages of infection, age-structure, spatial spread and vertical transmission, without completely losing their analytical tractability. Deterministic models have been employed to perform parameter estimation and fit surveillance data (Chowell et al., 2003; Wang and Ruan, 2004; Nishiura et al., 2010) and to assess the impact of control measures in the SARS epidemic (Lipsitch et al., 2003), to investigate how best to use antibiotics in populations harbouring drug-resistant organisms (Austin et al., 1997) and for the analysis of the transmission dynamics of multiple strains pathogens (Gupta et al., 1998; Andreasen et al., 8 15) causing infectious diseases such

1.2. An overview on the mathematical models used in epidemiology

as malaria (Gupta et al., 1994), dengue (Ferguson et al., 1999) and influenza (Minayev and Ferguson, 2009). Multi-strain models have been often analysed through the use of computer simulations and a theoretical analysis of the equilibria and the relative stability of a multi-strain model has been proposed by Gog and Grenfell (2002). The effect of different vaccination policies in the presence of two competing HIV strains conferring partial cross-immunity has been proposed and numerically analysed by Porco and Blower (1998, 2000) and it is in this framework that the work presented in Chapter 2 can be placed. In the next chapter we indeed present a mathematical analysis (in terms of equilibria and their stability) of the 2-HIV strains competition model proposed in (Porco and Blower, 1998).

Deterministic models are rapid to simulate, relatively easy to parametrize and capture the average epidemic behaviour, i.e. they can be considered a valid tool for predictions in large populations. On the contrary, in presence of low levels of infections (i.e. near the start and the end of an epidemic) or of small populations, the deterministic approach fails to catch the random nature of transmission events. Another limitation of the deterministic approach consists in an oversimplified description of the interactions between individuals. Due to mathematical convenience, it is indeed assumed that either an individual has an equal chance of contacting anyone in the population (homogeneous mixing at the population level) or random mixing occurs between each pairs of subgroups into which the population is structured.

1.2.2 Stochastic models

Stochastic models can be considered the counterpart of the deterministic case, are particularly apt to model the spread of a disease in small populations or in the early and final stages of an epidemic (i.e. when the number of cases is small) and can be added of various forms of heterogeneity in contact patterns. In the stochastic modelling framework the modeller attributes a probability of occurrence to each single event and counts (in terms of discrete units) their occurrence. The study of the evolution (in time) of the probabilities of the state of the system and the investigation of the distribution of quantities of interest is much more complex. Due to the analytical complexity

of the study of a stochastic process, often increased by the need (or wish) to include very detailed information on the contact structures of a population, computer simulations offer an alternative approach to explore the behaviour of the system. Such an example is given by Cooper et al. (1999), where computer simulations are used to explore the properties and the behaviour of a stochastic compartmental model that had been set up to study the spread of hand-borne nosocomial pathogens within a general medical-surgical ward. Other examples of stochastic models used in epidemiology are given by (Keeling et al., 2001; Riley et al., 2003; Chis Ster and Ferguson, 2007) and many others more extensively discussed in the rest of the chapter.

1.2.3 Beyond the homogeneous mixing assumption

It is widely recognized that heterogeneity in contact patterns due for example to age differences between individuals, the spatial distribution of individuals and the presence of social structures in the population play an important role in disease spread. Structured, microsimulation, meta-population, network models and models with multiple levels of mixing are all examples of population models that attempt to achieve an increased realism by going beyond the rather unrealistic homogeneous mixing assumption. In principle, all the models presented below can be placed within both a deterministic and a stochastic framework; as a matter of fact, the most of the works reported as examples belong to the stochastic category.

Microsimulation or individual-based models

Microsimulation models, also called individual-based models, are stochastic simulations of contact patterns and disease progression operating at an individual level, where the individual unit needs to be specified by the modeller. The 2001 UK foot-and-mouth epidemic generated a unique data-set describing the spatial spread of the infection between livestock farms and offered the opportunity to explore, mainly using micro-simulation models, the impact of spatial and individual heterogeneities on the course of the epidemic (Keeling et al., 2001; Morris et al., 2001; Chis Ster and Ferguson, 2007). The full spatio-temporal dynamics of the foot-and-mouth disease (FMD) epidemic

1.2. An overview on the mathematical models used in epidemiology

has indeed been explored by Keeling et al. (2001) and Chis Ster and Ferguson (2007) using stochastic, spatial, individual farm-based models incorporating heterogeneity in farm size and species composition.

The transmission of the highly virulent H5N1 influenza virus to humans in South-East Asia triggered the development of individual-based models investigating the strategies to be applied for containing an emerging influenza pandemic (Ferguson et al., 2005, 2006; Longini et al., 2005; Germann et al., 2006). Individual-based model have also been employed to understand the role of population heterogeneity and human mobility in the spread of the recent 2009-2010 H1N1 pandemic influenza virus (Merler and Ajelli, 2010).

Network models

The historical study of networks has its bases in two disparate fields: social sciences and graph theory. While the research in graph theory and social sciences generally considers an understanding of the network itself to be the ultimate goal, in epidemiology the interest is focused on the spread of the disease, in which case the network forms a constraining background to the transmission dynamics.

In order to understand the role of network structure on epidemic dynamics, a range of idealized networks, defined in terms of how individuals are distributed in space and how connections are formed, have been developed and analyzed. Random networks, lattices, small-world, spatial and scale-free networks have been used to describe different aspects of the population mixing behaviour, on the basis of the different levels of clustering, degree distribution and path length, intrinsically defined by the network structure itself (Keeling and Eames, 2005).

The spread of infection on generic networks can also be modelled through the pairwise approximation which, as the name suggests, takes the number of different pair types as variable of the model and requires some form of moment closure approximation. A pairwise model has been used for example to provide real-time predictions during the 2001 foot-and-mouth epidemic in the UK (Ferguson et al., 2001) and the extent to which the ensemble behaviour of stochastic spatial epidemic models may be captured by modelling disease processes as occurring on networks derived from the underlying spa-

tial structure has been deeply analysed by Parham and Ferguson (2006) and Parham et al. (2008). As a matter of fact, the differential equation formulation of pairwise models represents a more rapid parametrization alternative to computationally intensive microsimulation models and may be amenable to obtain an analytical understanding into spatio-temporal dynamics.

In the presence of an emerging infection, three techniques have been mainly employed to gather network information: infection tracing, contact tracing and diary-based studies. Infection tracing consists in the reconstruction of the transmission network (or epidemic-tree), consisting in all the links through which transmission occurred. Such tracing has been employed for example to analyze the foot-and-mouth disease outbreak of 2001 (Haydon et al., 2003), to gather information about the individuals most involved in disease transmission (the so called “super-spreader”) during the 2003 SARS outbreak in Hong Kong (Riley et al., 2003) and to investigate the transmission properties of the new H1N1 strain in the first few hundred cases study in England, Wales and Scotland in the 2009-2010 influenza pandemic (Ghani et al., 2009). Contact tracing aims to identify all potential transmission contacts from a source individual, has been largely applied in the study of sexually transmitted diseases and relies on individuals providing complete and accurate data about personal relationships. At a farm level, explicit contact structures have been used to analyze the spread of animal diseases such as the foot-and-mouth disease (Ferguson et al., 2001; Green et al., 2006; Kao et al., 2006) and avian influenza (Le Menach et al., 2006).

Models with multiple levels of mixing

Household models are a natural starting point if the attempt of the modeller is to include a more realistic social structure than those assumed under the homogeneous mixing hypothesis, still remaining within an analytically tractable (but rather more complex) framework. Household models partition the population into households and homogeneous mixing within each household is superimposed on homogeneous mixing (typically at a smaller rate) in the population at large (Becker, 1995; Hall and Becker, 1996; Ball et al., 1997). Most effort has been typically placed into analysing, within a probabilistic framework, the asymptotic behaviour, the epidemic final sizes

1.2. An overview on the mathematical models used in epidemiology

and the impact of targeted intervention strategies such as vaccination (Ball et al., 1997). Recently, household models have been fruitfully applied to approximate the disease dynamics of an influenza pandemic (Dodd and Ferguson, 2007; Fraser, 2007) and to explore the effectiveness of public health intervention scenarios (Wu et al., 2006; House and Keeling, 2009).

An even more realistic social structure is achieved when individuals belong to more than one type of mixing group and different groups are allowed to overlap. Such a generalization of the households model is given for example in the two (i.e. households and workplaces) levels of mixing models defined by Ball and Neal (2002) and Pellis et al. (2009).

Patch models

Patch or metapopulation models are characterized by the presence of a large population which is divided into a finite number of groups, also called patches. Within each group individuals are assumed to mix homogeneously and the different patches can be connected either deterministically or randomly. Patches usually represent geographical areas at various spatial scales (Rvachev and Longini, 1985; Hollingsworth et al., 2006; Colizza et al., 2007; Rizzo et al., 2008; Balcan et al., 2009) and the connections among the patches typically represent the existing transportation, air travel or commuting network. In comparison to individual-based models, metapopulation models require less information and computational effort so that they can be placed within an inferential framework. One of the first metapopulation models has been proposed by Rvachev and Longini (1985) to describe the global spread of influenza. In this work the authors consider 52 big cities of the world interconnected via air transport. This model has been later updated (Grais et al., 2003), revisited and extended to evaluate the effectiveness of intervention strategies as travel restrictions (Flahault et al., 2006). Stochastic metapopulation models have been proposed by Riley et al. (2003) to model the 2003 SARS outbreak in Hong Kong and by Colizza et al. (2007) and Balcan et al. (2009) to investigate the role played by the airline transportation system versus the short range connections in the global spatio-temporal spread of an influenza pandemic. A first comparison between individual-based and metapopulation models has been recently proposed by Ajelli et al.

(2010). The good agreement between the two modelling frameworks (in terms of epidemic profile and spatio-temporal patterns) could be the first step towards the future development of hybrid models combining the computational efficiency of patch models to the high detail resolution provided by the individual-based approach in specific locations of interest.

1.3 Statistical Inference

Statistical inference is the process of drawing conclusions from data that are subject to random variation due to the nature of the phenomenon itself, observational errors or sampling variation and is based on the definition of a probabilistic model that usually provides a simplified but adequate representation of the phenomenon.

Two statistical approaches can be distinguished: the non-parametric one aims at estimating the distribution underlying the phenomenon under minimal assumptions, generally using functional estimation. Conversely, the parametric approach represents the distribution of the observations through a density function in which only the parameter is unknown.

A parametric statistical model consists of the observation of a random variable x , distributed according to $f(x|\theta)$ where only the parameter θ is unknown and belongs to a vector space Θ of finite dimension. Making inference on parameter θ means that we use observation x to improve our knowledge on parameter θ . Compared with probabilistic modelling, statistical analysis has fundamentally an inversion purpose which is obvious in the notion of the likelihood function $l(\theta|x)$ (a function of the unknown θ given the observed value x) which is just the sample density $f(x|\theta)$ rewritten in the “proper” order

$$l(\theta|x) = f(x|\theta)$$

Within the parametric approach, statistical inference on the unknown parameter can be performed either within a classical (or frequentist) framework or within a Bayesian setting. In the next few lines I am going to briefly recall the two approaches. Indeed, a significant portion of the work presented in this thesis deals with parameter estimation which has been conducted within

both the frequentist approach (on Chapter 3 and 5) and the Bayesian framework (on Chapter 4).

The classical approach makes inference on the unknown parameter by the method of maximum likelihood which was promoted by R.A. Fisher in his classical 1925 paper. Once fixed the underlying probability model $f(x|\theta)$, the method of maximum likelihood selects the values of the model parameter $\hat{\theta}$ that produce the distribution most likely to have resulted in the observed data (i.e. the parameters that maximize the likelihood function)

$$\hat{\theta} = \arg \max_{\theta \in \Theta} l(\theta|x)$$

The maximum likelihood method is widely applied partly because of the intuitive motivation of maximizing the probability of occurrence and partly because of the strong asymptotic properties of the maximum likelihood estimator (consistency, normality, efficiency) and functional invariance (i.e. for any function $h(\theta)$ the maximum likelihood estimator of h is $h(\hat{\theta})$) (Becker, 1989; Zacks, 1971). Maximum likelihood has been applied to a wide range of epidemiological models facing a variety of different problems (Keeling et al., 2001; Boender et al., 2007; Le Menach et al., 2006; Nishiura et al., 2009; Lessler et al., 2009). Also microsimulation models have been placed within a maximum-likelihood inference scheme for example by Pelupessy et al. (2002) to identified the most important routes of transmission of resistant pathogens among the patients of a hospital and by Matthews et al. (2006) to investigate the reasons underlying the substantial variations in the on-farm prevalence of *E.Coli* O157 both between farms and between sampling events on the same farm observed in a cross-sectional study conducted on Scottish cattle farms between 1998 and 2000. Despite the increasing computational power available today, the dimensionality of the problem (i.e. the size of the simulated population) poses severe restrictions on the extent to which one can use individual-based models to make inference on unknown quantities. As a matter of fact, it is nowadays unfeasible to place very complex individual-based models acting on large populations (of the order of million of individuals) within any inference scheme and the effort of the modeller is hence focused on the parametrization of the model. We applied the maximum likelihood

theory to estimate the unknown parameters and the relative confidence intervals of a stochastic spatially-explicit model for the farm-to-farm transmission of the highly pathogenic H7N1 avian influenza virus in Italy (see Chapter 3). In this context, we can insert the least square method, also known as trajectory matching method (Turchin, 2003). The parameter values are estimated by minimising the sum of the squares of the residuals, meant as the difference between the observed and simulated epidemics and can be interpreted as a maximum likelihood criterion if the measurement errors are assumed to be normally distributed. From this point of view, it represents a simplified maximum likelihood approach, in which the dynamics of the epidemic are simulated and then the likelihood of the observed data is evaluated. The least square method has often implemented to perform parameter estimation especially in deterministic settings (Chowell et al., 2006, 2004, 2007b,a) and is the inferential method at the basis of the whole Chapter 5, where we estimate the within school reproduction number of real and simulated school outbreaks.

The main difference brought by the Bayesian approach is to consider a probability distribution on the parameters. By definition (Robert, 1996) a Bayesian statistical model is made of a parametric model $f(x|\theta)$ and a *prior* distribution on the parameters, $\pi(\theta)$. Within this framework, statistical inference is based on the distribution of θ conditional on x

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int f(x|\theta)\pi(\theta)d\theta}$$

which is called *posterior* distribution. By Bayes's Theorem, the information on θ is actualized with the information contained in the observation x . Notice that, from a Bayesian viewpoint, there is little difference between observations and parameters, since conditional manipulations allow for an interplay of their respective roles. Whenever the posterior distribution $\pi(\theta|x)$ cannot be directly simulated, inference on the posterior distribution can be obtained by Markov Chain Monte Carlo (MCMC) methods, which are able to construct Markov chains whose stationary distribution is the distributions of interest. Markov Chain Monte Carlo techniques have been often employed in the field of mathematical epidemiology and examples are given by (Lip-

1.3. Statistical Inference

sitch et al., 2003; Chis Ster and Ferguson, 2007; Cauchemez et al., 2009a). On Chapter 4 we present a novel modelling approach which has been applied to the recent 2009-2010 H1N1 influenza pandemic in Italy. In this work we couple together a deterministic description of the infection dynamics with a statistical model for the reporting process where, by the application of Markov Chain Monte Carlo techniques, we obtain the estimates (in terms of posterior distribution) of epidemiological relevant parameters such as the reproduction number R_0 , the age-dependent reporting rates and susceptibility.

Chapter 2

Analysis of a vaccine model with cross-immunity: when can two competing infectious strains coexist?

2.1 Introduction

Control policies of infectious diseases can lead to unexpected outcomes when the infectious agents consist of a variety of different strains. In fact, it has often be argued that more pathogenic strains are in competition with less pathogenic ones (Bremermann and Thieme, 1989), so that the application of control policies may shift the competitive balance in favour of the less fit strains (McLean, 1995; Martcheva et al., 2008) that might however be more virulent.

It has been observed in previous studies and in practice that vaccination, one of the most powerful control policies, can have very dramatic effect on the outcome of the competition between more pathogens. This topic has been examined with the use of mathematical models in several papers (Porco and Blower, 1998; Lipsitch, 1999; Iannelli et al., 2005; Martcheva, 2006).

Vaccination can destabilize the existing host-pathogen evolutionary equilibria, accelerate pathogen evolution and also lead to the emergence or domi-

nance of a once-rare pathogen, a mechanism also known as strain replacement (Porco and Blower, 2000; Iannelli et al., 2005; Martcheva, 2006). Vaccines differ for their mode of action; vaccines with differential effectiveness provide different degrees of protection against infection by the different strains of the pathogen and their efficacy has been extensively discussed in the literature (Smith et al., 1984; Halloran et al., 1992; Blower and McLean, 1994; Martcheva, 2006).

Porco and Blower (1998) showed that vaccination can indeed shift the competitive balance in favour of a strain that, without vaccination, would be out-competed and that vaccination can also promote coexistence of different strains, something that normally is not expected (Bremermann and Thieme, 1989). The results by Porco and Blower (1998) have been mainly obtained through numerical simulations, so that the conditions under which a shift in competitive balance or coexistence occurs have not been fully established.

Here we examine in detail the “Vaccine Model with Cross-Immunity” or “Differential Degree Model” proposed in Porco and Blower (1998) to describe the spread of 2-HIV strains and the subsequent progression into AIDS in a population of potential sex partners. More in general, the model can be thought as describing the spread of two competing pathogens within a population in presence of vaccination and cross-immunity. We analyse the impact of vaccination at the community level and give a rather complete description of the model behavior, at least in terms of equilibria. We find the exact conditions under which vaccination may lead to a shift in competitive balance and also show that, under these conditions, there always exist a range of vaccination rates under which a coexistence equilibrium exists. We find that the Coexistence Equilibrium may be stable or unstable, depending on another condition. The former case corresponds to what had already been observed numerically. In the latter case, the parameter region in which a coexistence equilibrium exists is actually a ‘bi-stability’ region in which both monomorphic equilibria are stable, so that asymptotic behavior depends on initial conditions. This fact, that would lead to a sort of hysteresis cycle if vaccination rates were increased then decreased, has rarely been demonstrated in models of competition between pathogen strains.

2.2 Model Formulation

The “Vaccine Model with Cross-Immunity” proposed by Porco and Blower (1998) is a particular transmission dynamics model of HIV in presence of two subtypes and a vaccine that provides a degree of protection against infection by both subtypes.

The state variables are X (the number of susceptible individuals), V (the number of effectively vaccinated individuals), Y_1 and Y_2 (the number of individuals infected with subtype 1 and subtype 2 respectively and have not developed AIDS), A_1 and A_2 (the number of individuals who have been infected with subtype 1 and subtype 2 respectively and have developed AIDS). The state variables are supposed to be C^1 functions of the time variable t . Individuals are part of a community of potential sex partners and we assume that individuals with AIDS do not acquire new sex partners. This means that the sexually active community N is given by $N = X + V + Y_1 + Y_2$. We assume that individuals enter the community at a constant rate π and a fraction p of these are vaccinated. The vaccine induces a protective immune response in a fraction e of the vaccinated individuals, that is the vaccine takes only in a fraction pe of the new entries.

Uninfected individuals either not vaccinated or who were vaccinated but in whom the vaccine did not take, are referred as being completely susceptible. The degree of protection conferred by the vaccine against subtype i is indicated with ξ_i ($0 \leq \xi_i \leq 1$); $\xi_i = 0$ corresponds to no protection and $\xi_i = 1$ corresponds to complete protection against infection.

Individuals leave each class at a constant per capita rate μ when they cease acquiring new sex partners.

The transmission probability of subtype i per partnership is indicated with β_i , the number of new sex partners per unit time is indicated by c , γ_i is for the rate of progression to AIDS and α indicates the death rate due to AIDS. The flow diagram in Figure 2.1 describes the dynamics of the “Vaccine Model with Cross-Immunity”.

2.2. Model Formulation

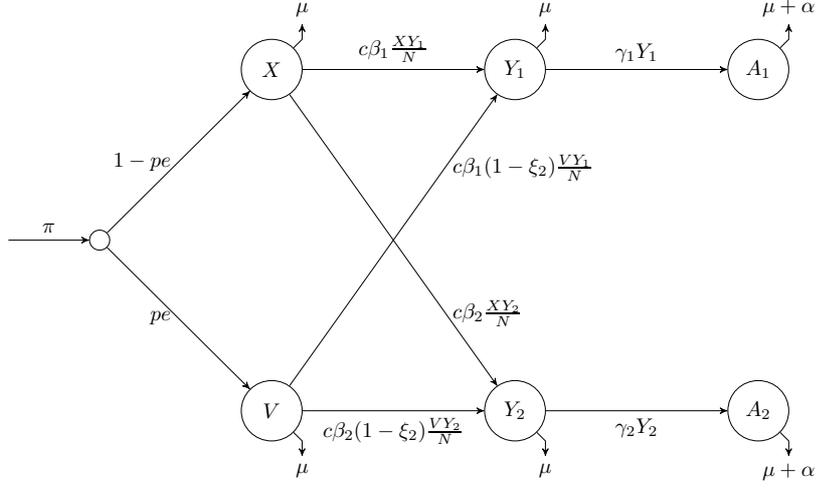


Figure 2.1: The flow chart of the model

The differential equations describing the corresponding dynamics are:

$$\dot{X} = \pi(1 - pe) - \mu X - c\beta_1 X \frac{Y_1}{N} - c\beta_2 X \frac{Y_2}{N} \quad (2.1)$$

$$\dot{V} = \pi pe - \mu V - (1 - \xi_1)c\beta_1 V \frac{Y_1}{N} - (1 - \xi_2)c\beta_2 V \frac{Y_2}{N} \quad (2.2)$$

$$\dot{Y}_1 = c\beta_1 X \frac{Y_1}{N} + (1 - \xi_1)c\beta_1 V \frac{Y_1}{N} - (\mu + \gamma_1)Y_1 \quad (2.3)$$

$$\dot{Y}_2 = c\beta_2 X \frac{Y_2}{N} + (1 - \xi_2)c\beta_2 V \frac{Y_2}{N} - (\mu + \gamma_2)Y_2 \quad (2.4)$$

$$\dot{A}_1 = \gamma_1 Y_1 - (\mu + \alpha)A_1 \quad (2.5)$$

$$\dot{A}_2 = \gamma_2 Y_2 - (\mu + \alpha)A_2 \quad (2.6)$$

where $N = X + V + Y_1 + Y_2$.

We observe that equations (2.1)–(2.4) are sufficient to describe the behavior of the system. Furthermore, these equations can be suitable for any infection of *SI* type, where γ_1 and γ_2 denote disease-induced mortality rates, and vaccination occurs at birth. The assumption of a constant (independent of population size) input rate π in the population may then need to be amended. We introduce the reproduction numbers (Anderson and May, 1991)

$$R_0^1 = \frac{c\beta_1}{\mu + \gamma_1} \quad R_0^2 = \frac{c\beta_2}{\mu + \gamma_2}. \quad (2.7)$$

Then, performing the change of variables

$$x = \frac{X}{N} \quad v = \frac{V}{N} \quad y_1 = \frac{Y_1}{N} \quad y_2 = \frac{Y_2}{N}$$

and using R_0^1 and R_0^2 as parameters, system (2.1)–(2.4) can be equivalently written as

$$\dot{x} = \frac{\pi}{N}(1 - x - pe) - x [(R_0^1(\mu + \gamma_1) - \gamma_1)y_1 + (R_0^2(\mu + \gamma_2) - \gamma_2)y_2] \quad (2.8)$$

$$\dot{v} = \frac{\pi}{N}(pe - v) - v [(R_0^1(\mu + \gamma_1)(1 - \xi_1) - \gamma_1)y_1 + (R_0^2(\mu + \gamma_2)(1 - \xi_2) - \gamma_2)y_2] \quad (2.9)$$

$$\dot{y}_1 = y_1 \left[R_0^1(\mu + \gamma_1)(x + (1 - \xi_1)v) - \gamma_1(1 - y_1) + \gamma_2 y_2 - \frac{\pi}{N} \right] \quad (2.10)$$

$$\dot{y}_2 = y_2 \left[R_0^2(\mu + \gamma_2)(x + (1 - \xi_2)v) - \gamma_2(1 - y_2) + \gamma_1 y_1 - \frac{\pi}{N} \right] \quad (2.11)$$

$$\dot{N} = \pi - N(\mu + \gamma_1 y_1 + \gamma_2 y_2) \quad (2.12)$$

In (2.8), we have dropped the dependency on c , β_1 and β_2 using instead the non-dimensional quantities R_0^1 and R_0^2 as parameters. It would be possible to reduce the parameters to a smaller number of non-dimensional quantities; we prefer to keep them all, while later showing that different behaviours depend on the ratios μ/γ_i and γ_1/γ_2 .

By adding together (2.8)–(2.11) we get

$$\dot{x} + \dot{v} + \dot{y}_1 + \dot{y}_2 = \left(\frac{\pi}{N} - \gamma_1 y_1 - \gamma_2 y_2 \right) [1 - (x + v + y_1 + y_2)] \quad (2.13)$$

Since

$$x + v + y_1 + y_2 = 1$$

is invariant for (2.8)–(2.11), as intuitively obvious, we can drop (for instance) the equation for v and consider the system

$$\begin{cases} \dot{N} = \pi - N(\mu + \gamma_1 y_1 + \gamma_2 y_2) \\ \dot{x} = \frac{\pi}{N}(1 - x - pe) - x [(R_0^1(\mu + \gamma_1) - \gamma_1)y_1 + (R_0^2(\mu + \gamma_2) - \gamma_2)y_2] \\ \dot{y}_1 = y_1 \left[R_0^1(\mu + \gamma_1)(x + (1 - \xi_1)[1 - (x + y_1 + y_2)]) - \gamma_1(1 - y_1) + \gamma_2 y_2 - \frac{\pi}{N} \right] \\ \dot{y}_2 = y_2 \left[R_0^2(\mu + \gamma_2)(x + (1 - \xi_2)[1 - (x + y_1 + y_2)]) - \gamma_2(1 - y_2) + \gamma_1 y_1 - \frac{\pi}{N} \right]. \end{cases} \quad (2.14)$$

We can then obtain the fraction of vaccinated individuals by subtraction

$$v = 1 - (x + y_1 + y_2).$$

2.3 Existence and Stability of Equilibria

We study here the equilibria of (2.14); when this makes the derivation shorter, we will consider also (2.8)–(2.11).

Note first that, from (2.12), any steady state $(\tilde{x}, \tilde{v}, \tilde{y}_1, \tilde{y}_2, \tilde{N})$ of (2.14) satisfies

$$\frac{\pi}{\tilde{N}} = \mu + \gamma_1 \tilde{y}_1 + \gamma_2 \tilde{y}_2. \quad (2.15)$$

2.3.1 Disease Free Equilibrium

Existence

The Disease Free Equilibrium (DFE) occurs when the fraction of infected individuals is null $y_1^* = y_2^* = 0$ and there are positive fractions of susceptible and vaccinated individuals $x^* \neq 0, v^* \neq 0$.

From (2.15), we obtain $N^* = \frac{\pi}{\mu}$. Setting the right-hand side of (2.8)–(2.9) equal to 0 with $y_1^* = y_2^* = 0$, we immediately obtain for the DFE

$$x^* = 1 - pe \quad \text{and} \quad v^* = pe.$$

This is always a feasible solution under the constraints $0 \leq p \leq 1, 0 \leq e \leq 1$ arising from their definition. Otherwise said, the DFE always exists.

Stability

We study the local stability of the DFE through the Jacobian matrix of system (2.14) at the DFE $(x^*, 0, 0, N^*) = (1 - pe, 0, 0, \frac{\pi}{\mu})$. The eigenvalues of the Jacobian at the DFE are

$$\begin{aligned} \lambda_1^* &= -\mu \\ \lambda_2^* &= -\mu \\ \lambda_3^* &= (\mu + \gamma_1) (R_0^1(1 - pe) + (1 - \xi_1)R_0^1pe - 1) \\ \lambda_4^* &= (\mu + \gamma_2) (R_0^2(1 - pe) + (1 - \xi_2)R_0^2pe - 1) \end{aligned}$$

Therefore, the DFE is stable if and only if $\lambda_3^* < 0$ and $\lambda_4^* < 0$.

Rearranging the terms, the necessary and sufficient conditions for the DFE to be stable can be written as

$$\begin{cases} R_p^1 = R_0^1(x^* + (1 - \xi_1)v^*) < 1 \\ R_p^2 = R_0^2(x^* + (1 - \xi_2)v^*) < 1. \end{cases} \quad (2.16)$$

or

$$\begin{cases} R_p^1 = R_0^1(1 - \xi_1 pe) < 1 \\ R_p^2 = R_0^2(1 - \xi_2 pe) < 1 \end{cases} \iff \begin{cases} pe > pe_1^{DF} \\ pe > pe_2^{DF} \end{cases}$$

where

$$\begin{cases} pe_1^{DF} = \frac{R_0^1 - 1}{R_0^1 \xi_1} \\ pe_2^{DF} = \frac{R_0^2 - 1}{R_0^2 \xi_2}. \end{cases} \quad (2.17)$$

Observe that if $R_0^1(1 - \xi_1) > 1$ or $R_0^2(1 - \xi_2) > 1$, then the DFE is never stable.

Notice moreover that if

$$\begin{cases} R_0^1 < 1 \\ R_0^2 < 1 \end{cases}$$

then the DFE is stable independently the choice of pe , ξ_1 and ξ_2 .

For this reason we assume from now on that

$$\begin{cases} R_0^1 > 1 \\ R_0^2 > 1 \end{cases}$$

2.3.2 Subtype-i-Only Equilibrium

Existence

We analyse here the Subtype-1-Only Equilibrium.

By definition, at the Subtype-1-Only Equilibrium there are no individuals infected by subtype 2 (i.e. $\bar{y}_2 = 0$) and there are positive fractions of individuals infected by subtype 1 ($\bar{y}_1 > 0$), susceptible ($\bar{x}_1 > 0$) and vaccinated individuals ($\bar{v}_1 > 0$).

Setting equal to 0 equation (2.10), together with (2.15) and $\bar{y}_1 > 0 = \bar{y}_2$, one

2.3. Existence and Stability of Equilibria

obtains

$$R_0^1(\bar{x}_1 + (1 - \xi_1)\bar{v}_1) = 1. \quad (2.18)$$

The equilibrium fractions of susceptible \bar{x}_1 can be computed by setting equal to 0 the right hand side of (2.8) so that it can be expressed as function of \bar{y}_1 as

$$\bar{x}_1 = \frac{(\mu + \gamma_1\bar{y}_1)(1 - pe)}{\mu + R_0^1(\mu + \gamma_1)\bar{y}_1} \quad (2.19)$$

Substituting $\bar{v}_1 = 1 - \bar{x}_1 - \bar{y}_1$ and (2.19) into (2.18) we obtain that \bar{y}_1 must solve $G(\bar{y}_1) = 1$, where

$$G(y) = R_0^1 \left[\xi_1 \frac{(\mu + \gamma_1 y)(1 - pe)}{\mu + R_0^1(\mu + \gamma_1)y} + (1 - \xi_1)(1 - y) \right]$$

Since we assumed $R_0^1 > 1$, we obtain

$$G'(y) = -\frac{R_0^1 \xi_1 \mu (1 - pe) [R_0^1 \mu + (R_0^1 - 1) \gamma_1]}{[\mu + R_0^1(\mu + \gamma_1)y]^2} - R_0^1(1 - \xi_1) < 0 \quad (2.20)$$

and

$$G(1) = \frac{R_0^1(\mu + \gamma_1)(1 - pe)\xi_1}{R_0^1(\mu + \gamma_1) + \mu} < 1$$

Hence, $G(\bar{y}_1) = 1$ has a unique solution in $(0, 1)$ if and only if

$$G(0) > 1 \iff R_0^1((1 - pe)\xi_1 + (1 - \xi_1)) = R_0^1(1 - \xi_1 pe) = R_p^1 > 1$$

We have then proved

Proposition 1. *A sufficient and necessary condition for a Subtype-1-Only Equilibrium to exist is $R_p^1 > 1$, i.e, $pe < pe_1^{DF}$ defined in (2.17). Moreover, under the assumption $R_p^1 > 1$, the Subtype-1-Only Equilibrium is unique.*

The equilibrium fraction of vaccinated individuals \bar{v}_1 can be computed by setting equal to 0 the right hand side of (2.9) and it can be expressed as function of \bar{y}_1 as

$$\bar{v}_1 = \frac{(\mu + \gamma_1\bar{y}_1)pe}{\mu + (1 - \xi_1)R_0^1(\mu + \gamma_1)\bar{y}_1} \quad (2.21)$$

For future use, we prove the following

Proposition 2. *\bar{y}_1 at the Subtype-1-Only Equilibrium is a decreasing func-*

tion of pe on $[0, pe_1^{DF})$.

Proof. We write explicitly the dependence of G on pe as $G(pe, \bar{y}_1(pe)) = 1$. Since by (2.20) we know that

$$\frac{\partial G(pe, \bar{y}_1(pe))}{\partial \bar{y}_1} < 0$$

and also

$$\frac{\partial G(pe, \bar{y}_1(pe))}{\partial pe} = -\frac{R_0^1 \xi_1 (\mu + \gamma_1 \bar{y}_1)}{\mu + R_0^1 (\mu + \gamma_1) \bar{y}_1} < 0$$

by the Implicit Function Theorem we obtain

$$\bar{y}_1'(pe) = -\frac{\frac{\partial G(pe, \bar{y}_1)}{\partial pe}}{\frac{\partial G(pe, \bar{y}_1)}{\partial \bar{y}_1}} < 0 \quad (2.22)$$

thus proving that \bar{y}_1 is a decreasing function of pe . \square

Completely similar arguments lead us to state that a Subtype-2-Only Equilibrium $(\bar{x}_2, \bar{v}_2, 0, \bar{y}_2)$ exists and is unique under the necessary and sufficient condition $R_p^2 > 1$.

The equilibrium fractions of susceptible and vaccinated individuals at the equilibrium are given by

$$\bar{x}_2 = \frac{(\mu + \gamma_2 \bar{y}_2)(1 - pe)}{\mu + R_0^2 (\mu + \gamma_2) \bar{y}_2} \quad \bar{v}_2 = \frac{(\mu + \gamma_2 \bar{y}_2) pe}{\mu + (1 - \xi_2) R_0^2 (\mu + \gamma_2) \bar{y}_2}$$

where \bar{y}_2 is the unique solution of equation $H(\bar{y}_2) = 1$ where

$$H(\bar{y}_2) = R_0^2 \left[\frac{(\mu + \gamma_2 \bar{y}_2)(1 - pe)}{\mu + R_0^2 (\mu + \gamma_2) \bar{y}_2} + (1 - \xi_2)(1 - \bar{y}_2) \right]$$

provided that $R_p^2 > 1$.

In terms of pe , we get that the Subtype-2-Only Equilibrium exists for

$$pe < pe_2^{DF} \quad \text{where} \quad pe_2^{DF} = \frac{R_0^2 - 1}{R_0^2 \xi_2}.$$

Finally, with the same argument used above, it can be proved that \bar{y}_2 is a

2.3. Existence and Stability of Equilibria

decreasing function of pe .

Stability

We examine now the stability of the Subtype-1-Only Equilibrium. In order to do that, we consider the Jacobian matrix of (2.14) at the Subtype-1-Only Equilibrium $E_1 = (\bar{N}_1, \bar{x}_1, \bar{y}_1, 0)$ and obtain a matrix of the form:

$$\mathbf{J}(E_1) = \begin{pmatrix} E & F \\ 0 & R_0^2(\mu + \gamma_2)[\bar{x}_1 + (1 - \xi_2)\bar{v}_1] - (\mu + \gamma_2) \end{pmatrix}$$

where E is the 3×3 matrix

$$\mathbf{E} = \begin{pmatrix} -(\mu + \gamma_1\bar{y}_1) & 0 & -\bar{N}\gamma_1 \\ (\gamma_1 - R_0^1(\gamma_1 + \mu))\frac{\bar{x}_1\bar{y}_1}{\bar{N}} & -[\mu + R_0^1(\mu + \gamma_1)\bar{y}_1] & -[R_0^1(\mu + \gamma_1) - \gamma_1]\bar{x}_1 \\ (\mu + \gamma_1\bar{y}_1)\frac{\bar{y}_1}{\bar{N}} & R_0^1(\mu + \gamma_1)\xi_1\bar{y}_1 & [\gamma_1 - R_0^1(\mu + \gamma_1)(1 - \xi_1)]\bar{y}_1 \end{pmatrix}$$

We first show the following

Lemma 1. *All the eigenvalues of E have negative real part.*

The lemma implies that the Subtype-1-Only Equilibrium is always asymptotically stable when it exists ($R_p^1 > 1$) in absence of individuals infected with subtype 2, as has been obtained in similar models with one strain and vaccination (Pugliese, 1990).

Proof. The characteristic polynomial of E (after a change of sign) can be written as

$$\lambda^3 + a_1\lambda^2 + a_2\lambda + a_3 = 0 \quad (2.23)$$

Routh-Hurwitz criterion states that all solutions of (2.23) have negative real part if and only if $a_1, a_2, a_3 > 0$ and $a_1a_2 - a_3 > 0$ (Murray, 2002).

If we set

$$\begin{aligned} K &= R_0^1(\mu + \gamma_1)(1 - \xi_1)\bar{y}_1 \\ L &= R_0^1(\mu + \gamma_1)\bar{y}_1 \end{aligned}$$

after some computations (see the Appendix) we obtain

$$\begin{aligned}
 a_1 &= 2\mu + K + L \\
 a_2 &= (\mu + L)(\mu + K) + [R_0^1(\mu + \gamma_1) - \gamma_1]L\xi_1\bar{x}_1 + (\mu + \gamma_1\bar{y}_1)K \\
 a_3 &= [R_0^1(\mu + \gamma_1) - \gamma_1]L\mu\xi_1\bar{x}_1 + (\mu + \gamma_1\bar{y}_1)(\mu + L)K \\
 a_1a_2 - a_3 &= (\mu + L)(\mu + K)(2\mu + L + K) + [R_0^1(\mu + \gamma_1) - \gamma_1]L\xi_1\bar{x}_1(\mu + L + K) + \\
 &\quad + (\mu + \gamma_1\bar{y}_1)(\mu + K)K.
 \end{aligned}$$

Since $R_0^1 > 1$ and hence $R_0^1(\mu + \gamma_1) > \gamma_1$ all these quantities are positive, thus proving that Routh-Hurwitz conditions are satisfied. \square

Since $J(E_1)$ is block-triangular, the set of eigenvalues of $J(E_1)$ is given by the union of the set of eigenvalues of E and

$$\bar{\lambda}_4 = R_0^2(\mu + \gamma_2)[\bar{x}_1 + (1 - \xi_2)\bar{v}_1] - (\mu + \gamma_2)$$

Hence, the Subtype-1-Only Equilibrium is stable for (2.14) if and only if

$$\bar{\lambda}_4 = R_0^2(\mu + \gamma_2)[\bar{x}_1 + (1 - \xi_2)\bar{v}_1] - (\mu + \gamma_2) < 0.$$

Rearranging the terms, the Subtype-1-Only Equilibrium is stable if and only if

$$R_p^{2:1} = R_0^2(\bar{x}_1 + (1 - \xi_2)\bar{v}_1) < 1. \quad (2.24)$$

We wish now to express (2.24) in terms of \bar{y}_1 only. To this aim, one can immediately insert (2.19) into (2.24). Instead, to obtain a simpler expression that does not contain pe , we start by rewriting (2.18) as

$$\frac{R_0^1(\mu + \gamma_1\bar{y}_1)}{\mu + R_0^1(\mu + \gamma_1)\bar{y}_1} \left(1 - \frac{pe\mu\xi_1}{\mu + R_0^1(\mu + \gamma_1)(1 - \xi_1)\bar{y}_1} \right) = 1$$

and by algebraic manipulation of the expression we may write pe as function of \bar{y}_1

$$pe = \frac{[R_0^1(1 - \bar{y}_1) - 1][\mu + R_0^1(\mu + \gamma_1)(1 - \xi_1)\bar{y}_1]}{R_0^1(\mu + \gamma_1\bar{y}_1)\xi_1} \quad (2.25)$$

Substituting (2.25) into (2.19) we obtain the following expression for equi-

2.3. Existence and Stability of Equilibria

librium fraction of effectively vaccinated individuals

$$\bar{v}_1 = \frac{R_0^1(1 - \bar{y}_1) - 1}{R_0^1\xi_1} \quad (2.26)$$

Finally we obtain from (2.24)

$$\begin{aligned} R_p^{2:1} &= R_0^2[\bar{x}_1 + (1 - \xi_2)\bar{v}_1] \\ &= R_0^2[\bar{x}_1 + (1 - \xi_1)\bar{v}_1 + (\xi_1 - \xi_2)\bar{v}_1] \\ &= (\text{using (2.18)}) R_0^2 \left[\frac{1}{R_0^1} + (\xi_1 - \xi_2)\bar{v}_1 \right] \\ &= \frac{R_0^2}{R_0^1} \left[1 + \frac{\xi_1 - \xi_2}{\xi_1} [R_0^1(1 - \bar{y}_1) - 1] \right]. \end{aligned} \quad (2.27)$$

Summarizing, we have obtained:

Proposition 3. *The Subtype-1-Only Equilibrium $E_1 = (\bar{N}_1, \bar{x}_1, \bar{y}_1, 0)$ is asymptotically stable [unstable] if $R_p^{2:1} < [>]1$, where $R_p^{2:1}$ is given by expression (2.24) or (2.27).*

We now wish to express condition $R_p^{2:1} < 1$ in terms of pe .

In all the rest of the paper let us assume, without loss of generality, that $R_0^1 > R_0^2$.

If $\xi_1 \leq \xi_2$, (2.27) implies that $R_p^{2:1} < 1$. In other words, if $\xi_1 \leq \xi_2$, the Subtype-1-Only Equilibrium is asymptotically stable, when it exists.

Therefore, we study the condition $R_p^{2:1} < 1$ under the additional assumption $\xi_1 > \xi_2$.

Since $\frac{(\xi_1 - \xi_2)}{\xi_1} R_0^2 \bar{y}_1 > 0$ and $\frac{R_0^2}{R_0^1} < 1$, expression (2.27) implies that

$$\begin{aligned} R_p^{2:1} &= \frac{R_0^2}{R_0^1} \left[1 + \frac{(\xi_1 - \xi_2)}{\xi_1} R_0^1 - \frac{(\xi_1 - \xi_2)}{\xi_1} \right] - \frac{(\xi_1 - \xi_2)}{\xi_1} R_0^2 \bar{y}_1 \\ &< \frac{R_0^2}{R_0^1} \left[R_0^1 - \frac{\xi_2}{\xi_1} R_0^1 + \frac{\xi_2}{\xi_1} \right]. \end{aligned} \quad (2.28)$$

Since

$$\frac{R_0^2}{R_0^1} \left[R_0^1 - \frac{\xi_2}{\xi_1} R_0^1 + \frac{\xi_2}{\xi_1} \right] \leq 1 \iff C := R_0^1 R_0^2 (\xi_1 - \xi_2) + R_0^2 \xi_2 - R_0^1 \xi_1 \leq 0. \quad (2.29)$$

inequality (2.28) shows that, if $R_0^1 > R_0^2$, $\xi_1 > \xi_2$ and $C \leq 0$, then $R_p^{2:1} < 1$ for every value of $0 \leq pe \leq 1$.

In order to study when $R_p^{2:1} > 1$, we then add the assumption $C > 0$, using then the assumptions

$$R_0^1 > R_0^2, \quad \xi_1 > \xi_2 \quad \text{and} \quad C > 0 \quad (2.30)$$

Let us now set $R_p^{2:1} = 1$ and find the corresponding fraction of infected individuals

$$\bar{y}_1^{BP} = \frac{R_0^1 R_0^2 (\xi_1 - \xi_2) + R_0^2 \xi_2 - R_0^1 \xi_1}{R_0^1 R_0^2 (\xi_1 - \xi_2)} = \frac{C}{R_0^1 R_0^2 (\xi_1 - \xi_2)}. \quad (2.31)$$

The superscript BP is related to the fact that this value corresponds to a branching point of equilibrium curves, as will be seen later.

Because of the monotonic dependence of $R_p^{2:1}$ on \bar{y}_1 (2.27), we have

$$R_p^{2:1} < [>]1 \quad \iff \quad \bar{y}_1 > [<]\bar{y}_1^{BP}$$

By (2.25) and (2.31) we can compute the pe values at which branching occurs.

We see that $R_p^{2:1} = 1$ for

$$pe_1^{BP} = \hat{v} \frac{R_0^1 R_0^2 (\xi_1 - \xi_2) [R_0^1 (\mu + \gamma_1) (1 - \xi_1) + \mu] + R_0^1 (\mu + \gamma_1) (1 - \xi_1) (R_0^2 \xi_2 - R_0^1 \xi_1)}{[R_0^1 R_0^2 (\xi_1 - \xi_2) (\mu + \gamma_1) + \gamma_1 (R_0^2 \xi_2 - R_0^1 \xi_1)]} \quad (2.32)$$

where

$$\hat{v} = \frac{(R_0^1 - R_0^2)}{R_0^1 R_0^2 (\xi_1 - \xi_2)}. \quad (2.33)$$

Since \bar{y}_1 is a decreasing function of pe (see (2.22)), we conclude that

$$R_p^{2:1} < [>]1 \quad \iff \quad pe < [>]pe_1^{BP}$$

By algebraic manipulation of (2.32), pe_1^{BP} may be written as

$$pe_1^{BP} = \frac{(R_0^1 - R_0^2)}{R_0^2 (\xi_1 - \xi_2)} \left[1 - \xi_1 + \frac{\mu \xi_1 (R_0^2 (1 - \xi_2) - R_0^1 (1 - \xi_1))}{(C \gamma_1 + \mu R_0^1 R_0^2 (\xi_1 - \xi_2))} \right] \quad (2.34)$$

where C is defined in (2.29).

2.3. Existence and Stability of Equilibria

We have the following

Lemma 2. *Assume (2.30).*

a) *If $R_0^2(1 - \xi_2) > R_0^1(1 - \xi_1)$, then $0 < pe_1^{BP} < 1$ at least for $\mu/\gamma_1 > 0$ small enough.*

b) *If $R_0^2(1 - \xi_2) \leq R_0^1(1 - \xi_1)$, then $pe_1^{BP} \geq 1$ for all $\mu > 0$.*

The proof is in the Appendix.

The assumption $R_0^2(1 - \xi_2) > R_0^1(1 - \xi_1)$ is then necessary for strain 2 to be able to invade the Subtype-1-Only Equilibrium.

Summing up, necessary assumptions for having $R_p^{2:1} > 1$ with $pe \leq 1$ are

$$\begin{cases} \xi_1 > \xi_2 \\ C > 0 \\ R_0^2(1 - \xi_2) > R_0^1(1 - \xi_1) \end{cases} \quad (2.35)$$

We conclude the following

Proposition 4. *Under the assumption $R_0^1 > R_0^2$, if any of the conditions (2.35) is violated, then the Subtype-1-Only Equilibrium is asymptotically stable for all $0 \leq pe \leq 1$ in which this equilibrium is defined. If all of (2.35) are satisfied, then the Subtype-1-Only Equilibrium (when it is defined) is asymptotically stable for $0 \leq pe \leq pe_1^{BP}$ and unstable for $pe > pe_1^{BP}$, where pe_1^{BP} is defined by (2.32) or (2.34). Under (2.35) the quantity $pe_1^{BP} < 1$ at least for μ small enough.*

Completely similar (but reversed) arguments apply to the Subtype-2-Only Equilibrium. It is asymptotically stable [unstable] if

$$R_p^{1:2} = R_0^1(\bar{x}_2 + (1 - \xi_1)\bar{v}_2) < [>]1.$$

As before, we may write $R_p^{1:2}$ as

$$R_p^{1:2} = \frac{R_0^1}{R_0^2} \left[1 + \frac{\xi_2 - \xi_1}{\xi_2} [R_0^2(1 - \bar{y}_2) - 1] \right] \quad (2.36)$$

Again, if we assume that $R_0^1 > R_0^2$, then (2.36) together with

$$\bar{v}_2 = \frac{R_0^2(1 - \bar{y}_2) - 1}{R_0^2 \xi_1} > 0 \quad (2.37)$$

implies that

$$\xi_2 \geq \xi_1 \implies R_p^{1:2} > 1.$$

That is, if $R_0^1 > R_0^2$ and $\xi_2 \geq \xi_1$, then Subtype-1 invades the Subtype-2-Only Equilibrium, wherever it exists.

Assume now $\xi_2 < \xi_1$ together with $R_0^1 > R_0^2$. By the same reasoning made before, expression (2.36) implies

$$R_p^{1:2} = \frac{R_0^1}{R_0^2} \left[1 + \frac{\xi_2 - \xi_1}{\xi_2} (R_0^2 - 1) \right] + \frac{\xi_1 - \xi_2}{\xi_2} \bar{y}_2 > \frac{R_0^1}{R_0^2} \left[R_0^2 - \frac{\xi_1}{\xi_2} R_0^2 + \frac{\xi_1}{\xi_2} \right]. \quad (2.38)$$

The right hand side of (2.38) is greater or equal than 1, if and only if $C \leq 0$ with C defined in (2.35). Hence

$$C \leq 0 \implies R_p^{1:2} > 1,$$

i.e. strain 1 invades the Subtype-2-Only Equilibrium whenever this exists. To proceed, we also assume $C > 0$.

As before, we find the fraction of individuals infected with strain 2 at the equilibrium corresponding to $R_p^{1:2} = 1$:

$$\bar{y}_2^{BP} = \frac{R_0^1 R_0^2 (\xi_1 - \xi_2) + R_0^2 \xi_2 - R_0^1 \xi_1}{R_0^1 R_0^2 (\xi_1 - \xi_2)} = \frac{C}{R_0^1 R_0^2 (\xi_1 - \xi_2)}$$

Notice that, since (2.36) is an increasing function of \bar{y}_2 (remember $\xi_1 > \xi_2$), we have

$$R_p^{1:2} < 1 \iff \bar{y}_2 < \bar{y}_2^{BP}.$$

Writing, analogously to (2.25), pe as function of \bar{y}_2 as

$$pe = \frac{[R_0^2(1 - \bar{y}_2) - 1][\mu + R_0^2(\mu + \gamma_2)(1 - \xi_2)\bar{y}_2]}{R_0^2 \xi_2 (\mu + \gamma_2 \bar{y}_2)} \quad (2.39)$$

2.3. Existence and Stability of Equilibria

we see that $R_p^{1:2} = 1$ for

$$pe_2^{BP} = \hat{v} \frac{R_0^1 R_0^2 (\xi_1 - \xi_2) [R_0^2 (\mu + \gamma_2) (1 - \xi_2) + \mu] + R_0^2 (R_0^2 \xi_2 - R_0^1 \xi_1) (\mu + \gamma_2) (1 - \xi_2)}{[R_0^1 R_0^2 (\xi_1 - \xi_2) (\mu + \gamma_2) + \gamma_2 (R_0^2 \xi_2 - R_0^1 \xi_1)]} \quad (2.40)$$

using the definition (2.33) for \hat{v} .

Since \bar{y}_2 is a decreasing function of pe (by (2.22)), we conclude that

$$R_p^{1:2} < 1 \iff pe > pe_2^{BP}$$

By manipulation of (2.40) we find that

$$pe_2^{BP} = \frac{(R_0^1 - R_0^2)}{R_0^1 (\xi_1 - \xi_2)} \left[1 - \xi_2 + \frac{\mu \xi_2 (R_0^2 (1 - \xi_2) - R_0^1 (1 - \xi_1))}{\gamma_2 C + \mu R_0^1 R_0^2 (\xi_1 - \xi_2)} \right]. \quad (2.41)$$

Analogously to Lemma 2, we have

Lemma 3. *Assume (2.30).*

- a) *If $R_0^2 (1 - \xi_2) > R_0^1 (1 - \xi_1)$, then $0 < pe_2^{BP} < 1$ at least for $\mu/\gamma_2 > 0$ small enough.*
- b) *If $R_0^2 (1 - \xi_2) \leq R_0^1 (1 - \xi_1)$, then $pe_2^{BP} \geq 1$ for all $\mu > 0$.*

The proof is identical to that of Lemma 2 and is skipped.

Symmetrically to Proposition 4, we obtain

Proposition 5. *Under the assumption $R_0^1 > R_0^2$, if any of the conditions (2.35) is violated, then the Subtype-2-Only Equilibrium is unstable for all $0 \leq pe \leq 1$ in which this equilibrium is defined. If all of (2.35) are satisfied, then the Subtype-2-Only Equilibrium is unstable for $0 \leq pe \leq pe_2^{BP}$ and asymptotically stable for $pe > pe_2^{BP}$ (when the equilibrium itself is defined), where pe_2^{BP} is defined by (2.40) or (2.41). Under (2.35) the quantity $pe_2^{BP} < 1$ at least for μ small enough.*

2.3.3 Coexistence Equilibrium

Existence

At the Coexistence Equilibrium completely susceptible ($\hat{x} > 0$), effectively vaccinated ($\hat{v} > 0$), individuals infected by subtype 1 ($\hat{y}_1 > 0$) and individu-

als infected by subtype 2 ($\hat{y}_2 > 0$) are all present in the community.

Setting equal to 0 equations (2.10)–(2.11), together with (2.15) one obtains

$$\begin{cases} R_0^1[\hat{x} + (1 - \xi_1)\hat{v}] = 1 \\ R_0^2[\hat{x} + (1 - \xi_2)\hat{v}] = 1 \end{cases} \quad (2.42)$$

The equilibrium fractions \hat{x} , \hat{v} can be computed solving (2.42):

$$\hat{x} = \frac{R_0^2(1 - \xi_2) - R_0^1(1 - \xi_1)}{R_0^1 R_0^2 (\xi_1 - \xi_2)} \quad \hat{v} = \frac{R_0^1 - R_0^2}{R_0^1 R_0^2 (\xi_1 - \xi_2)} \quad (2.43)$$

Fraction \hat{v} is positive under the condition that

$$\text{if } R_0^i > R_0^j, \quad \text{then } \xi_i > \xi_j$$

which means that the Coexistence Equilibrium exists only if the vaccine induces a higher degree of protection against the subtype with the higher fitness in a completely susceptible population.

Without loss of generality, let's assume $R_0^1 > R_0^2$ and require $\xi_1 > \xi_2$.

The susceptible fraction \hat{x} is positive if and only if

$$R_0^2(1 - \xi_2) > R_0^1(1 - \xi_1)$$

By substitution of (2.43) into $\hat{x} + \hat{v} < 1$ one obtains

$$\frac{R_0^1 \xi_1 - R_0^2 \xi_2}{R_0^1 R_0^2 (\xi_1 - \xi_2)} < 1 \quad \iff \quad C > 0$$

where C is given by (2.29).

Conditions (2.35) are then necessary for the existence of a Coexistence Equilibrium.

Setting equal to 0 equations (2.8)–(2.9) together with (2.15) and using matrix notation, one obtains

$$A \begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \end{pmatrix} = \mu \begin{pmatrix} 1 - pe - \hat{x} \\ pe - \hat{v} \end{pmatrix} \quad (2.44)$$

2.3. Existence and Stability of Equilibria

where

$$A = \begin{pmatrix} R_0^1(\mu + \gamma_1)\hat{x} - \gamma_1(1 - pe) & R_0^2(\mu + \gamma_2)\hat{x} - \gamma_2(1 - pe) \\ R_0^1(\mu + \gamma_1)(1 - \xi_1)\hat{v} - \gamma_1pe & R_0^2(\mu + \gamma_2)(1 - \xi_2)\hat{v} - \gamma_2pe \end{pmatrix}$$

System (2.44) admits a unique solution if and only if

$$|A| = \mu[R_0^2(\mu + \gamma_2)(1 - \xi_2)\hat{v} - R_0^1(\mu + \gamma_1)(1 - \xi_1)\hat{v} + pe(\gamma_1 - \gamma_2)] \neq 0 \quad (2.45)$$

Observation 1. *If $\gamma_1 = \gamma_2$, then by (2.35) $|A| > 0$ for $0 \leq pe \leq 1$.*

Under the further assumption that $|A| \neq 0$, we can explicitly solve (2.44) by Cramer's rule

$$\begin{aligned} \hat{y}_1 &= \frac{\begin{vmatrix} \mu(1 - pe - \hat{x}) & R_0^2(\mu + \gamma_2)\hat{x} - \gamma_2(1 - pe) \\ \mu(pe - \hat{v}) & R_0^2(\mu + \gamma_2)(1 - \xi_2)\hat{v} - \gamma_2pe \end{vmatrix}}{|A|} \\ &= \frac{pe\mu[\gamma_2(\hat{x} + \hat{v}) - (\mu + \gamma_2)] + \mu\hat{v}[R_0^2(\mu + \gamma_2)(1 - \xi_2 + \xi_2\hat{x}) - \gamma_2]}{|A|} \end{aligned} \quad (2.46)$$

$$\begin{aligned} \hat{y}_2 &= \frac{\begin{vmatrix} R_0^1(\mu + \gamma_1)\hat{x} - \gamma_1(1 - pe) & \mu(1 - pe - \hat{x}) \\ R_0^1(\mu + \gamma_1)(1 - \xi_1)\hat{v} - \gamma_1pe & \mu(pe - \hat{v}) \end{vmatrix}}{|A|} \\ &= \frac{pe\mu[\gamma_1(1 - (\hat{x} + \hat{v})) + \mu] + \mu\hat{v}[\gamma_1 - R_0^1(\mu + \gamma_1)(1 - \xi_1 + \xi_1\hat{x})]}{|A|} \end{aligned} \quad (2.47)$$

We conclude the following

Proposition 6. *Under the assumption $R_0^1 > R_0^2$, necessary conditions for a Coexistence Equilibrium to exist are given by (2.35). Moreover, if $\gamma_1 = \gamma_2$ the Coexistence Equilibrium is unique. If $\gamma_1 \neq \gamma_2$ the Coexistence Equilibrium is unique under the assumption that $|A| \neq 0$, where $|A|$ is given by (2.45).*

Let's now prove the following

Lemma 4. *The equilibrium fractions \hat{x} , \hat{v} , \hat{y}_1 , \hat{y}_2 satisfy condition*

$$\hat{x} + \hat{v} + \hat{y}_1 + \hat{y}_2 = 1 \quad (2.48)$$

Proof. Equations (2.8)–(2.9) together with (2.48) can be written as

$$B \begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \\ 1 \end{pmatrix} = 0 \quad (2.49)$$

where

$$B = \begin{pmatrix} R_0^1(\mu + \gamma_1)\hat{x} - \gamma_1(1 - pe) & R_0^2(\mu + \gamma_2)\hat{x} - \gamma_2(1 - pe) & \mu(1 - pe - \hat{x}) \\ R_0^1(\mu + \gamma_1)(1 - \xi_1)\hat{v} - \gamma_1pe & R_0^2(\mu + \gamma_2)(1 - \xi_2)\hat{v} - \gamma_2pe & \mu(pe - \hat{x}) \\ 1 & 1 & 1 - \hat{x} - \hat{v} \end{pmatrix}$$

By (2.35) matrix B can be reduced to the form

$$\begin{pmatrix} R_0^1(\mu + \gamma_1)\hat{x} - \gamma_1(1 - pe) & R_0^2(\mu + \gamma_2)\hat{x} - \gamma_2(1 - pe) & \mu(1 - pe - \hat{x}) \\ \mu & \mu & \mu(1 - \hat{x} - \hat{v}) \\ 1 & 1 & 1 - \hat{x} - \hat{v} \end{pmatrix}$$

thus proving our claim. □

In order to find sufficient conditions for the existence of a positive equilibrium, we start with the assumption $|A| > 0$.

By (2.47) one obtains that $\hat{y}_2 > 0$ for

$$pe > \frac{\hat{v}[R_0^1(\mu + \gamma_1)(1 - \xi_1 + \xi_1\hat{x}) - \gamma_1]}{\mu + \gamma_1(1 - (\hat{x} + \hat{v}))}$$

2.3. Existence and Stability of Equilibria

Substituting (2.43) and rearranging the terms one gets

$$\begin{aligned}
pe &> \hat{v} \frac{\left[R_0^1(\mu + \gamma_1)(1 - \xi_1) + R_0^1 \xi_1(\mu + \gamma_1) \frac{R_0^2(1 - \xi_2) - R_0^1(1 - \xi_1)}{R_0^1 R_0^2(\xi_1 - \xi_2)} - \gamma_1 \right]}{R_0^1 R_0^2(\xi_1 - \xi_2)\mu + \gamma_1 [R_0^1 R_0^2(\xi_1 - \xi_2) - R_0^1 \xi_1 + R_0^2 \xi_2]} \\
pe &> \hat{v} \frac{R_0^1 R_0^2(\xi_1 - \xi_2) [R_0^1(\mu + \gamma_1)(1 - \xi_1) - \gamma_1] + R_0^1 \xi_1(\mu + \gamma_1) [R_0^2(1 - \xi_2) - R_0^1(1 - \xi_1)]}{R_0^1 R_0^2(\xi_1 - \xi_2)(\mu + \gamma_1) + \gamma_1 (R_0^2 \xi_2 - R_0^1 \xi_1)} \\
pe &> \hat{v} \frac{R_0^1 R_0^2(\xi_1 - \xi_2) [R_0^1(\mu + \gamma_1)(1 - \xi_1) + \mu] + R_0^1(\mu + \gamma_1)(1 - \xi_1) (R_0^2 \xi_2 - R_0^1 \xi_1)}{R_0^1 R_0^2(\xi_1 - \xi_2)(\mu + \gamma_1) + \gamma_1 (R_0^2 \xi_2 - R_0^1 \xi_1)} \quad (2.50)
\end{aligned}$$

By (2.32) inequality (2.50) can be written as

$$pe > pe_1^{BP}$$

Similarly, by (2.46), condition $\hat{y}_1 > 0$ can be expressed in terms of pe

$$pe < \frac{\hat{x}\hat{v}[R_0^2 \xi_2(\mu + \gamma_2)] + \hat{v}[R_0^2(\mu + \gamma_2)(1 - \xi_2) - \gamma_2]}{\mu + \gamma_2(1 - (\hat{x} + \hat{v}))}$$

Substituting (2.43) and rearranging the terms one gets

$$\begin{aligned}
pe &< \hat{v} \frac{[R_0^2(1 - \xi_2) - R_0^1(1 - \xi_1)][R_0^2 \xi_2(\mu + \gamma_2)] + [R_0^2(\mu + \gamma_2)(1 - \xi_2) - \gamma_2]}{[R_0^1 R_0^2(\xi_1 - \xi_2)(\mu + \gamma_2) + \gamma_2(R_0^1 \xi_1 - R_0^2 \xi_2)]} \\
pe &< \hat{v} \frac{R_0^1 R_0^2(\xi_1 - \xi_2) [R_0^2(\mu + \gamma_2)(1 - \xi_2) - \gamma_2] + R_0^2 \xi_2(\mu + \gamma_2) [R_0^2(1 - \xi_2) - R_0^1(1 - \xi_1)]}{R_0^1 R_0^2(\xi_1 - \xi_2)(\mu + \gamma_2) + \gamma_2 (R_0^2 \xi_2 - R_0^1 \xi_1)} \\
pe &< \hat{v} \frac{R_0^1 R_0^2(\xi_1 - \xi_2) [R_0^2(\mu + \gamma_2)(1 - \xi_2) + \mu] + R_0^2 (R_0^2 \xi_2 - R_0^1 \xi_1)(\mu + \gamma_2)(1 - \xi_2)}{[R_0^1 R_0^2(\xi_1 - \xi_2)(\mu + \gamma_2) + \gamma_2 (R_0^2 \xi_2 - R_0^1 \xi_1)]} \quad (2.51)
\end{aligned}$$

By (2.40) inequality (2.51) can be written as

$$pe < pe_2^{BP}$$

With similar (but reversed) arguments, one finds that under the assumption $|A| < 0$, $\hat{y}_2 > 0$ and $\hat{y}_1 > 0$ for

$$pe_2^{BP} < pe < pe_1^{BP}$$

We have then proved the following

Proposition 7. *Under the assumptions $R_0^1 > R_0^2$ and (2.35), sufficient and*

necessary conditions for the Coexistence Equilibrium to exist are

(a) if $|A| > 0$, $R_p^{1:2} > 1$ and $R_p^{2:1} > 1$ (i.e. $pe_1^{BP} < pe < pe_2^{BP}$);

(b) if $|A| < 0$, $R_p^{1:2} < 1$ and $R_p^{2:1} < 1$ (i.e. $pe_2^{BP} < pe < pe_1^{BP}$).

where pe_1^{BP} and pe_2^{BP} are given by (2.34) and (2.41) respectively.

Conditions for sub- or super-critical bifurcations

It is therefore relevant finding whether $pe_1^{BP} < pe_2^{BP}$ or vice versa.

Lemma 5. Under the assumption $R_0^1 > R_0^2$ and (2.35), pe_1^{BP} and pe_2^{BP} , given by (2.34) and (2.41), are decreasing functions of γ_1 and γ_2 respectively.

If $\gamma_1 \geq \gamma_2$ then $pe_1^{BP} < pe_2^{BP}$.

If

$$R_0^2(\xi_1 - \xi_2) \geq \xi_1 \quad (2.52)$$

then $pe_1^{BP} < pe_2^{BP}$ for all values of γ_1 and γ_2 .

If $R_0^2(\xi_1 - \xi_2) < \xi_1$ then $pe_1^{BP} > pe_2^{BP}$ for γ_1 small enough, and γ_2 large enough.

Proof. The fact that pe_1^{BP} and pe_2^{BP} are decreasing functions of γ_1 and γ_2 is an immediate consequence of expressions (2.34) and (2.41) and assumptions (2.35).

Consider now $\gamma_1 = \gamma_2 = \gamma$. With some simple algebraic manipulations, one obtains

$$\begin{aligned} pe_2^{BP} - pe_1^{BP} &= \frac{(R_0^1 - R_0^2)(1 - \xi_2)}{R_0^2(\xi_1 - \xi_2)} - \frac{(R_0^1 - R_0^2)(1 - \xi_1)}{R_0^2(\xi_1 - \xi_2)} \\ &+ \frac{\mu(R_0^2(1 - \xi_2) - R_0^1(1 - \xi_1))}{(C\gamma + \mu R_0^1 R_0^2(\xi_1 - \xi_2))} \left[\frac{(R_0^1 - R_0^2)\xi_2}{R_0^2(\xi_1 - \xi_2)} - \frac{(R_0^1 - R_0^2)\xi_1}{R_0^2(\xi_1 - \xi_2)} \right] \\ &= \frac{(R_0^1 - R_0^2)(R_0^2(1 - \xi_2) - R_0^1(1 - \xi_1))(\mu + \gamma)C}{R_0^1 R_0^2(\xi_1 - \xi_2)(C\gamma + \mu R_0^1 R_0^2(\xi_1 - \xi_2))} > 0. \end{aligned}$$

From the fact that pe_1^{BP} and pe_2^{BP} are decreasing functions of γ_1 and γ_2 respectively, we may conclude that the inequality $pe_1^{BP} < pe_2^{BP}$ holds also for every $\gamma_1 > \gamma_2$.

As for the final claim, if $\lim_{\gamma_1 \rightarrow 0} pe_1^{BP} \leq \lim_{\gamma_2 \rightarrow +\infty} pe_2^{BP}$, then $pe_1^{BP} < pe_2^{BP}$ for all finite, positive γ_1 and γ_2 .

2.3. Existence and Stability of Equilibria

Vice versa, if $\lim_{\gamma_1 \rightarrow 0} pe_1^{BP} > \lim_{\gamma_2 \rightarrow +\infty} pe_2^{BP}$, by continuity $pe_1^{BP} > pe_2^{BP}$ over some range of γ_1 and γ_2 values.

$$\lim_{\gamma_1 \rightarrow 0} pe_1^{BP} = \frac{(R_0^1 - R_0^2)}{R_0^2(\xi_1 - \xi_2)} \left[1 - \xi_1 + \frac{\xi_1(R_0^2(1 - \xi_2) - R_0^1(1 - \xi_1))}{R_0^1 R_0^2(\xi_1 - \xi_2)} \right]$$

while

$$\lim_{\gamma_2 \rightarrow +\infty} pe_2^{BP} = \frac{(R_0^1 - R_0^2)(1 - \xi_2)}{R_0^2(\xi_1 - \xi_2)}.$$

Hence, with some algebra

$$\begin{aligned} \lim_{\gamma_1 \rightarrow 0} pe_1^{BP} - \lim_{\gamma_2 \rightarrow +\infty} pe_2^{BP} &= \frac{(R_0^2(1 - \xi_2) - R_0^1(1 - \xi_1))(R_0^1 - R_0^2)}{R_0^1 R_0^2(\xi_1 - \xi_2)} \left(-1 + \frac{\xi_1}{R_0^2(\xi_1 - \xi_2)} \right). \end{aligned}$$

This quantity is positive if and only if $R_0^2(\xi_1 - \xi_2) < \xi_1$, yielding the conclusion of the proof. \square

Finally, we show

Lemma 6. (a) if $pe_1^{BP} < pe_2^{BP}$, then $|A| > 0$ for all $pe \in [pe_1^{BP}, pe_2^{BP}]$;

(b) if $pe_2^{BP} < pe_1^{BP}$, then $|A| < 0$ for all $pe \in [pe_2^{BP}, pe_1^{BP}]$.

(c) if $pe_2^{BP} = pe_1^{BP}$, then $|A| = 0$ for $pe = pe_1^{BP} = pe_2^{BP}$.

Through a) and b), we will be able to draw a clear bifurcation pattern (see for example Britton (2003)) of the system, with transcritical bifurcations occurring at E_1 for $pe = pe_1^{BP}$, and at E_2 for $pe = pe_2^{BP}$.

Proof. We denote by $|A|_{BPi}$ the determinant of A computed with $pe = pe_i^{BP}$, $i = 1, 2$. Through simple, but tedious, algebraic manipulations, one arrives at

$$\begin{aligned} |A|_{BP1} &= \frac{\mu(R_0^2(1 - \xi_2) - R_0^1(1 - \xi_1))(R_0^1 - R_0^2)}{R_0^1 R_0^2(\xi_1 - \xi_2)} \\ &\quad \times \left[\mu + \gamma_2 - \frac{\gamma_2 \mu R_0^1 \xi_1}{C\gamma_1 + \mu R_0^1 R_0^2(\xi_1 - \xi_2)} + \frac{\gamma_1 \mu R_0^1 \xi_1}{C\gamma_1 + \mu R_0^1 R_0^2(\xi_1 - \xi_2)} \right]. \end{aligned} \tag{2.53}$$

It is immediate to see that $|A|_{BP1}$ is an increasing function of γ_1 . We already know (Observation 1) that $|A|_{BP1} > 0$ for $\gamma_1 \geq \gamma_2$.

Setting (2.53) equal to 0, we see that

$$|A|_{BP1} = 0 \iff \gamma_1 = \Psi_1(\gamma_2) \quad (2.54)$$

with

$$\Psi_1(\gamma_2) = \frac{\mu(\gamma_2 R_0^1(\xi_1 - R_0^2(\xi_1 - \xi_2)) - \mu R_0^1 R_0^2(\xi_1 - \xi_2))}{C(\mu + \gamma_2) + \mu R_0^1 \xi_1}. \quad (2.55)$$

If $\Psi_1(\gamma_2) < 0$, then $|A|_{BP1} > 0$ for all $\gamma_1 > 0$. In particular $\Psi_1(\gamma_2) < 0$ for all $\gamma_2 > 0$ if $R_0^2(\xi_1 - \xi_2) \geq \xi_1$, i.e. (2.52) holds.

Otherwise, $|A|_{BP1} > 0$ for $\gamma_1 > \Psi_1(\gamma_2)$ and $|A|_{BP1} < 0$ for $\gamma_1 < \Psi_1(\gamma_2)$.

Similarly, we obtain

$$\begin{aligned} |A|_{BP2} &= \frac{\mu(R_0^2(1 - \xi_2) - R_0^1(1 - \xi_1))(R_0^1 - R_0^2)}{R_0^1 R_0^2(\xi_1 - \xi_2)} \\ &\quad \times \left[\mu - \frac{\gamma_2 \mu R_0^2 \xi_2}{C\gamma_2 + \mu R_0^1 R_0^2(\xi_1 - \xi_2)} + \gamma_1 + \frac{\gamma_1 \mu R_0^2 \xi_2}{C\gamma_2 + \mu R_0^1 R_0^2(\xi_1 - \xi_2)} \right]. \end{aligned} \quad (2.56)$$

Now it is immediate to see that $|A|_{BP2}$ is a decreasing function of γ_2 and

$$\lim_{\gamma_2 \rightarrow \infty} |A|_{BP2} = \frac{\mu(R_0^2(1 - \xi_2) - R_0^1(1 - \xi_1))(R_0^1 - R_0^2)}{R_0^1 R_0^2(\xi_1 - \xi_2)} \left(\mu + \gamma_1 - \frac{\mu R_0^2 \xi_2}{C} \right).$$

Hence, if

$$C(\mu + \gamma_1) \geq \mu R_0^2 \xi_2 \quad (2.57)$$

$|A|_{BP2} > 0$ for all $\gamma_2 > 0$.

Otherwise, when (2.57) does not hold, setting (2.56) equal to 0, we see that

$$|A|_{BP2} = 0 \iff \gamma_2 = \Psi_2(\gamma_1) \quad (2.58)$$

with

$$\Psi_2(\gamma_1) = \frac{\mu(\gamma_1 R_0^2(\xi_2 + R_0^1(\xi_1 - \xi_2)) + \mu R_0^1 R_0^2(\xi_1 - \xi_2))}{\mu R_0^2 \xi_2 - C(\mu + \gamma_1)}. \quad (2.59)$$

2.3. Existence and Stability of Equilibria

We see that, if (2.57) does not hold, $\Psi_2(\gamma_1) > 0$, and $|A|_{BP2} > 0$ for $\gamma_2 < \Psi_2(\gamma_1)$ and $|A|_{BP2} < 0$ for $\gamma_2 > \Psi_2(\gamma_1)$.

Finally, we observe that the conditions for $|A|_{BPi} = 0$, $i = 1, 2$, are actually the same; more precisely $\Psi_2(\gamma_1)$ is the inverse of Ψ_1 , defined on the appropriate domain. Indeed, solving the equation $\gamma_1 = \Psi_1(\gamma_2)$ for γ_2 , we obtain

$$\gamma_2 = \frac{\mu(\gamma_1(C + R_0^1\xi_1) + \mu R_0^1 R_0^2(\xi_1 - \xi_2))}{\mu R_0^1(\xi_1 - R_0^2(\xi_1 - \xi_2)) - C\gamma_1} = \Psi_2(\gamma_1)$$

where the last identity comes from the definition of C , so that

$$C + R_0^1\xi_1 = R_0^2(\xi_2 + R_0^1(\xi_1 - \xi_2)) \quad \text{and} \quad R_0^1(\xi_1 - R_0^2(\xi_1 - \xi_2)) = -C + R_0^2\xi_2.$$

Summarizing, we have obtained that if (2.52) does not hold, the function $\gamma_2 = \Psi_2(\gamma_1)$ divides the plane into two regions (see Figure 2.2) such that below and to the right both $|A|_{BP1}$ and $|A|_{BP2}$ are positive; above and to the left both are negative.

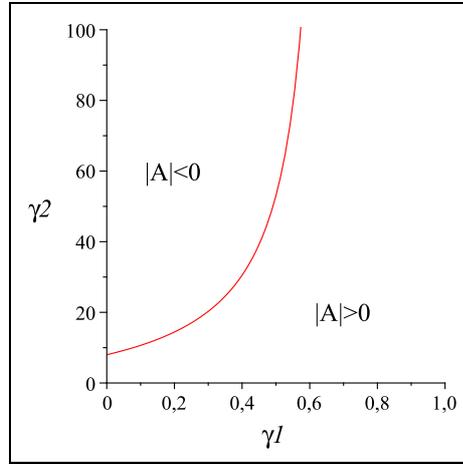


Figure 2.2: The function $\gamma_2 = \Psi_2(\gamma_1)$ and the corresponding regions in the plane (γ_1, γ_2) where cases (a) or (b) of Lemma 6 hold. Parameter values are $R_0^1 = 4$, $R_0^2 = 2$, $\mu = 1$, $\xi_1 = 0.9$, $\xi_2 = 0.5$.

Since $|A|$ is an affine function of pe (see (2.45)), if it has the same sign at both ends of a segment, it will have the same sign also within, yielding a) and b).

To show c), through long computations, one arrives at

$$pe_2^{BP} - pe_1^{BP} = \frac{(R_0^1 - R_0^2)(R_0^2(1 - \xi_2) - R_0^1(1 - \xi_1))C}{R_0^1 R_0^2 (\xi_1 - \xi_2)} \\ \times \frac{\gamma_2(C(\mu + \gamma_1) - \mu R_0^2 \xi_2) + \mu R_0^2(\mu R_0^1(\xi_1 - \xi_2) + \gamma_1(R_0^1(\xi_1 - \xi_2) + \xi_2))}{(C\gamma_1 + \mu R_0^1 R_0^2(\xi_1 - \xi_2))(C\gamma_2 + \mu R_0^1 R_0^2(\xi_1 - \xi_2))}.$$

It is then easy to see that $pe_2^{BP} - pe_1^{BP} = 0$ if and only if $\gamma_2 = \Psi_2(\gamma_1)$. \square

We can now summarise the conclusions about the existence of the Coexistence Equilibrium.

Proposition 8. *Assume $R_0^1 > R_0^2$ and (2.35). Then there occur transcritical bifurcations at E_1 for $pe = pe_1^{BP}$, and at E_2 for $pe = pe_2^{BP}$ with the emergence of a positive coexistence equilibrium. Either*

- (a) $pe_1^{BP} < pe_2^{BP}$ and the coexistence equilibrium is unique and feasible for all $pe \in [pe_1^{BP}, pe_2^{BP}]$;
- (b) $pe_2^{BP} < pe_1^{BP}$ and the coexistence equilibrium is unique and feasible for all $pe \in [pe_2^{BP}, pe_1^{BP}]$.
- (c) $pe_2^{BP} = pe_1^{BP}$, and there is a continuum of positive equilibria for $pe = pe_1^{BP} = pe_2^{BP}$.

If (2.52) holds, (a) is true for all values of γ_1 and γ_2 .

Otherwise, (c) is true for $\gamma_2 = \Psi_2(\gamma_1)$; (b) is true for $\gamma_2 > \Psi_2(\gamma_1) > 0$; (a) is true for $\gamma_2 < \Psi_2(\gamma_1)$ and for all γ_2 when $\Psi_2(\gamma_1) < 0$, where $\Psi_2(\gamma_1)$ is defined in (2.59).

Note that $\Psi_2(\gamma_1) > \gamma_1$, so that, if $\gamma_1 \geq \gamma_2$, (a) is always true.

Stability

It is easy to show that in case (b) the coexistence equilibrium is always unstable. This can be proved by bifurcation theory, but can also be checked directly using

Lemma 7. *Let conditions (2.35) hold and let J be the Jacobian of (2.8)–(2.12) computed at the coexistence equilibrium. Then*

$$\text{sign}(|J|) = \text{sign}(|A|).$$

The proof is given in the Appendix.

It follows that in case (b), the Routh-Hurwitz stability conditions are violated, and the coexistence equilibrium is unstable.

As for case (a), bifurcation theory shows that the coexistence equilibrium is asymptotically stable for pe close to pe_1^{BP} and pe_2^{BP} .

We were not able to prove that Routh-Hurwitz stability conditions are satisfied for all $pe \in (pe_1^{BP}, pe_2^{BP})$. We then performed a numerical study drawing 1 million sets of parameters ($R_0^1, R_0^2, \xi_1, \xi_2, \gamma_1/\mu, \gamma_2/\mu$) satisfying conditions (2.35) and (2.52) or $\Psi_2(\gamma_1) < 0$ or $\gamma_2 < \Psi_2(\gamma_1)$. For each such draw, we divided the (pe_1^{BP}, pe_2^{BP}) interval into 10000 sub-intervals and, for each value of pe in this mesh, computed, through standard routines (Press et al., 1992), the eigenvalues of the Jacobian at the coexistence equilibrium. All the computed eigenvalues had negative real parts, suggesting that the coexistence equilibrium never loses its stability through Hopf bifurcations in the intervals (pe_1^{BP}, pe_2^{BP}) .

2.4 Examples

The case $pe_1^{BP} < pe_2^{BP}$ had already been numerically observed by Porco and Blower (1998). In this case coexistence occurs in the parameter region where the other existing equilibria are unstable. The case $pe_1^{BP} < pe_2^{BP}$ occurs for $\gamma_1 = \gamma_2$, $\gamma_1 > \gamma_2$ and may occur also for certain $\gamma_1 < \gamma_2$ as shown by the following example.

Example 1: Let $\gamma_1 = 0.015 < \gamma_2 = 0.517$ and

$$R_0^1 = 8.363, R_0^2 = 3.790, \mu = 0.423, \xi_1 = 0.990, \xi_2 = 0.020 \text{ and } \pi = 1$$

By substitution into (2.17), (2.32) and (2.40) we get

$$pe_1^{DF} = 0.888 \quad pe_2^{DF} = 36.496 \quad pe_1^{BP} = 0.153 \quad pe_2^{BP} = 0.552$$

The Subtype-1-Only Equilibrium is stable for $0 \leq pe < 0.153$, the Subtype-2-Only Equilibrium is stable for $0.552 < pe \leq 1$ and the Coexistence Equilibrium exists into the range $0.153 < pe < 0.552$, where the DFE, the Subtype-1-Only and the subtype-2-Only Equilibria exist but are unstable. Numerical computation of the eigenvalues of the linearized system confirm that the Coexistence Equilibrium is stable where it exists. Figure 2.3 shows the equilibrium fractions y_1 and y_2 as function of parameter pe .

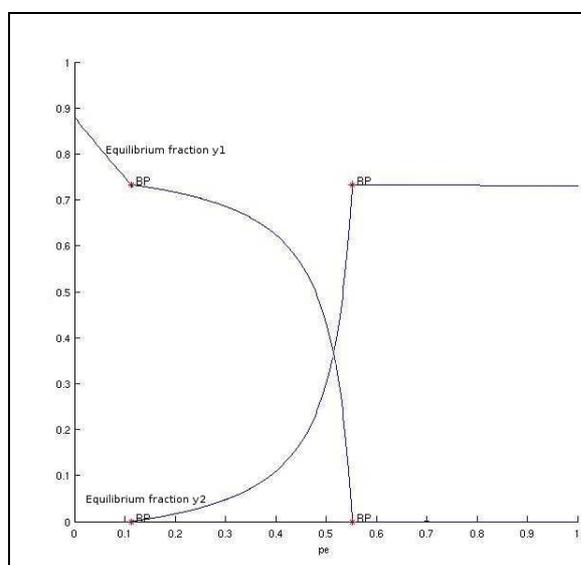


Figure 2.3: Equilibrium fractions y_1 and y_2 as function of pe for fixed $\gamma_1 = 0.015, \gamma_2 = 0.517, R_0^1 = 8.363, R_0^2 = 3.790, \mu = 0.423, \xi_1 = 0.990, \xi_2 = 0.020, \pi = 1$. Coexistence of the strains occurs for $0.153 < pe < 0.552$.

The case $pe_1^{BP} > pe_2^{BP}$ had never been observed before. In this case the Coexistence Equilibrium exists in a ‘bi-stability’ region in which both the Subtype-1-Only and the Subtype-2-Only Equilibrium are stable and hence the asymptotic behavior of the system depends on the initial conditions. This latter case occurs only for certain $\gamma_1 < \gamma_2$.

Example 2: Consider the case $\gamma_1 = 0.026 < \gamma_2 = 0.966$ and let

$$R_0^1 = 4.723, R_0^2 = 2.293, \mu = 0.235, \xi_1 = 0.923, \xi_2 = 0.650 \text{ and } \pi = 1.$$

2.4. Examples

By substitution into (2.17), (2.32) and (2.40) we get

$$pe_1^{DF} = 0.853 \quad pe_2^{DF} = 0.866 \quad pe_1^{BP} = 0.829 \quad pe_2^{BP} = 0.822$$

The DFE is stable for $pe > 0.866$, the Subtype-1-Only Equilibrium is stable for $0 < pe < 0.829$ and the Subtype-2-Only Equilibrium is stable for $0.822 < pe < 0.866$. The Coexistence Equilibrium exists for $0.822 < pe < 0.829$ and is unstable. Figure 2.4 shows the equilibrium fractions y_1 and y_2 as function of parameter pe . Figure 2.5 shows two trajectories for the equilibrium fractions y_1 and y_2 starting close to the Coexistence Equilibrium at $pe = 0.8234$ and converging one to the Subtype-1-Only Equilibrium and the other to the Subtype-2-Only Equilibrium. The bifurcation and trajectory graphs have been obtained by the graphical package MatCont of the MATLAB software.

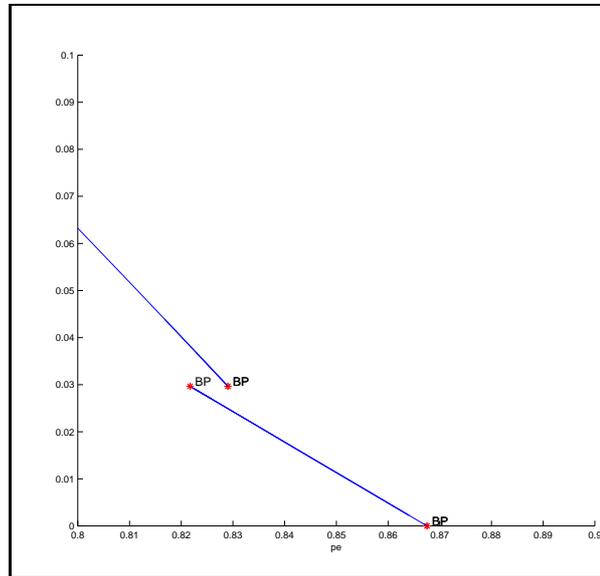


Figure 2.4: Equilibrium fractions y_1 and y_2 as function of pe for fixed $\gamma_1 = 0.026, \gamma_2 = 0.966, R_0^1 = 4.723, R_0^2 = 2.293, \mu = 0.235, \xi_1 = 0.923, \xi_2 = 0.650, \pi = 1$. A bi-stability region occurs for $0.822 < pe < 0.829$; in this region unstable coexistence of the strains occurs.

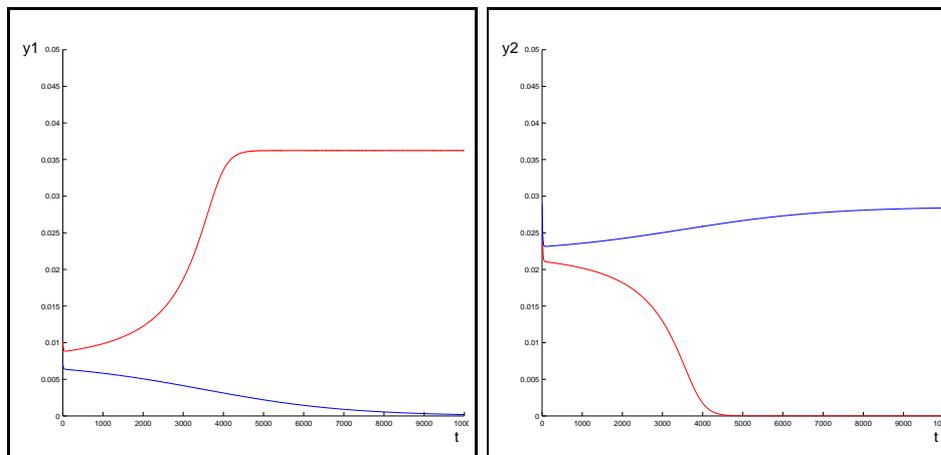


Figure 2.5: Trajectories of the fractions y_1 (left panel) and y_2 (right panel) as functions of time; parameter values are $\gamma_1 = 0.026$, $\gamma_2 = 0.966$, $R_0^1 = 4.723$, $R_0^2 = 2.293$, $\mu = 0.235$, $\xi_1 = 0.923$, $\xi_2 = 0.650$, $\pi = 1$ and $pe = 0.8234$. Both trajectories start close to the Coexistence Equilibrium $x = 0.148$, $y_1 = 0.007$, $y_2 = 0.022$, $N = 3.896$; the starting point of the red one, converging to the Subtype-1-Only Equilibrium, is $(0.148, 0.010, 0.029, 3.896)$; the starting point of the blue one, converging to the Subtype-2-Only Equilibrium, is $(0.148, 0.007, 0.029, 3.896)$.

2.5 Discussion

In this chapter we have analysed a model for competition between two viral strains with complete cross-immunity and imperfect vaccination. The model was first proposed by Porco and Blower (1998) with different HIV strains as case system; the authors showed through simulations the possibility that vaccination shifted the competitive hierarchy, with potential side-effects on public health.

Here we have examined the same model in greater detail, finding for instance the exact conditions under which vaccination may lead to coexistence of two strains; these are given by (2.35). It is worth commenting on their biological interpretation.

The first $\xi_1 > \xi_2$ means that the vaccine reduces more the susceptibility to the strain with the higher reproduction number (the better competitor in absence of vaccination) since we assumed $R_0^1 > R_0^2$.

2.5. Discussion

The second condition, that can be written as

$$\frac{\xi_2}{\xi_1} < \frac{R_0^1(R_0^2 - 1)}{R_0^2(R_0^1 - 1)}$$

specifies that the ratio of susceptibilities under vaccination must be decreased sufficiently relative to a ratio of reproduction numbers.

The third condition $R_0^2(1-\xi_2) > R_0^1(1-\xi_1)$ means that, if every individual were vaccinated, the second strain would have a higher reproduction number (note that the third condition implies the first one, which is then pleonastic).

Under these conditions there always exists a range of vaccination rates under which a (unique) coexistence equilibrium exists, at least if μ/γ_i is small enough, i.e. natural mortality is sufficiently lower than that induced by the infection (or, in case of HIV, than the rate of progressing into AIDS).

The relative values of γ_1 and γ_2 (i.e., of the expected lengths of sojourn in the classes I_1 and I_2) determine the ordering between pe_1^{BP} given by (2.32) and pe_2^{BP} given by (2.40). This in turn affects the qualitative behavior of system (2.14).

The case $pe_1^{BP} < pe_2^{BP}$ had already been numerically observed (Porco and Blower, 1998). In this case, coexistence occurs in the parameter region where all the other equilibria are unstable. Numerically, we found that the coexistence equilibrium is asymptotically stable for parameter values in this region, but the possibility of destabilization via Hopf bifurcations cannot be totally excluded, since an analytical proof is missing. The unconditional stability of the coexistence equilibrium has been proven in another model with coexistence of totally cross-immune pathogen strains (Andreasen and Pugliese, 1995).

On the other hand, the case $pe_1^{BP} > pe_2^{BP}$ is also possible, giving rise to phenomena that had not been anticipated. In this case there exists a parameter region in which both monomorphic equilibria (i.e. the Subtype-1-Only and the Subtype-2-Only Equilibrium) are stable and the coexistence equilibrium exists unstable (see Figure 2.4). In this ‘bi-stability’ region the asymptotic behavior of system (2.8)–(2.12) depends on the initial conditions. The presence of the bi-stability region implies that, with a gradual increase of vaccination rates, one may encounter a sudden shift from a situation with

only strain 1 present in appreciable proportion, to one with only strain 2. Moreover, decreasing again vaccination rates, one would see a hysteresis-type behavior.

As shown in the main text, bi-stability may occur only if $\gamma_1 < \gamma_2$. This means that the mortality rate (or rate of developing AIDS, in case of HIV) must be larger for strain 2 (the one that is out-competed without vaccination) than for strain 1. In other words, bi-stability may occur only if vaccination shifts the competitive balance in favour of a more virulent strain, a rather unpleasant scenario (Massad et al., 2006). Note that the model is definitely not realistic for HIV, mainly because its structure implies that the duration of the infectious stage is exponential, which is certainly not plausible, whether infectious are treated or not. The goal of our analysis is mainly exploratory to suggest possible phenomena that may be then examined (probably with the help of numerical software) in more realistic and complex models. On the other hand, the model can be applied to many other fatal diseases of S-I type, as long as one can assume that the entrance in the community (with or without vaccination) is constant and independent from the population size. One can reasonably expect that similar results would be obtained also under other assumptions for the birth rate, but the analysis would be more complex. Thus, these results should be of interest in the analysis of several emerging and re-emerging fatal infectious diseases.

2.6 Appendix

2.6.1 Computation of the Routh-Hurwitz coefficients

We report here the computations of a_1 , a_2 , a_3 and $a_1a_2 - a_3$ that lead us to prove that the Subtype-1-Only Equilibrium is stable, wherever it exists.

Remind that we set

$$\begin{aligned} K &= R_0^1(\mu + \gamma_1)(1 - \xi_1)\bar{y}_1 \\ L &= R_0^1(\mu + \gamma_1)\bar{y}_1 \end{aligned}$$

$$\begin{aligned} a_1 &= (\mu + \gamma_1\bar{y}_1) + [\mu + R_0^1(\mu + \gamma_1)\bar{y}_1] - \gamma_1\bar{y}_1 + R_0^1(\mu + \gamma_1)(1 - \xi_1)\bar{y}_1 \\ &= \mu + \mu + R_0^1(\mu + \gamma_1)\bar{y}_1 + R_0^1(\mu + \gamma_1)(1 - \xi_1)\bar{y}_1 \\ &= 2\mu + K + L \end{aligned}$$

$$\begin{aligned} a_2 &= (\mu + \gamma_1\bar{y}_1) [\mu + R_0^1(\mu + \gamma_1)\bar{y}_1] + \\ &\quad - [\mu + R_0^1(\mu + \gamma_1)\bar{y}_1] [\gamma_1\bar{y}_1 - R_0^1(\mu + \gamma_1)(1 - \xi_1)\bar{y}_1] + \\ &\quad + [R_0^1(\mu + \gamma_1) - \gamma_1] R_0^1(\mu + \gamma_1)\xi_1\bar{x}_1\bar{y}_1 + \\ &\quad - (\mu + \gamma_1\bar{y}_1) [\gamma_1\bar{y}_1 - R_0^1(\mu + \gamma_1)(1 - \xi_1)\bar{y}_1] + \\ &\quad + (\mu + \gamma_1\bar{y}_1)\gamma_1\bar{y}_1 \\ &= [\mu + R_0^1(\mu + \gamma_1)\bar{y}_1] [\mu + R_0^1(\mu + \gamma_1)(1 - \xi_1)\bar{y}_1] + \\ &\quad + [R_0^1(\mu + \gamma_1) - \gamma_1] R_0^1(\mu + \gamma_1)\xi_1\bar{x}_1\bar{y}_1 + \\ &\quad + (\mu + \gamma_1\bar{y}_1)R_0^1(\mu + \gamma_1)(1 - \xi_1)\bar{y}_1 \\ &= (\mu + L)(\mu + K) + [R_0^1(\mu + \gamma_1) - \gamma_1] L\xi_1\bar{x}_1 + (\mu + \gamma_1\bar{y}_1)K \end{aligned}$$

$$\begin{aligned} a_3 &= (\mu + \gamma_1\bar{y}_1) \left\{ [R_0^1(\mu + \gamma_1) - \gamma_1] R_0^1(\mu + \gamma_1)\xi_1\bar{x}_1\bar{y}_1 - [\mu + R_0^1(\mu + \gamma_1)\bar{y}_1] [\gamma_1\bar{y}_1 + \right. \\ &\quad \left. - R_0^1(\mu + \gamma_1)(1 - \xi_1)\bar{y}_1] \right\} + \gamma_1\bar{N} [\gamma_1\bar{y}_1 - R_0^1(\mu + \gamma_1)\bar{y}_1] R_0^1(\mu + \gamma_1)\xi_1 \frac{\bar{x}_1\bar{y}_1}{\bar{N}} + \\ &\quad + (\mu + \gamma_1\bar{y}_1) [\mu + R_0^1(\mu + \gamma_1)\bar{y}_1] \frac{\bar{y}_1}{\bar{N}} \left. \right\} \end{aligned}$$

$$\begin{aligned}
 &= (\mu + \gamma_1 \bar{y}_1) \{ [R_0^1(\mu + \gamma_1) - \gamma_1] R_0^1(\mu + \gamma_1) \xi_1 \bar{x}_1 \bar{y}_1 - [\mu + R_0^1(\mu + \gamma_1) \bar{y}_1] [\gamma_1 \bar{y}_1 + \\
 &- R_0^1(\mu + \gamma_1)(1 - \xi_1) \bar{y}_1] \} - [R_0^1(\mu + \gamma_1) - \gamma_1] R_0^1(\mu + \gamma_1) \xi_1 \bar{x}_1 \bar{y}_1 \gamma_1 \bar{y}_1 + \\
 &+ (\mu + \gamma_1 \bar{y}_1) [\mu + R_0^1(\mu + \gamma_1) \bar{y}_1] \gamma_1 \bar{y}_1 \\
 &= [R_0^1(\mu + \gamma_1) - \gamma_1] L \mu \xi_1 \bar{x}_1 + (\mu + \gamma_1 \bar{y}_1)(\mu + L)K
 \end{aligned}$$

$$\begin{aligned}
 a_1 a_2 - a_3 &= [\mu + R_0^1(\mu + \gamma_1) \bar{y}_1] [\mu + R_0^1(\mu + \gamma_1)(1 - \xi_1) \bar{y}_1] [2\mu + R_0^1(\mu + \gamma_1) \bar{y}_1 + \\
 &+ R_0^1(\mu + \gamma_1)(1 - \xi_1) \bar{y}_1] + [R_0^1(\mu + \gamma_1) - \gamma_1] R_0^1(\mu + \gamma_1) \xi_1 \bar{x}_1 \bar{y}_1 [2\mu + \\
 &+ R_0^1(\mu + \gamma_1) \bar{y}_1 + R_0^1(\mu + \gamma_1)(1 - \xi_1) \bar{y}_1] + \\
 &+ (\mu + \gamma_1 \bar{y}_1) R_0^1(\mu + \gamma_1)(1 - \xi_1) \bar{y}_1 \cdot \\
 &[2\mu + R_0^1(\mu + \gamma_1) \bar{y}_1 + R_0^1(\mu + \gamma_1)(1 - \xi_1) \bar{y}_1] + \\
 &- [R_0^1(\mu + \gamma_1) - \gamma_1] R_0^1(\mu + \gamma_1) \xi_1 \bar{x}_1 \bar{y}_1 + \\
 &- (\mu + \gamma_1 \bar{y}_1) [\mu + R_0^1(\mu + \gamma_1) \bar{y}_1] R_0^1(\mu + \gamma_1)(1 - \xi_1) \bar{y}_1 \\
 &= (\mu + L)(\mu + K)(2\mu + L + K) + \\
 &+ [R_0^1(\mu + \gamma_1) - \gamma_1] L \xi_1 \bar{x}_1 (\mu + L + K) + \\
 &+ (\mu + \gamma_1 \bar{y}_1)(\mu + K)K
 \end{aligned}$$

2.6.2 Proof of Lemma 2

Proof. a) It is clear that, under the assumptions (2.30), if $R_0^2(1 - \xi_2) > R_0^1(1 - \xi_1)$, then $0 < pe_1^{BP}$. As for the other inequality, if $\mu = 0$,

$$pe_1^{BP} = \frac{(R_0^1 - R_0^2)(1 - \xi_1)}{R_0^2(\xi_1 - \xi_2)}$$

and

$$\frac{(R_0^1 - R_0^2)(1 - \xi_1)}{R_0^2(\xi_1 - \xi_2)} < 1 \iff R_0^1(1 - \xi_1) < R_0^2(1 - \xi_2).$$

By continuity, if $R_0^2(1 - \xi_2) > R_0^1(1 - \xi_1)$, $pe_1^{BP} < 1$ for $\mu > 0$ small enough.

b) If $R_0^2(1 - \xi_2) = R_0^1(1 - \xi_1)$, $pe_1^{BP} \equiv 1$ for all $\mu > 0$.

If $R_0^2(1 - \xi_2) < R_0^1(1 - \xi_1)$, pe_1^{BP} is a decreasing continuous function of μ on $[0, +\infty)$.

2.6. Appendix

Hence

$$pe_1^{BP} > p_\infty := \lim_{\mu \rightarrow +\infty} pe_1^{BP} = \frac{(R_0^1 - R_0^2)}{R_0^2(\xi_1 - \xi_2)} \left[1 - \xi_1 + \frac{\xi_1(R_0^2(1 - \xi_2) - R_0^1(1 - \xi_1))}{R_0^1 R_0^2(\xi_1 - \xi_2)} \right].$$

Now

$$\begin{aligned} p_\infty - 1 &= \frac{(R_0^1 - R_0^2)(1 - \xi_1) - R_0^2(\xi_1 - \xi_2)}{R_0^2(\xi_1 - \xi_2)} + \frac{(R_0^1 - R_0^2)\xi_1(R_0^2(1 - \xi_2) - R_0^1(1 - \xi_1))}{R_0^2(\xi_1 - \xi_2)R_0^1 R_0^2(\xi_1 - \xi_2)} \\ &= \frac{R_0^1(1 - \xi_1) - R_0^2(1 - \xi_2)}{R_0^2(\xi_1 - \xi_2)} \left(1 - \frac{(R_0^1 - R_0^2)\xi_1}{R_0^1 R_0^2(\xi_1 - \xi_2)} \right). \end{aligned}$$

We use the inequality $(R_0^1 - R_0^2)\xi_1 < R_0^1\xi_1 - R_0^2\xi_2$ in the bracketed term to have

$$1 - \frac{(R_0^1 - R_0^2)\xi_1}{R_0^1 R_0^2(\xi_1 - \xi_2)} > 1 - \frac{R_0^1\xi_1 - R_0^2\xi_2}{R_0^1 R_0^2(\xi_1 - \xi_2)} = \frac{C}{R_0^1 R_0^2(\xi_1 - \xi_2)} > 0,$$

proving $p_\infty - 1 > 0$. □

2.6.3 Proof of Lemma 7

Proof. Let J be the Jacobian matrix at the coexistence equilibrium:

$$J = \begin{pmatrix} -\frac{\pi}{\hat{N}} & 0 & -\gamma_1 \hat{N} & -\gamma_2 \hat{N} \\ -\frac{\pi}{\hat{N}^2}(1 - \hat{x} - pe) & -\frac{\pi}{\hat{N}} - a & -\hat{x}[R_0^1(\mu + \gamma_1) - \gamma_1] & -\hat{x}[R_0^2(\mu + \gamma_2) - \gamma_2] \\ \frac{\pi}{\hat{N}^2} \hat{y}_1 & \hat{y}_1 R_0^1(\mu + \gamma_1) \xi_1 & -\hat{y}_1 b + \gamma_1 \hat{y}_1 & -\hat{y}_1 b + \gamma_2 \hat{y}_1 \\ \frac{\pi}{\hat{N}^2} \hat{y}_2 & \hat{y}_2 R_0^2(\mu + \gamma_2) \xi_2 & -\hat{y}_2 c + \gamma_1 \hat{y}_2 & -\hat{y}_2 c + \gamma_2 \hat{y}_2 \end{pmatrix}.$$

where

$$\begin{aligned} a &= R_0^1(\mu + \gamma_1)\hat{y}_1 - \gamma_1 \hat{y}_1 + R_0^2(\mu + \gamma_2)\hat{y}_2 - \gamma_2 \hat{y}_2 \\ b &= R_0^1(\mu + \gamma_1)(1 - \xi_1) \\ c &= R_0^2(\mu + \gamma_2)(1 - \xi_2) \end{aligned} \tag{2.60}$$

We apply the Gauss-Jordan algorithm in the following steps:

- 1) substitute the fourth row of J with the sum of its fourth row multiplied times \hat{y}_1 and its third row multiplied times $-\hat{y}_2$, thus obtaining matrix J_1
- 2) substitute the third column of matrix J_1 with the sum of the its third column and its fourth column multiplied times -1 , thus obtaining matrix J_2
- 3) substitute the third row of matrix J_2 with the sum of its third row multiplied

times \hat{N} and its first row multiplied times \hat{y}_1 , thus obtaining matrix

$$J_3 = \begin{pmatrix} -\frac{\pi}{\hat{N}^2} & -\frac{\pi}{\hat{N}} & 0 & (\gamma_2 - \gamma_1)\hat{N} & -\gamma_2\hat{N} \\ -\frac{\pi}{\hat{N}^2}(1 - \hat{x} - pe) & -\frac{\pi}{\hat{N}} - a & \hat{x}d & -\hat{x}[R_0^2(\mu + \gamma_2) - \gamma_2] & \\ 0 & \hat{N}R_0^1\hat{y}_1(\mu + \gamma_1)\xi_1 & 0 & -\hat{N}R_0^1\hat{y}_1(\mu + \gamma_1)(1 - \xi_1) & \\ 0 & \hat{y}_1\hat{y}_2e & 0 & \hat{y}_1\hat{y}_2(b - c) & \end{pmatrix}.$$

where a , b and c are given in (2.60) and

$$\begin{aligned} d &= R_0^2(\mu + \gamma_2) - \gamma_2 - R_0^1(\mu + \gamma_1) + \gamma_1 \\ e &= R_0^2(\mu + \gamma_2)\xi_2 - R_0^1(\mu + \gamma_1)\xi_1 \end{aligned}$$

Due to the properties of the determinant, we have

$$|J| = |J_3|. \quad (2.61)$$

We compute $|J_3|$ expanding through its first column obtaining

$$|J_3| = \frac{|A|\pi R_0^1 R_0^2(\mu + \gamma_1)(\mu + \gamma_2)\hat{y}_1^2\hat{y}_2(\xi_1 - \xi_2)}{\mu} \quad (2.62)$$

where $|A|$ is given in (2.45). Conditions (2.35) and identities (2.61) and (2.62) imply our claim. \square

Chapter 3

Modelling the Spatial Spread of H7N1 Avian Influenza Virus among Poultry Farms in Italy

3.1 Introduction

In 1999 – 2000 the Italian industrial poultry production was disrupted by an epidemic of Highly Pathogenic Avian Influenza (HPAI) caused by a H7N1 virus subtype. Since March 1999, the Low Pathogenic (LPAI) H7N1 virus subtype was endemically circulating in the North of Italy, where more than 65% of the Italian poultry production is concentrated, and the currently accepted hypothesis is that a H7N1 LPAI strain mutated into a HPAI strain (Busani et al., 2009; Mannelli et al., 2007). This hypothesis has been widely discussed in the literature (Webster et al., 1992; Alexander et al., 2000; Stegeman, 2004). HPAI virus was first detected in a poultry farm on November 28th, 1999; after that, the measures provided by the European Union (EU) legislation ¹ were applied, at different times at various spatial scales and were continued until the infection was officially eradicated on April 10th, 2000.

The first goal of this study was to investigate whether a spatial transmission kernel was adequate for describing the actual epidemic spread in Northern Italy, considering also the implemented control measures. We were furthermore interested in

¹CEC. Council Directive 92/40/EEC of 19 May 1992 introducing community measures for the control of avian influenza. Official Journal of the European Commission 1992 L167:1-15

3.2. Data

analysing potential differences in susceptibility among poultry species, consistent with the association found between AI virus infection and poultry species (Busani et al., 2009; Mannelli et al., 2006; Thomas et al., 2005), and with other reports of species differences in susceptibility to high pathogenicity viruses (Tumpey et al., 2004). Mannelli et al. (2007) found a reduction in transmissibility during the course of the epidemic, using a non-spatial model. We then analysed whether this claim could be upheld using a more detailed spatial model. Finally, we assessed the effectiveness of the measures implemented in first containing and then eradicating the infection in order to discuss the relative merit of each specific measure, and to study whether a different implementation of the measures could have been more effective.

Here we analyse the spatio-temporal spread of the infection first using a SEIR model with a spatial kernel similar to the one proposed by Boender et al. (2007), to which all control measures were added just as they were actually implemented. We use maximum likelihood methods to estimate parameters and to establish their confidence intervals.

We then extend the SEIR model allowing for species differences in susceptibility, and test the improvement of fit relative to the basic model. We also allow for changes in transmissibility during the course of the epidemic (Mannelli et al., 2007), considering different epidemic phases, corresponding to steps in the implementation of control measure, and to awareness of the ongoing epidemic.

3.2 Data

The study area of this work consists of the North-Eastern regions of Lombardia and Veneto, where 392 out of 413 (94.9%) outbreaks occurred. Due to the lack of data on 10 infected backyard farms located in the study area, we considered 382 outbreaks in our analysis (Capua and Marangon, 2000; Mannelli et al., 2007). In these regions there is a densely populated poultry area where different avian species (laying hens, broilers, breeders of different species, meat turkeys, geese, quails, ostriches and others) are bred.

Poultry production consists of repeated cycles. A productive cycle starts with the stocking of the one day old chicks (typically all of the same poultry type) and, after a period whose length depends on the species (on average: 42 days for broilers, 95 – 145 days for female and male meat turkeys, up to 2 years for laying hens), it ends with the slaughtering of the whole flock. Between successive production

cycles there is usually an “empty period” during which no birds are stocked since the buildings have to be cleaned and sanitized, and maintenance procedures need to be performed.

The study period started on November 28th, 1999 (i.e. the day that HPAI virus was first suspected of having infected a poultry farm) and ended on April 10th 2000 (the day that infection was eradicated). The data-sets upon which this work is based have been collected by the Istituto Zooprofilattico Sperimentale delle Venezie (IZSVe) and have already been subject to analysis (Busani et al., 2009; Mannelli et al., 2006, 2007). Data on species and production type, duration of each production cycle and geographical location of the farms in the study area were collected by veterinarians working for the Regional Veterinary Service (Busani et al., 2009). The geographic distance between every pair of farms in the data-set has also been computed.

To contain the epidemic, the following measures outlined by EU legislation¹ were applied starting from December 17th, 1999: the stamping-out of infected or suspected of being infected farms (IF) and the ban of restocking (BR) on emptied farms (either because they ended a production cycle in the at-risk area during the epidemic or because they were culled) (Busani et al., 2009; Mannelli et al., 2007). Pre-emptive culling (PEC) of farms located at less than 1 km from an infected farm started in Veneto from January 20th, 2000 (Busani et al., 2009) and in Lombardia from February 10th, 2000 (Mannelli et al., 2007). Further measures such as the pre-emptive slaughter at farms that had at-risk contacts with an IF and strict limitations to the movements of live poultry, products, vehicles and staff were also applied in the whole study area (Busani et al., 2009). As the epidemic unfolded, the IZSVe recorded the date of onset of clinical signs (for every confirmed case) and the date of culling (either because infected or because pre-emptively slaughtered) of every farm that underwent this measure.

During the study period 382 farms were infected (red dots in Figure 3.1), 72 (65 in Veneto and 7 in Lombardia) were pre-emptively slaughtered (blue dots in Figure 3.1), the ban on restocking was imposed on 1486 farms (yellow dots in Figure 3.1) and 1307 escaped the infection (green dots in Figure 3.1). H7N1 virus spread to the maximum distance of 176.18 km from the source farm. For every farm in the study area, starting and ending dates of each production cycle have been recorded.

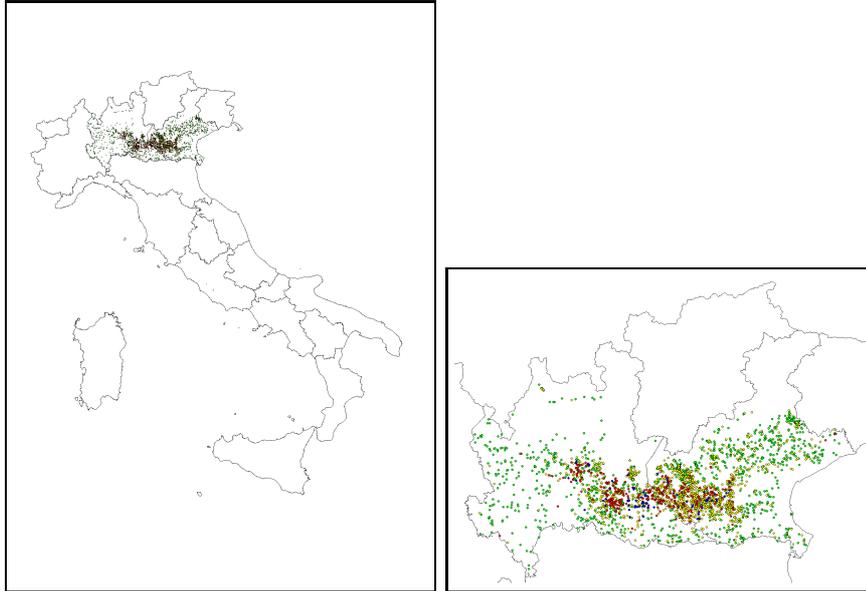


Figure 3.1: Infected farms (red dots), not infected farms (green dots), farms banned from restocking (yellow dots) and pre-emptive culled farms (blue dots) in the HPAI epidemic of years 1999-2000 in Italy (left panel) and in the study area (Veneto and Lombardia) (right panel)

3.3 Models Analysed

The SEIR models are defined on a farm level (i.e. the farms are the individual units) and our assumptions are similar to those made by Boender et al. (2007) to model the diffusion of HPAI in The Netherlands. Time is discrete and for each of the 135 days (November 28th, 1999 - April 10th, 2000) farms in a production cycle are classified as susceptible (S), latently infected (i.e. infected but not yet infectious) (E), infected (I) or removed (either because they were culled or because they were banned from restocking) (R). Farms are considered removed (R) also when they are in the “empty period” between successive production cycles. Following Busani et al. (2009), we assumed that infection occurred 7 days before the detection of first symptoms and this included a period of latency of 2 days (Van der Goot et al., 2003); the infectious period lasted until the day of culling (Busani et al., 2009). With these assumptions, the average length of the infectious period was $T = 11.82$ days (5 – 95 percentile interval (6, 26)).

In the Basic Model, transmission of infection between an infectious farms j and a

susceptible farm i at distance r_{ij} can occur (in a given day) with probability

$$h(r_{ij}) = \frac{h_0}{1 + \left(\frac{r_{ij}}{r_0}\right)^\alpha}. \quad (3.1)$$

This is the same transmission kernel as used by Boender et al. (2007).

The parameters h_0 , r_0 and α have been estimated by maximum likelihood (ML), while other parameters have been kept fixed. Sensitivity analysis (not shown here) on the lengths of the incubation period (3, 5, 7 days) and of the latently infected period (1, 2 days) show that the results obtained are robust to the exact choice of incubation and latency period.

In the Susceptibility Model, farms are divided into 5 groups, according to the species produced during the epidemic. Precisely, the species are: laying hens (1), meat turkeys (2), broilers (3), breeders (turkeys and chickens considered together) (4), others (5). The transmission kernel (3.1) is modified by substituting the constant h_0 with h_k where k ($= 1 \dots 5$) represents the species.

In the models with varying transmission rate, the transmission constant (h_0 or h_k) changes with time, according to the epidemic phase (see details in next Section). In all models, the force of infection on a susceptible farm i at time t $\lambda_i(t)$ is given by

$$\lambda_i(t) = \sum_{j \neq i} h(r_{ij})$$

where the sum is performed over all infectious farms j at time t .

The overall model is a simple discrete stochastic model, where, given the state of the system at time t , each non-infected farm i independently becomes infected with probability $1 - e^{-\lambda_i(t)}$, while infected farms progress through the latent or infectious period.

The likelihood of the observed events can then be computed by multiplying (for each time t) the probabilities of becoming infected for each farm infected that day, times the probabilities of not becoming infected for each farm not infected that day. This can be computed in an efficient way (Boender et al., 2007) by dividing farms into the following sets: M (farms infected at time t_{inf}), K (farms not infected and not pre-emptively culled within the end of the epidemic at time t_{max}), Λ (farms not infected and pre-emptively culled at time t_{cul}) and B (farms not infected and banned from restocking at time t_{ban}). Then the log-likelihood function can be

3.4. Parameter Estimates

written as follows

$$\begin{aligned}
 l = & - \sum_{k \in K} \sum_{t=1}^{t_{max}-1} \lambda_k(t) - \sum_{l \in \Lambda} \sum_{t=1}^{t_{cul,l}-1} \lambda_l(t) - \sum_{b \in B} \sum_{t=1}^{t_{ban,b}-1} \lambda_b(t) + \\
 & - \sum_{m \in M} \sum_{t=1}^{t_{inf,m}-1} \lambda_m(t) + \sum_{m \in M} \log[1 - e^{-\lambda_m(t_{inf,m})}]
 \end{aligned}$$

3.4 Parameter Estimates

We computed the maximum likelihood estimates (MLE) of the parameters, the relative 95% confidence intervals and AIC index for the Basic and the Susceptibility Model. Confidence intervals have been computed by finite difference approximation of the inverse of the Hessian matrix of the log-likelihood function, which is the natural plug-in estimator of the the Fisher information matrix (Rice, 2004). The MLE of the parameters of interest, which have been computed by implementing the simulated annealing algorithm given by Press et al. (2002), the value of the log-likelihood function at the MLE and the AIC indexes are given in Tables 3.1 and 3.2.

Table 3.1: MLE and 95% Confidence Intervals of the Basic Model's parameters

	Estimate	95% Confidence Interval
h_0	0.0064	(0.0037, 0.0090)
r_0	2.1524	(1.3943, 2.9106)
α	2.0760	(1.8711, 2.2809)
log-likelihood	-2430.4558	
AIC	4866.9116	

According to Akaike's Criterion (Akaike, 1974), we find that the SEIR model with different susceptibility according to the species better explains the data. As for the changes in transmissibility during the course of the epidemic, we divided the study period into 4 phases: the first 19 days (Phase 1), during which no control or containment measures were undertaken; the next 34 days ($20 \leq t \leq 53$, Phase 2) during which stamping-out of IF and a ban on restocking (BR) were applied on the whole study area; the successive 20 days ($54 \leq t \leq 74$, Phase 3) during which pre-emptive culling (PEC) of farms located at less than 1 km from an IF, beyond IF culling and BR, was applied in Veneto; the remaining 61 days ($75 \leq t \leq 135$, Phase 4) during which culling of IF, BR and PEC were applied in the whole study

Table 3.2: MLE and 95% Confidence Intervals of the Susceptibility Model's parameters

	Estimate	95% Confidence Interval
h_1	0.009562	(0.0049, 0.0141)
h_2	0.007010	(0.0034, 0.0105)
h_3	0.001000	(0.0004, 0.0015)
h_4	0.005273	(0.0022, 0.0083)
h_5	0.001190	(0.0003, 0.0020)
r_0	3.0908	(1.7853, 4.3963)
α	2.1850	(1.9037, 2.4663)
log-likelihood	-2294.6860	
AIC	4603.372	

area.

The temporal changes in transmissibility were first explored on the Basic SEIR Model (i.e. without distinction among the different species). In the 4-Phases Basic Model each phase had a different transmissibility coefficient h_0 (h_0^i , $i = 1 \dots 4$) in equation (3.1). A 2-Phases Basic Model has also been analysed where only Phase 1 had a different transmissibility coefficient (h_0^1 vs. h_0^2 for all subsequent phases); the 2-Phases Model was considered on the basis of the results from the 4-Phases Model, but can also be justified because the change in transmissibility may be due to the limitations introduced to the movements of live poultry, products, vehicles and staff.

The maximum likelihood estimates of the parameters of the three variations of the Basic SEIR Model, together with the value of the log-likelihood function at the MLE and the AIC index are given in Table 3.3. By means of the log-likelihood ratio test and the assumption that the test statistic is asymptotically χ^2 distributed with the degree of freedom equal to the difference in dimensionality of the parameters' space of the tested models, we see that both the 2-Phases Basic Model and the 4-Phases Basic Model better explain our data at a significance level of 0.01, compared to the Basic SEIR Model. On the contrary, the 4-Phases Basic Model does not produce a (significantly) better fit when compared to the 2-Phases Basic Model. Akaike's Criterion is slightly lower for the 4-Phases Basic Model than the 2-Phases Basic Model, but the difference is too small to justify a more complex model (Table 3.3).

When considering the model with temporal phases and different host susceptibility, the number of parameters becomes too large to obtain reliable ML estimates. We defined a 2-Phases Susceptibility Model associating a constant reduction of

3.4. Parameter Estimates

Table 3.3: MLE of the Basic SEIR Model's parameters

	Basic Model	2-Phases Model	4-Phases Model
h_0^1	0.0064	0.0107	0.0104
h_0^2		0.0062	0.0056
h_0^3			0.0067
h_0^4			0.0078
r_0	2.1524	2.1340	2.1824
α	2.0760	2.0717	2.0830
log-likelihood	-2430.4558	-2426.9912	-2424.6035
AIC	4866.9116	4861.9824	4861.2070

transmissibility between the 2 phases, independently from the species. The reduction factor c between the 2 phases was fixed at 0.58, which is the value obtained with the 2-Phases Basic Model. We moreover fixed the proportionalities among the susceptibility of the species at the values obtained with the Susceptibility Model

$$r_2 = \frac{h_2}{h_1} = 0.73, \quad r_3 = \frac{h_3}{h_1} = 0.10, \quad r_4 = \frac{h_4}{h_1} = 0.55, \quad r_5 = \frac{h_5}{h_1} = 0.12.$$

With these assumptions, the transmissibility constant h_k^t to species k at time t is given by

$$h_k^t = \begin{cases} r_k h_1 & \text{if } t \leq 19 \\ cr_k h_1 & \text{if } t \geq 20 \end{cases} \quad (3.2)$$

where $r_1 = 1$, and the only unknown quantity to estimate is h_1 . The MLE of the 2-Phases Susceptibility Models and the relative 95% confidence intervals are given in Table 3.4 while the corresponding values of the transmissibility constants using (3.2) are given in Table 3.5. By Akaike's Criterion the data are better explained by the 2-Phases Susceptibility Model than by the Susceptibility Model. However, the 2-Phases Susceptibility Model improves the log-likelihood estimate of just 2.1 units (see Tables 3.2 and 3.4) while the large AIC reduction comes mainly from its low number of parameters; the low number of parameters (just 3) comes out of the fact that we have fixed several proportionality factors (see (3.2)) at the values obtained from previous analysis. Hence, the statistical assumptions needed to compare models through the AIC are not met, and it is not possible to choose the model on this basis solely.

Table 3.4: MLE and 95% Confidence Intervals of the 2-Phases Susceptibility Model's parameters

	Estimate	95% Confidence Interval
h_1	0.0155	(0.0078, 0.0232)
r_0	3.1595	(1.7703, 4.5487)
α	2.1921	(1.8937, 2.4904)
log-likelihood	-2292.5117	
AIC	4591.0234	

Table 3.5: Values of the transmissibility constants for the 2-Phases Susceptibility Model

	Phase 1	Phase 2
h_1	0.0155	0.0090
h_2	0.0113	0.0065
h_3	0.0015	0.0009
h_4	0.0085	0.0049
h_5	0.0018	0.0010

3.5 Simulations

3.5.1 How data are reproduced by the model

We simulated AI epidemics using the Basic SEIR Model, the Susceptibility Model and their 2-Phases versions in order to compare them to observed data and assess their behavior.

We assumed the observed spatial distribution of farms in the study area and the start of the epidemic from the first infected farm at time $t = 1$. We assigned the observed production cycles to the farms which were not infected during the 1999-2000 outbreak. Infected farms were assigned the observed production cycle until the day of infection; the production cycle was then completed from the distribution of the observed production cycles, according to the species. The length of the infectious period of each IF was randomly drawn from the observed infectious periods. Note that in this way we were accounting for Ban on Restocking (BR) and IF stamping-out since time $t = 20$. We assumed that PEC started at time $t = 54$ in Veneto and at time $t = 75$ in Lombardia and that it took a random number of days between 0 and 4 to cull an identified contiguous farm. Finally, we let epidemics evolve until extinction according to each of the considered models.

We tested the simulation results on the following indicators: the mean number of infected farms (or total case) (IF), the mean number of pre-emptively culled

3.5. Simulations

farms (PEC), the mean number of farms banned from restocking (BR), the average extinction time (T_{ext}) and the average maximum distance to which H7N1 spread (D_{max}). The number of culled farms (because either infected or pre-emptively culled) gives a measure of the total losses (TL) due to the epidemic.

For each model, we averaged over 100 realizations that generated at least 10 cases each; the number of replicates was suggested by similar analyses in the literature (Keeling et al., 2001; Matthews et al., 2003). Tables 3.6 and 3.7 show the average values, together with the relative 5 – 95 percentile intervals, obtained using either the Basic SEIR Model or the Susceptibility Model, with the respective 2-Phases variations. Note that, according to the criteria for applying the PEC, a total number of 129 (instead of 72) farms should have been pre-emptively culled; thus we expect that models that include culling at all farms within a radius of 1 km from an IF will produce a higher number of culling than actually observed.

Table 3.6: Mean numbers and 5 – 95 percentile intervals computed on 100 realizations that generated at least 10 cases, using the Basic Model and the 2-Phases Basic Model

	Basic SEIR Model	2-Phases Basic Model	Observed Data
IF	169.82 (21, 361)	366.25 (117, 575)	382
PEC	100.47 (10, 180)	147.72(70, 203)	72
BR	1447.26(1105, 1638)	1307.13 (953, 1523)	1486
T_{ext}	123.53(77, 162)	133.89 (104, 177)	135
D_{max}	118.22(58.69, 181.35)	144.89(102.23, 190.09)	176.18

Table 3.7: Mean numbers and 5 – 95 percentile intervals computed on 100 realizations that generated at least 10 cases, using the Susceptibility Model and the 2-Phases Susceptibility Model

	Susceptibility Model	2-Phases Susceptibility Model	Observed Data
IF	266.53(90, 448)	385.01(196, 530)	382
PEC	139.34(79, 196)	136.51(82, 179)	72
BR	1403.06(1089, 1587)	1383.41(1125,1535)	1486
T_{ext}	135.64(104, 181)	130.17(109,162)	135
D_{max}	139.33(89.97, 193.38)	151.07(108.48, 196.32)	176.18

The numbers reported in Tables 3.6 and 3.7 show that the indicators produced by the models are reasonably consistent with the data. The 2-Phases versions of the

models predict mean values of indicators closer to observed data; the agreement is further improved when taking into account difference by species in the susceptibility to HPAI infection. The maximum predicted distance of virus spread, on average, is of 151.07 km from the source farm, which is less than the observed distance (176.18 km).

In order to have a more complete comparison between data and simulations, we plot (Figure 3.2) the 3-day running (moving) averages (to remove extreme fluctuations) of the data against the 3-day running averages of the 100 realizations of the the 2-Phases Susceptibility Model. In Figure 3.3 we compare the 3-day running average of the data to the trajectories in time (3-day running averages) of four realizations: those yielding the 20th, 40th 60th and 80th percentiles of the total number of cases. Finally, Figure 3.4 shows one simulation of the spatial diffusion of infection generated with the 2-Phases Susceptibility Model.

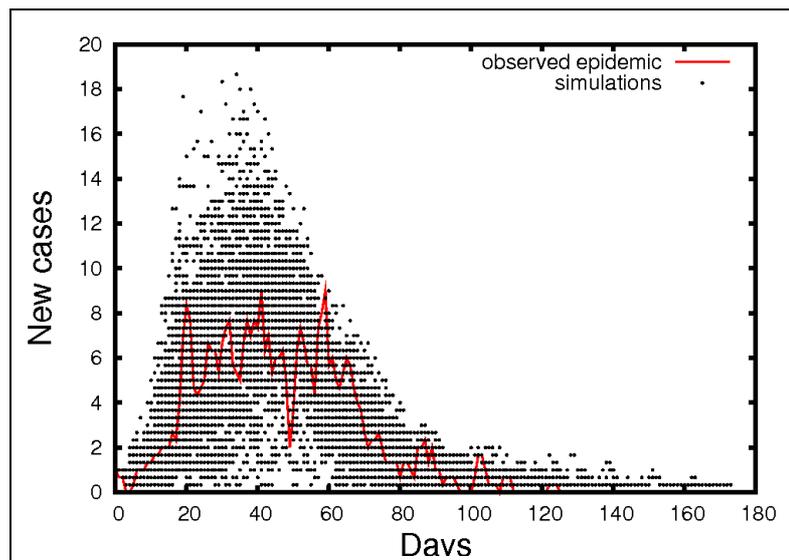


Figure 3.2: Comparison of the number of new cases between the 3-day running average of the observed epidemic and of 100 replicates of the stochastic 2-Phases Susceptibility Model

3.5.2 Assessment of the effectiveness of the intervention measures

In order to assess the effectiveness of pre-emptive culling (PEC) and of the imposition of the ban on restocking (BR), we first explored the effect of neglecting them. Every test in this section was conducted on the 2-Phases Susceptibility Model. The

3.5. Simulations

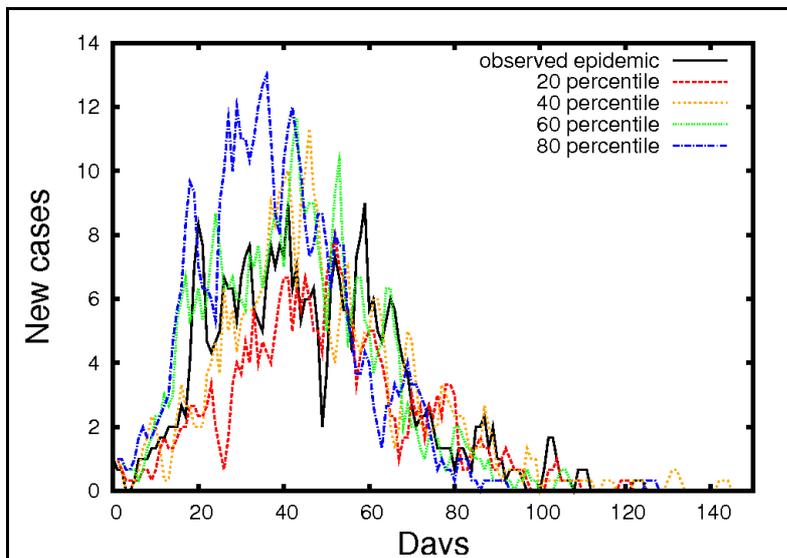


Figure 3.3: Comparison of the number of new cases between the 3-day running average of the observed epidemic and the 20th, 40th, 60th and 80th of 100 replicates of the stochastic 2-Phases Susceptibility Model

average quantities obtained with the 2-Phases Susceptibility Model (see Table 3.7) constitute our baseline.

When neglecting BR, we assumed that every farm is susceptible for the whole course of the epidemic. In Table 3.8 we report the average values of the chosen indicators when neglecting the imposition of ban of restocking (NO-BR) and when neglecting the application of pre-emptive culling of farms close to an infectious premise (NO-PEC). From our results (see Table 3.8) we conclude that the most effective intervention measure in stopping the infection was the imposition of ban of restocking on emptied farms.

Table 3.8: Mean numbers and 5–95 percentile intervals computed on 100 realizations that generated at least 10 cases, using the 2-Phases Susceptibility Model with and without BR or PEC

	NO-BR	NO-PEC	baseline
IF	984.02(817,1103)	421.81(202, 592)	385.01(196, 530)
PEC	496.89(368, 671)	0	136.51(82, 179)
BR	0	1326.67(1055, 1573)	1383.41(1125,1535)
T_{ext}	147.13(123, 170)	143.95(111, 179)	130.17(109,162)
D_{max}	168.19(135.21, 203.17)	148.69(114.96,189.35)	151.07(108.48, 196.32)
TL	1480.91	421.81	521.52

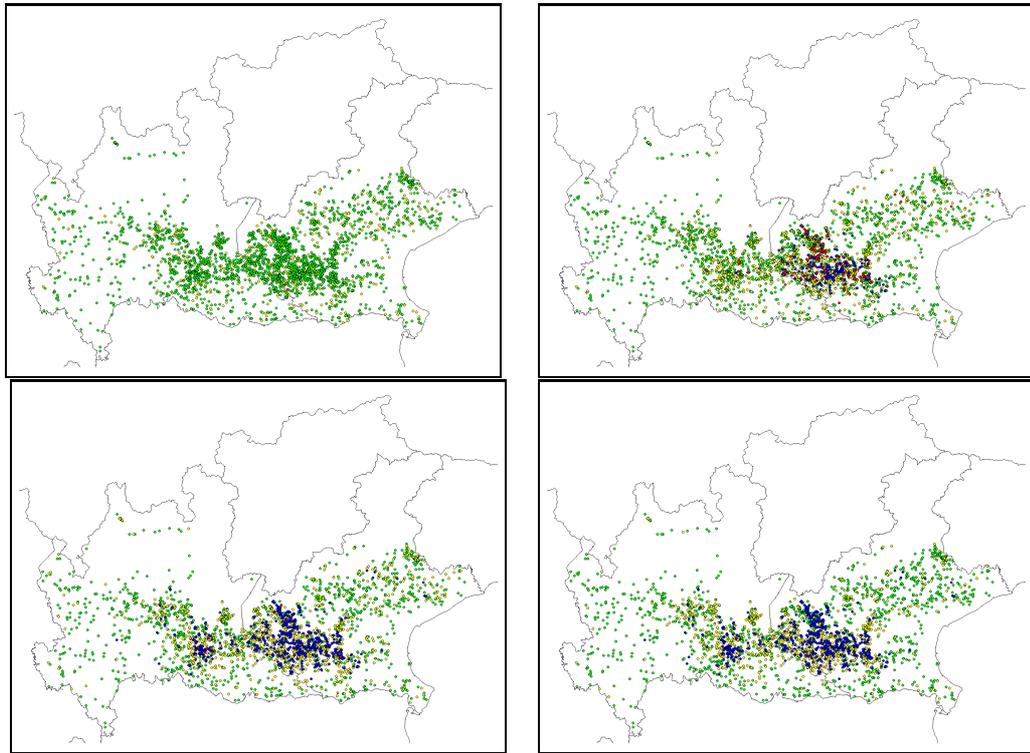


Figure 3.4: Status of farms in the study area at time $t = 1$ (top left), $t = 50$ (top right), $t = 100$ (bottom left), $t = 150$ (bottom right) in one simulation of the 2-Phases Susceptibility Model. Yellow dots represent empty farms, green dots represent susceptible units, red dots represent infectious units, blue dots represent (either pre-emptive or previously infected) culled farms.

We also explored the effect of applying PEC with some variations. An earlier (i.e. since time $t = 20$) application of pre-emptive culling (earlier-PEC) on the whole study area produces on average less infected cases and a higher number of pre-emptive culled farms, for a total number of losses which is slightly lower than those produced by the basic scenario (see Table 3.9).

The application of pre-emptive culling for farms within a radius of 0.5 km from an IF (restricted-PEC) produces on average a higher number of infected farms and a lower number of pre-emptive culled farms, for a smaller number of total losses in comparison to the base scenario (see Table 3.8). Note that the NO-PEC strategy produces on average an even lower number of total losses (equal to IF) (see Table 3.8), which is currently the lowest among the simulated strategies, and could then be considered to be the best one from this point of view. On the other hand, the time required to eradicate the disease in the NO-PEC scenario would be longer (about 10% on average) than with the baseline scenario. From the point of view

3.6. Results and Discussion

Table 3.9: Mean numbers and 5 – 95 percentile intervals computed on 100 realizations that generated at least 10 cases, using the 2-Phases Susceptibility Model with different intervention strategies

	earlier-PEC	restricted-PEC	earlier&restricted PEC
IF	222.05(98, 326)	409.33(208, 578)	334.66(174, 450)
PEC	269.68(127, 379)	58.78(32, 83)	138.34 (83, 192)
BR	1349.68(1074, 1505)	1321.70(1023, 1512)	1356.46(1080, 1522)
T_{ext}	125.11(96,163)	142.84(110, 199)	133.91(107, 169)
D_{max}	132.82(94.79,183.24)	150.79(115.93, 197.43)	146.23 (105.99, 185.14)
TL	491.73	468.11	473

of eradication time, the earlier-PEC strategy would have been the best one.

Note finally that the variation among simulations is rather high compared to the differences among strategies. The only strategy that produces results unequivocally different from the other ones is the NO-BR.

3.6 Results and Discussion

Our study confirms that proximity to an IF increases the risk of infection. This supports our choice to take transmission kernels as power law functions of the distance; moreover the exponent α and scale r_0 are rather similar to what was found by (Boender et al., 2007), despite the different context.

Reduction of virus transmissibility between Phase 1 and the subsequent phases and difference in susceptibility by species have been also observed in this analysis. Our estimates suggest a great difference in exposure and/or susceptibility among the poultry species. Since the model does not distinguish between differential levels of exposure and intrinsic susceptibility, the estimates show that laying hens and meat turkeys are most exposed and/or susceptible to H7N1 virus. Breeders seem to be less exposed and/or susceptible to H7N1 than laying hens and meat turkeys but more exposed and/or susceptible than broilers and all other species together (Table 3.2). These results are consistent with the cumulative probability of infection computed by Busani et al. (2009) on the same datasets.

We have also examined a model with differences in infectivity among species. The results (not shown) are on the border of significance for differences in infectivity. However, the fit to data and the agreement of simulations with observed data were much worse than in the model that accounts for the difference in susceptibility. Overall, we believe that the data cannot demonstrate with good confidence the

existence of differences in infectivity among species.

The 2-Phases Susceptibility Model turned out to be the model, among those tested here, whose simulated outputs (Tables 3.6 and 3.7) are most similar to the observed data. Figures 3.2 and 3.3 show that the observed epidemic falls within the range of the predictions obtained by the 2-Phases Susceptibility Model both in terms of number of new cases at each time t (Figure 3.2) and of the general profile of the epidemic curve over time (Figure 3.3). This supports the utility of the model as an adequate and useful tool for policy testing.

The results show that control measures such as culling of infectious farms, preemptive culling of contiguous premises, ban of restocking on emptied farms and restrictions to the movement of animals, vehicles and staff (i.e. decrease of the number of contacts among farms) have effectively reduced virus transmission over time, as observed also by Le Menach et al. (2006) for the epidemic in The Netherlands.

The BR resulted in the most effective intervention measure to control and eradicate the epidemic. Simulations without BR measure but applying only culling of infected farms and neighboring premises resulted in a larger number of infected or culled farms than the observed number. Simulations without BR did not take into account of the “empty period” between successive production cycles (i.e. every farm is assumed to be in production during the whole epidemic). For this reason the effect of BR may have been overestimated. On the other hand, the overestimation was presumably small, since only some of the farms would have not been in production during the study period and for only a few days.

The strategy that minimizes the total losses is the NO-PEC. However, its implementation would delay the eradication of the infection. As a consequence, the affected area would be submitted to the restriction measures longer, causing additional economic losses. Establishing the overall best strategy would entail an economic analysis beyond our aims. Comparative studies of the outcomes of alternative control strategies have been published for different disease outbreaks (Keeling et al., 2001; Henzler et al., 2003; Bouma et al., 2003; Matthews et al., 2003; Stegeman, 2004; Tildesley et al., 2009).

Earlier-PEC strategy achieves eradication more quickly than what observed in the actual schedule and has also smaller costs in terms of losses. Also, NO-PEC and restricted-PEC lower the total losses but at the cost of delaying eradication. Indeed, the data show that the actual policy has been a sort of restricted PEC (especially in Lombardia) because of the difference between the expected (129 farms, according

3.6. Results and Discussion

to the official policy) and observed (72 farms) pre-emptive cullings.

As a final observation, it can be seen that the real epidemic spread farther than most simulations thus suggesting a role of the long-range transmission, mainly related to human activities and poultry farming practices (movement of personnel, trucks, animals and birds in the infected area). Indeed, as shown by Figure 3.5, the farthest infected cases acquired infection relatively early in time. In our study

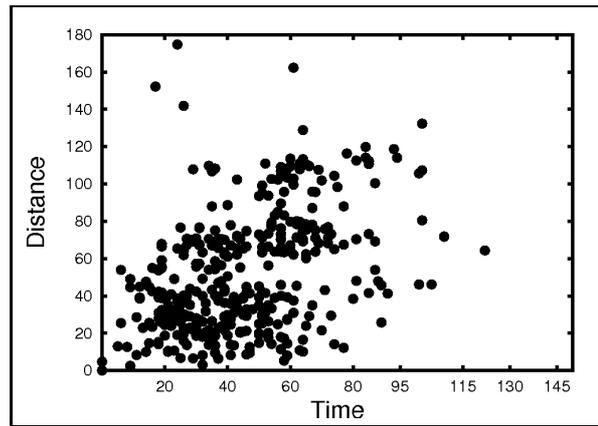


Figure 3.5: Distance (km) reached by infection in time (days) in the 1999-2000 epidemic in Italy

we have taken into account the distance between an IF and an uninfected farm, without any clue on the way of the HPAI viruses spreading. AI spreads mainly through direct or indirect contact with infected birds (flow of people, movement of materials and vehicles for instance) (Halvorson and Karunakaran, 1980; Thomas et al., 2005; Busani et al., 2009) but aerosol transmission, “contiguous spread” by poorly understood routes (Henzler et al., 2003; Sedlmaier et al., 2009) and inter-species transmission via pigs (Webster et al., 1992; Ninomiya et al., 2002) cannot be excluded.

More detailed data about the occurrence of at risk contacts between infected and uninfected poultry farms related to the movements of birds, people and vehicles would be necessary to include long-range transmission in the model.

Chapter 4

A new approach to estimate the spread and transmission of infectious diseases from Sentinel surveillance: application to the 2009-2010 A/H1N1 influenza pandemic in Italy

4.1 Introduction

The detection and control of existing, newly emerging or re-emerging infections in the human population often relies on the analysis of syndromic and virological surveillance data which are routinely collected by most developed and many developing countries. Surveillance data are often the only kind of data available in real time to inform decision makers and the analysis of these data provides important insights into the spread and transmission dynamics of diseases like influenza. During the 2009-2010 A/H1N1 influenza pandemic, syndromic and virological surveillance data were routinely collected by most of the countries affected by H1N1 and available in real time.

The analysis of syndromic and virological data poses many statistical challenges that have not been addressed yet. For example, the size of the population that is

monitored changes over time; only a fraction of syndromic cases who are detected by the surveillance system have been infected by the etiological agent of interest (e.g. H1N1 virus, in the past 2009-2010 influenza pandemic) and the others are due to other pathogens. These problems are usually either ignored or corrected by scaling the epidemic curve with multiplicative factors, something which is expected to bias the variance of the estimates.

Here we present a general framework to tackle these issues and analyze syndromic and virological data by taking explicitly into account the stochasticity in the surveillance system. This is done by coupling a deterministic mathematical ODE (ordinary differential equations) model with a statistical description of how the surveillance data is generated. Estimation of epidemiological parameters such as the reproduction number R_0 and the age-dependent reporting rates and susceptibility is then performed via Bayesian Markov Chain Monte Carlo (MCMC) sampling. The approach is applied to surveillance data collected in Italy during the 2009-2010 A/H1N1 influenza pandemic.

The general modelling framework proposed in this work can be applied to a variety of different infections detected by surveillance system in many countries and is potentially a powerful tool to be used in the future to provide policy makers with important information in real time.

4.2 Data

Since the 1999-2000 influenza season, the Italian influenza surveillance system relies on INFLUNET. During the 2009-2010 H1N1 pandemic influenza season, INFLUNET recruited on average 1094 (minimum 980, maximum 1165) volunteer GPs and paediatricians per week, covering on average 1.4 million people (2.3% of the Italian population). Data collected by INFLUNET on the weekly size of the monitored patients population and on the weekly number of observed Influenza-Like-Illness (ILI) cases, aggregated by age groups (0-4 years, 5-14 years, 15-64 years and 65+ years) are available online on the INFLUNET website (<http://www.iss.it/iflu/>). The virological surveillance of the 2009-2010 influenza season has been conducted by the Italian Ministry of Health, which coordinated the collection of the swabs through hospitals, laboratories operating within the national health service, sentinel GPs and paediatricians. Weekly reports are available online on the Italian Ministry of Health website under the voice “sorveglianza virologica” (<http://www.salute.gov.it/influenza/influenza.jsp>).

There is evidence that the number and structure of the contacts within an age-structured population significantly vary over time, in particular between holiday/week-end days and working days (Hens et al., 2009b,a). For this reason, using raw data from the Italian arm of the POLYMOD survey (a diary-based survey of daily contacts in eight European countries) (Mossong et al., 2008), we compute the daily mean number of contacts among the considered age classes during working days and holiday/week-end days. In the Supplementary Information (SI) we briefly discuss the methodology used to obtain the contact matrices used in this work. Finally, we use Italian demographic data for year 2008 which can be found on the Italian National Statistical Institute website (<http://www.istat.it/>).

We analyze here the data for the time period between week 38 of year 2009 (corresponding to mid September 2009, when the schools re-opened after the summer break) and week 7 of year 2010 (corresponding to the end of February, when the epidemic had clearly died out).

4.3 Model Formulation

4.3.1 Mathematical model

We defined an age-structured deterministic SEIR model, where individuals are successively Susceptible, Exposed (def), Infectious (def) and Removed (def), with five age classes (0-4, 5-14, 15-24, 25-64, 65+ years). The latent period (that is, the duration of stay in the Exposed state) and the infectious period are assumed to be Gamma distributed (this is achieved by splitting the Exposed and Infectious states in 2 compartments, each). The addition of one age-class to those considered by INFLUNET during the 2009 – 2010 influenza season is meant to allow a better specification of the contacts among younger age-classes which were particularly hit by H1N1 virus.

The model is coded in C and is numerically solved using standard routines with variable step size (Press et al., 2002). From the model we output C_i^t , the weekly (t) and age-specific ($i = 1, \dots, 5$) number of A/H1N1 infections in the Italian population and, by scaling down to the size of the monitored patients population, we get \bar{Z}_i^t , the expected number of A/H1N1 infections generated within class i during week t in the monitored patients population.

4.3.2 Statistical model

In what follows we adopt the notation graphically represented on Figure 4.1 for the purpose of clarity. Except for the variable C_t^i , which represents the age-structured weekly number of A/H1N1 cases in the Italian population, all the other variables are defined at the monitored patients population level. In particular, since no information on the patients age is provided for the samples tested in the virological analysis, we assume that π_t , defined as the probability that a swabs tests positive on week t , does not vary across the age-groups.

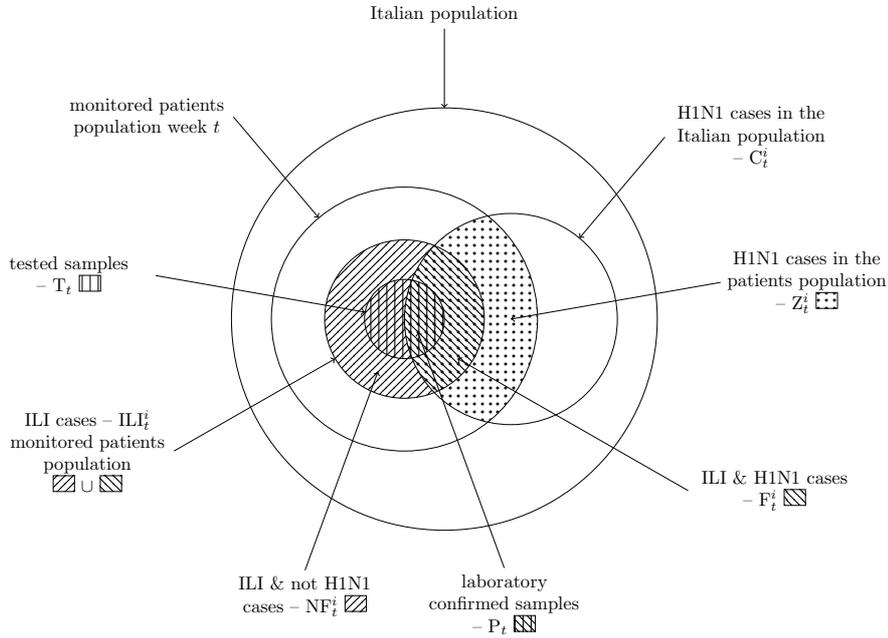


Figure 4.1: Graphical representation of the populations taken into account and notation adopted in the work. The Italian population is considered constant over the whole study period while the monitored patients population changes every week, due to the voluntary nature of the surveillance system. Index i denotes the age-class ($i = 1, \dots, 4$) and index t denotes the week, ranging from week 38 of year 2009 to week 7 of year 2010.

In the following we describe the assumptions that led us to the definition of the likelihood function L .

We divide the presentation in two parts. We first consider the case when the weekly number of H1N1 cases in the monitored patients population in the i -th age class Z_t^i is exactly predicted by the solution \bar{Z}_t^i of the deterministic system. Then, we extend our model to the situation when Z_t^i is a random variable of which \bar{Z}_t^i is the expected value.

4.3.3 Fixed Z_t^i

Let us start by considering the case $Z_t^i = \bar{Z}_t^i$.

Denote T_t the weekly (and age-unstructured) number of swabs sampled (within the monitored patients population and among the individuals showing ILI symptoms) to be laboratory tested (Figure 4.1). Denote P_t the laboratory confirmed H1N1 samples among those tested (T_t) on the corresponding week (Figure 4.1). Since we lack information about the precise timing of collection of the swabs, we assume that samples tested on week t had been collected during week $t - 1$.

Given π_t and T_t , the number of positive swabs P_t follows the Binomial distribution

$$P(P_t|T_t, \pi_t) = \binom{T_t}{P_t} \pi_t^{P_t} (1 - \pi_t)^{T_t - P_t} \quad (4.1)$$

Denote ILL_t^i the weekly number of ILI cases in the monitored patients population of age-class i and F_t^i the weekly number of H1N1 cases of age-class i that report ILI symptoms (Figure 4.1). Hence, if ρ_i represents the probability that a person infected with H1N1 reports ILI symptoms, the distribution of F_t^i is given by the Binomial model with parameters \bar{Z}_t^i and ρ_i

$$P(F_t^i|\bar{Z}_t^i, \rho_i) = \binom{\bar{Z}_t^i}{F_t^i} \rho_i^{F_t^i} (1 - \rho_i)^{\bar{Z}_t^i - F_t^i} \quad (4.2)$$

Denote NF_t^i the number of ILI cases (in the monitored patients population) that would result negative to the A/H1N1 virus, if tested

$$NF_t^i = ILL_t^i - F_t^i \quad (4.3)$$

Given F_t^i , we think of NF_t^i as the number of negative samples one gets in a sequence of Bernoulli trials before obtaining the F_t^i -th positive sample. Hence, given $F_t^i > 0$ and π_t , we assume that NF_t^i has a negative binomial distribution with parameters F_t^i and $1 - \pi_t$

$$P(NF_t^i|F_t^i, 1 - \pi_t) = \binom{NF_t^i + F_t^i - 1}{F_t^i - 1} \pi_t^{F_t^i} (1 - \pi_t)^{NF_t^i} \quad (4.4)$$

Equation 4.3 implies that

$$P(ILL_t^i|F_t^i, \pi_t) = P(NF_t^i = ILL_t^i - F_t^i|F_t^i, \pi_t)$$

4.3. Model Formulation

and the probability distribution of $ILLI_t^i$ is explicitly given by

$$P(ILLI_t^i | F_t^i, \pi_t) = \binom{ILLI_t^i - 1}{F_t^i - 1} \pi_t^{F_t^i} (1 - \pi_t)^{ILLI_t^i - F_t^i} \quad (4.5)$$

If $F_t^i = 0$, then the whole $ILLI_t^i$ set would test negative, that is

$$P(ILLI_t^i | F_t^i, \pi_t) = (1 - \pi_t)^{ILLI_t^i} \quad (4.6)$$

Once given a prior distribution $P(\pi_t)$ to π_t , using conditional probability and assumptions (4.4), (4.5) and (4.6), we define the (up to a normalising constant) probability of the data given the model

$$\begin{aligned} P(ILLI_t^i, P_t | T_t, \bar{Z}_t^i, \rho_i) &= \\ &= \sum_{j=0}^{\min(ILLI_t^i, \bar{Z}_t^i)} \int_0^1 P(ILLI_t^i | F_t^i = j, \pi_t) P(F_t^i = j | \bar{Z}_t^i, \rho_i) P(P_t | T_t, \pi_t) P(\pi_t) d\pi_t \end{aligned} \quad (4.7)$$

We assume a prior Beta distribution for π_t ,

$$P(\pi_t) = \frac{\pi_t^{\alpha-1} (1 - \pi_t)^{\beta-1}}{B(\alpha, \beta)} \quad (4.8)$$

where α and β are shape parameters, substitute (4.1), (4.2), (4.5), (4.6) and (4.8) into (4.7), and obtain (see the SI for the complete computation)

$$\begin{aligned} P(ILLI_t^i, P_t | T_t, \bar{Z}_t^i, \rho_i) &= \\ &= \frac{\binom{T_t}{P_t}}{B(a, b)} \left((1 - \rho_i)^{\bar{Z}_t^i} B(P_t + \alpha, ILLI_t^i + T_t - P_t + \beta) + \right. \\ &+ \left. \sum_{F_t^i=1}^{\min(ILLI_t^i, \bar{Z}_t^i)} \binom{ILLI_t^i - 1}{F_t^i - 1} \binom{\bar{Z}_t^i}{F_t^i} \rho_i^{F_t^i} (1 - \rho_i)^{\bar{Z}_t^i - F_t^i} B(F_t^i + P_t + \alpha, ILLI_t^i - F_t^i + T_t - P_t + \beta) \right) \end{aligned} \quad (4.9)$$

Denoting θ the parameter vector, the Bayesian model is defined by:

$$P(\{ILLI_t^i\}_{i,t}, \{P_t\}_t, \theta | T_t) = \prod_t \prod_i P(ILLI_t^i, P_t | T_t, \bar{Z}_t^i(\theta), \rho_i) P(\theta) \quad (4.10)$$

where $P(\theta)$ is the prior distribution.

4.3.4 Random Z_t^i

Instead of taking Z_t^i fixed to the value \bar{Z}_t^i , we assume that Z_t^i is drawn from a Negative Binomial distribution (Alexander et al., 2000; Lloyd-Smith et al., 2005; Lloyd-Smith, 2007; Mathews et al., 2007; Cauchemez and Ferguson, 2008)

$$Z_t^i \sim \text{NegBin}(r, \frac{\bar{Z}_t^i}{\bar{Z}_t^i + r}) \quad (4.11)$$

with (dispersion) parameter r to be defined. Decreasing values of r correspond to increasing levels of overdispersion. In this formulation, the expected value is fixed at \bar{Z}_t^i and the variance is given by $\bar{Z}_t^i \left(1 + \frac{\bar{Z}_t^i}{r}\right)$.

Under this assumption, it can be proved (see the SI) that

$$P(F_t^i | \bar{Z}_t^i, \rho_i, r) = \binom{F_t^i + r - 1}{r - 1} \left(\frac{\bar{Z}_t^i \rho_i}{\bar{Z}_t^i \rho_i + r}\right)^{F_t^i} \left(\frac{r}{\bar{Z}_t^i \rho_i + r}\right)^r \quad (4.12)$$

The (up to a normalising constant) probability of the data given the model is in this case given by

$$\begin{aligned} P(ILL_t^i, P_t | T_t, \bar{Z}_t^i, \rho_i, r) &= \\ &= \frac{\binom{T_t}{P_t}}{B(a, b)} \left((1 - q_t^i)^{\bar{Z}_t^i} B(P_t + \alpha, ILL_t^i + T_t - P_t + \beta) + \right. \\ &+ \left. \sum_{F_t^i=1}^{ILL_t^i} \binom{ILL_t^i - 1}{F_t^i - 1} \binom{F_t^i + r - 1}{r - 1} (q_t^i)^{F_t^i} (1 - q_t^i)^r B(F_t^i + P_t + \alpha, ILL_t^i - F_t^i + T_t - P_t + \beta) \right) \end{aligned} \quad (4.13)$$

where for simplicity of notation we set $q_t^i = \frac{\bar{Z}_t^i \rho_i}{\bar{Z}_t^i \rho_i + r}$.

Expression (4.13) has been obtained substituting (4.1), (4.12), (4.5), (4.6) and (4.8)

into formula

$$\begin{aligned}
 &P(ILL_t^i, P_t|T_t, \bar{Z}_t^i, \rho_i, r) = \\
 &= \sum_{j=0}^{ILL_t^i} \int_0^1 P(ILL_t^i|F_t^i = j, \pi_t)P(F_t^i = j|\bar{Z}_t^i, \rho_i, r)P(P_t|T_t, \pi_t)P(\pi_t)d\pi_t \quad (4.14)
 \end{aligned}$$

If we denote by θ the parameter vector, the Bayesian model is defined by (4.10) with (4.14) in place of (4.9).

4.4 Models definition and parametrisation

In the previous section we have explicitly defined two families of models, depending on the assumption on Z_t^i either exactly predicted by the deterministic model through the solution \bar{Z}_t^i or taken as a negative binomial random variable with expected value given by \bar{Z}_t^i . The first case will be referred as the “without overdispersion” variant of the model, the second as the “with overdispersion” one.

From early on in the 2009 pandemic, it was noticed that the young age-classes were particularly hit by the H1N1 virus (Fraser et al., 2009; Ghani et al., 2009). In order to quantify this observation, we use here the results from the cross-sectional serological study led by Miller et al. (2010) on serum samples collected in 2008 in England.

Details on how we used the results of the serological study by Miller et al. (2010) and on alternative assumptions and defined models are given in the SI. Table 4.1 summarizes the values of susceptibility we used in the “Susceptibility” model.

We first assume that, during the 2009-2010 H1N1 pandemic, the reporting rates were constant over time. In the “Basic” variant of the model we assume that the reporting rates did not vary across the age-groups (i.e. $\rho_1 = \dots = \rho_5$) whereas in the Age-Dependent Reporting (ADR) version we allow reporting rates to be age-specific. In the Time-Varying Reporting (TVR) version of the model we assume that the age-dependent reporting rate of each age-class changes over time t (weeks) proportionally among the age-classes as given by the piecewise linear function

$$\rho_i(t) = \rho_i g(t) \quad (4.15)$$

where $g(38) = 1$, $g(45) = a$, $g(52) = b$, g is linear on the whole domain and a , b and ρ_i with $i = 1, \dots, 4$ are parameters to be estimated.

Estimates of the infectivity h^1 and h^2 of the respective infectious stages I^1 and I^2 have been obtained through the fit of the infectivity function (after infection) of a SEIR model to the average of the daily titres collected from six volunteers who were experimentally infected with an H1N1 influenza virus, as described by Baccam et al. (2006). The values used for h^1 and h^2 are reported on Table 4.1 and a more extensive description of the methodology adopted for this estimation is given in the SI.

In agreement with some recent studies about H1N1 influenza (Cauchemez et al., 2009a; Ghani et al., 2009; Lessler et al., 2009), we fix the mean generation time T_g to 2.6 days and the mean latency period to 1 day as in (Baguelin et al., 2010).

In order to allow for a proper mixing, we seed the initial number of A/H1N1 cases I_0 (in the Italian population) on week 31 (mid August 2009) and fit the model to the data on the temporal window between week 38 of year 2009 and week 7 of 2010. The initial number of cases I_0 is distributed among the age classes proportionally to the vector (5%, 10%, 45%, 35%, 5%) which appears reasonable and comparable to the age distribution of reported cases over the summer (Rizzo et al., 2009). Sensitivity analysis on this assumption has been performed.

In Italy schools re-opened, after the summer break, on September 15th 2009. For this reason, until week 38, we assign holidays/week-end contacts to school-aged children (5 – 14 years). The same is done for Christmas holidays (December 23rd 2009-January 7th 2010) during which the other classes are assumed to have the average between week and holiday/weekend contacts.

4.5. Parameter estimation

	meaning of the parameter		model
η	latency rate	2.0/day	Basic, ADR
γ	infectious rate	0.833/day	Basic, ADR
σ_1	susceptibility of age-class 0 – 4 years	0.98	Basic, ADR, TVR
σ_2	susceptibility of age-class 5 – 14 years	0.96	Basic, ADR, TVR
σ_3	susceptibility of age-class 15 – 24 years	0.85	Basic, ADR, TVR
σ_4	susceptibility of age-class 25 – 64 years	0.87	Basic, ADR, TVR
σ_5	susceptibility of age-class 65+ years	0.73	Basic, ADR, TVR
h^1	infectivity of the infectious stage I^1	16.1	Basic, ADR, TVR
h^2	infectivity of the infectious stage I^2	9.6	Basic, ADR, TVR
α, β	shape parameters of the Beta distribution in (4.8)	1.0	Basic, ADR, TVR
p	probability of infection given an infectious contact	ind.comp.	Basic, ADR, TVR
R_0	reproduction number	estimated	Basic, ADR, TVR
I_0	number of H1N1 cases at week 31	estimated	Basic, ADR, TVR
ρ_1	ILI reporting rate of H1N1 cases of age 0 – 4	estimated	ADR, TVR
ρ_2	ILI reporting rate of H1N1 cases of age 5 – 24	estimated	ADR, TVR
ρ_3	ILI reporting rate of H1N1 cases of age 25 – 64	estimated	ADR, TVR
ρ_4	ILI reporting rate of H1N1 cases of age 65+	estimated	ADR, TVR
a, b	parameters defining the function in (4.15)	estimated	TVR

Table 4.1: Summary of the parameter values fixed and estimated in the models. With the expression “ind.comp.” we mean “indirectly computed” from R_0 , as explained in the main text.

4.5 Parameter estimation

In a Bayesian setting, we make inference on the parameters which are summarized in Table 4.1.

Given the likelihood function L and chosen a (in our case uniform) prior distribution of the parameters, the (target) posterior distribution is known up to a normalizing constant. MCMC methods construct Markov chains whose stationary distribution is the distribution of interest, when it cannot be directly simulated. We implemented the classical Metropolis-Hastings algorithm (Gilks et al., 1996; Tierney, 1994; Walsh, 2004; O’Neill, 2002) and, starting from arbitrary initial values in the parameter space, generated sequences of draws from the unknown (target) probability distribution of the parameters. We assume a flat prior distribution for π_t , thus setting the shape parameters α and β of (4.8) equal to 1. A log-scale has been used for sampling as the parameters were all positive definite and were ex-

pected to potentially vary by orders of magnitude. Parameters have been updated either separately (i.e. component by component) in the low dimensionality models or in blocks of 2 – 3 parameters each for the models with higher dimensionality, in order to improve the algorithm performance. We checked convergence by assigning different starting values in the parameter space (also far from the posterior mean) and by visual inspection of the trace plots. The algorithm was iterated for 500.000 times and we fixed a “burn-in” period of 100.000 steps. By tuning the variance of the proposal distribution, we adjusted the mixing of the chains and attempted to reach a rate of acceptance (number of accepted moves/number of proposed points) as closest as possible to the “golden” acceptance rate for the Random Walk Metropolis Hastings of 23% (Roberts et al., 1997). As expected, we found some correlations between certain parameters (like R_0 and I_0 , for example). We use the Deviance Information Criterion (DIC) for model comparison and selection (the preferred model is the one showing the lowest DIC) (Spiegelhalter et al., 2002).

4.6 Results

The ILI incidence curve peaked on week 46 (mid November), decreased over the next 6 weeks and then slowly increased again during the first weeks of 2010 (see Figure 4.2). The H1N1-attributable ILI-incidence curve (red dots) in Figure 4.2 has been simply obtained by multiplying the ILI incidence times the proportion of positive swabs collected in that week, under the assumption that the samples tested on week t had been collected during week $t - 1$.

Table 4.2 reports the mean and the equal-tailed 95% credible interval of the estimated parameters for the “Susceptibility” model “without overdispersion”. The estimated mean value of R_0 ranges from 1.36 to 1.42, respectively obtained by the “Basic” and “Age-Dependent Reporting” versions of the “Susceptibility” model “without overdispersion”.

Table 4.3 summarizes the estimates obtained by the “with overdispersion” variant of the “Age-Dependent Reporting Susceptibility model” with overdispersion parameter r estimated from the data; according to these estimates the estimated mean value of R_0 has been 1.29(1.27 – 1.32). The “Age-Dependent Reporting Susceptibility model” with overdispersion shows the lowest DIC among the models considered in this work and using this model we estimate that, on average, in the Italian population the 25.9% of H1N1 cases of 0-4 years, the 16.6% of H1N1 cases

4.6. Results

of 5-14 years, the 6.9% of H1N1 cases of 15-64 years and the 6.5% of H1N1 cases of 65+ years reported ILI symptoms to the surveillance system.

The different models exhibit different credibility interval ranges, which are reflected into the differently wide prediction bars of Figure 4.3. The “Age-Dependent Reporting Susceptibility” model with overdispersion is the one with the widest credibility interval range.

Figure 4.4 shows the age-specific estimated incidences (per thousand) of H1N1 cases within the Italian population obtained from the numerical solution of the SEIR model (in the Italian population) having fixed the parameters as obtained from each of 500 random draws from the joint posterior distribution estimated with the “Age-Dependent Reporting Susceptibility” model with estimated dispersion parameter r . Table 4.4 summarizes some statistics of the predictions plotted on Figure 4.4. The estimated peak-incidences of A/H1N1 cases show a fair variability both within and between the age-classes. At the community level the estimated peak-incidence is of 55.7 (30.8, 91.6) (per thousand). On Table 4.4 we also report the estimated age-specific and overall case attack rates, computed on the whole study period (weeks 31 – 7). In terms of A/H1N1 case attack rate, we estimate that the 5 – 14 years age-class was about 5 times more affected than the 65+ years age-group and that the overall attack rate was of 29.6% (27.7%, 31.6%).

Sensitivity analysis (see the SI) shows that the particular seeding does not affect the model output and that the estimates are also robust to the hypothesis on the length of the latent period.

	Basic Model	ADR Model	TVR Model
DIC	10104.5	2510.8	2468.4
R_0	1.362 (1.357, 1.368)	1.412 (1.405, 1.418)	1.384 (1.371, 1.398)
I_0	136 (116, 156)	37 (32, 44)	69 (45, 99)
ρ_1	0.084 (0.082, 0.086)	0.188 (0.182, 0.195)	0.191 (0.141, 0.247)
ρ_2	0.084 (0.082, 0.086)	0.175 (0.171, 0.179)	0.171 (0.128, 0.219)
ρ_3	0.084 (0.082, 0.086)	0.055 (0.054, 0.057)	0.055 (0.041, 0.071)
ρ_4	0.084 (0.082, 0.086)	0.035 (0.033, 0.038)	0.036 (0.027, 0.047)
a			1.180 (0.842, 1.657)
b			0.557 (0.377, 0.795)

Table 4.2: Susceptibility model without overdispersion: DIC score, mean and equal-tailed 95% credible interval of the marginal posterior distribution of the parameters for each specified model.

ADR Susceptibility model with overdispersion				
DIC	1234.0	1478.3	1693.9	2127.3
R_0	1.298 (1.275, 1.321)	1.304 (1.286, 1.322)	1.341 (1.332, 1.350)	1.385 (1.379, 1.392)
I_0	900 (476, 1536)	753 (466, 1146)	263 (202, 332)	76 (63, 90)
ρ_1	0.259 (0.205, 0.325)	0.254 (0.211, 0.305)	0.227 (0.211, 0.244)	0.201 (0.193, 0.210)
ρ_2	0.166 (0.132, 0.207)	0.164 (0.138, 0.196)	0.169 (0.159, 0.180)	0.176 (0.170, 0.182)
ρ_3	0.069 (0.056, 0.087)	0.069 (0.057, 0.082)	0.064 (0.059, 0.068)	0.058 (0.056, 0.060)
ρ_4	0.065 (0.050, 0.084)	0.062 (0.050, 0.076)	0.047 (0.043, 0.052)	0.038 (0.036, 0.042)
r	6.309 (3.950, 9.285)	fixed to 10	fixed to 100	fixed to 1000

Table 4.3: ADR Susceptibility model with overdispersion: DIC score, mean and equal-tailed 95% credible interval of the marginal posterior distribution of the parameters having fixed the dispersion parameter r to the specified value and having estimated r from the data.

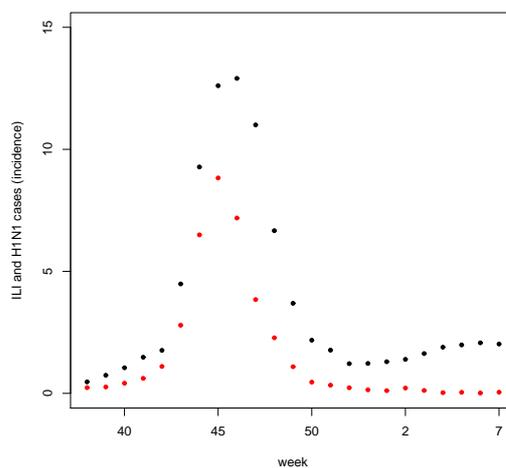


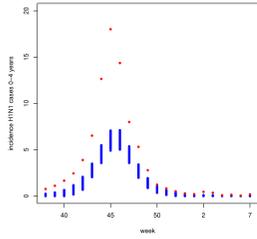
Figure 4.2: Incidence (per thousand) of the total number of reported ILI cases (black dots) and of the number of reported H1N1-attributable ILI-cases (red dots), obtained by multiplication of the weekly ILI datum times the proportion of positive samples on the corresponding week.

4.6. Results

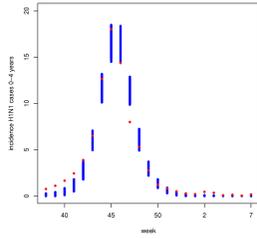
Estimated	peak-incidence	attack rate
0-4 years	53.2 (19.7, 102.5)	31.6% (29.4%, 33.9%)
5-14 years	99.2 (36.7, 192.1)	54.3% (51.6%, 57.1%)
15-64 years	57.1 (21.2, 110.6)	31.5% (29.4%, 33.7%)
65+ years	20.3 (7.5, 39.3)	11.4% (10.6%, 12.4%)
overall	55.7 (30.8, 91.6)	29.6% (27.7%, 31.6%)

Table 4.4: Estimated age-specific peak-incidence (per thousand) and attack rate of H1N1 cases caused by the A/H1N1 virus in the Italian population during the 2009-2010 pandemic as resulted from simulations of the ADR Susceptibility model with estimated overdispersion parameter r having fixed the parameters at the values obtained by 500 draws from the joint estimated posterior distribution. Mean and, in brackets, 5-95 percentile interval.

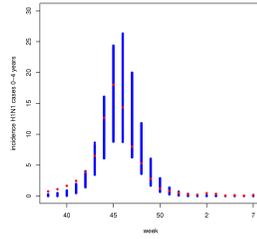
Basic model
without overdispersion



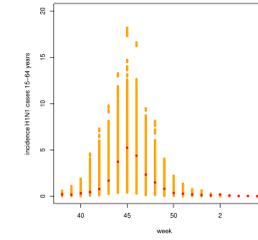
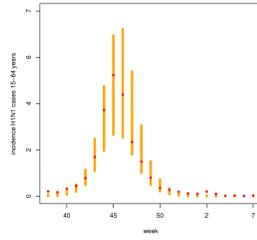
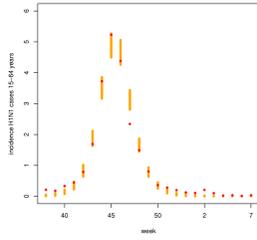
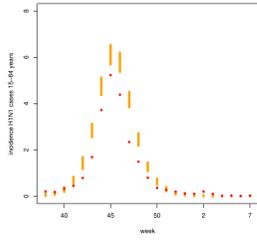
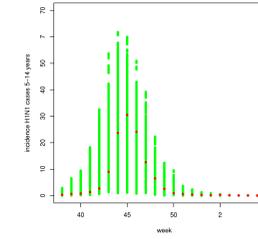
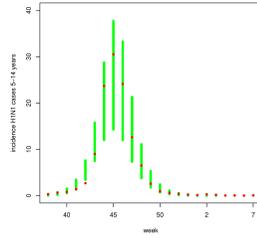
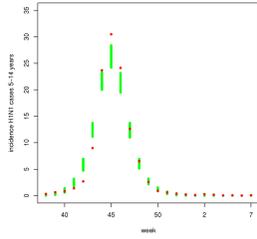
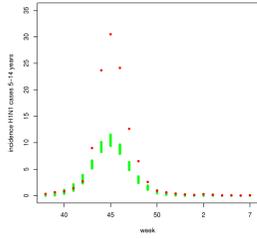
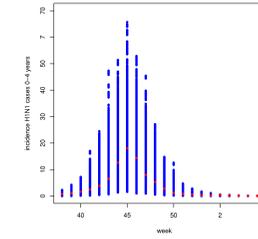
ADR model
without overdispersion



TVR model
without overdispersion



ADR model with
estimated overdispersion



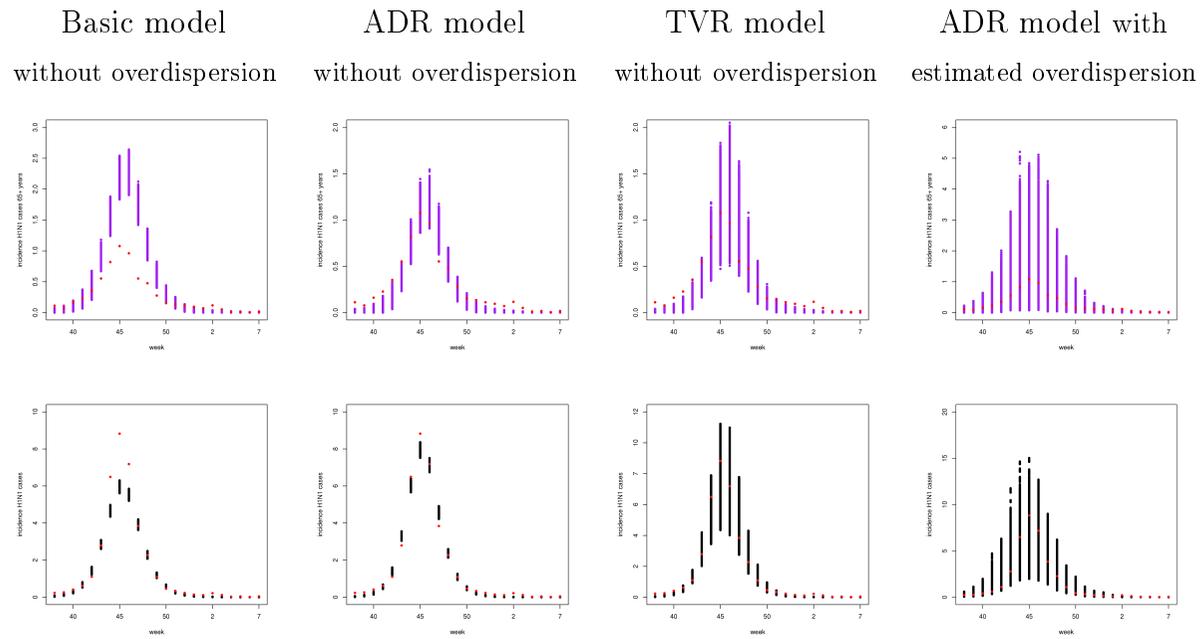


Figure 4.3: Susceptibility model (in the Basic, Age-Dependent Reporting and Time-Varying Reporting versions without overdispersion and in the Age-Dependent Reporting version with overdispersion parameter estimated from the data) : plot of the simulated weekly reported incidence (per thousand) of H1N1 cases in the 0 – 4 years age-class (blue), 5 – 14 years age-class (green), 15 – 64 years age-class (orange), 65+ years age-class (purple) and in the population as a whole (black) in comparison to the respective observed data (red).

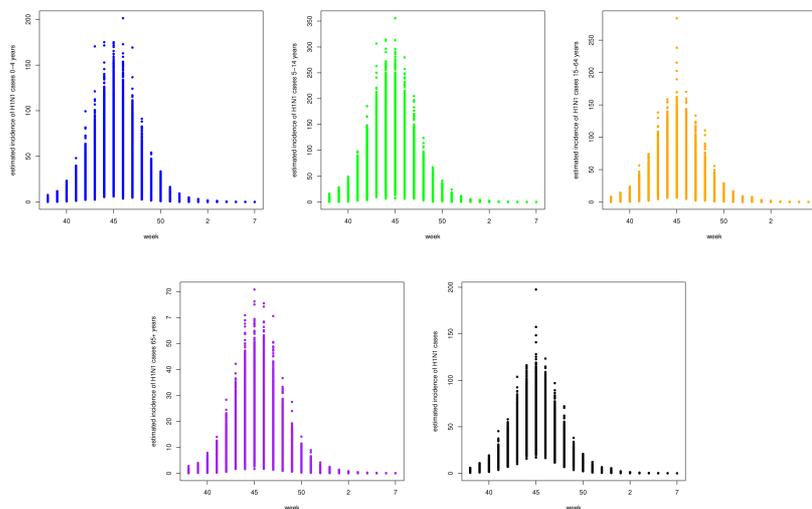


Figure 4.4: Estimated incidence (per thousand) of H1N1 cases in the Italian population using the ADR Susceptibility model with overdispersion parameter r estimated from the data: 0 – 4 years age-class (blue), 5 – 14 years age-class (green), 15 – 64 years age-class (orange), 65+ years age-class (purple), in the population as a whole (black). Predictions have been obtained from the numerical resolution of the SEIR model having fixed the parameters as resulted from 500 draws from the estimated joint posterior distribution.

4.7 Discussion

In this work we propose a general and rigorous statistical framework which explicitly takes into account the way surveillance data are generated. Our main objective was indeed to estimate the incidence of H1N1 cases at the national population level, without scaling the epidemic curve with some multiplicative factor.

We found that, when accounting for an age-specific susceptibility, the mean R_0 varies depending on the model into the range 1.298 – 1.412, where the first estimate has been obtained by the “Age-Dependent Reporting Susceptibility” model with dispersion parameter r estimated from the data and the latter one has been obtained by the “Age-Dependent Reporting Susceptibility” model “without overdispersion”. Our estimates are consistent with those of (Ajelli et al., 2010) derived from the exponential growth phase of the ILI number of cases.

Our finding that different age classes visited their GP with different rates is probably not surprising and the common sense would certainly have suggested this fact. Here we quantified this difference and found that infants and children reported

4.7. Discussion

the H1N1 symptoms and infection respectively around 3 and 4 times more often than adults. This was obtained under the simplifying assumptions that the reporting rates were either constant or changed linearly with time. We tried to test for reporting rates differing in time through the “Time-Varying Reporting Susceptibility” model but the results were not completely satisfying, since despite the relative width of the confidence intervals, a fair number of data points still fell outside the values predicted by the model. However, our estimates suggest a positive correlation of the reporting rate with the number of cases, which is considered a plausible phenomenon by Cauchemez and Ferguson (2008) and may reflect the high level of worry of the people towards the youngest age-classes induced by the media during the past swine flu pandemic, as suggested by Rubin et al. (266). Studies in the spirit of (Rubin et al., 266) able to measure and quantify the behavioural changes occurred during the last H1N1 pandemic in the Italian population are desirable. Table 4.4 and Figure 4.4 confirm that school-age children led and sustained the epidemic, followed by adults and the youngest children, whilst elder people were the less affected.

The introduction of a negative binomial distribution for Z_i^t increases the stochasticity of the models that was otherwise confined to act only in the reporting process. The choice of a negative binomial distribution for the distribution of infections has been suggested by several authors (Mathews et al., 2007; Cauchemez and Ferguson, 2008) and can be considered an approximation to a pure stochastic model. The selected values of r (mean value 6.30 and 95% CI = (3.95, 9.28)) are in the range already used by other authors and result in rather wide credible intervals for the number of infections in any given week.

In the “Susceptibility” model we assumed that at the beginning of the epidemic the whole population is susceptible to H1N1 and assign an age-dependent susceptibility to each age-class. A different assumption consists in assuming that a fraction of the population is immune since the beginning of the epidemic and that susceptibility to H1N1 does not vary among the age-classes (see the SI). The parameter estimates obtained when accounting for an immune fraction of population are consistent with the respective estimates obtained with the “Susceptibility” model so that either choices, to include a differential susceptibility depending on the age or to account for the presence of an immune fraction of the population at the beginning of the epidemic, fit the data equally well.

In the “Age-Dependent Reporting Susceptibility” model we fixed the age-dependent susceptibility to the values extrapolated by the study of Miller et al. (2010). In-

deed, the serological study on the cross-reactive antibody responses to the H1N1 influenza virus in the pre-pandemic period, led in Italy by Rizzo et al. (2010) could not be applied to our model due to incompatible divisions of the population into age-classes (1 – 55 years, 56 – 65 and 65+ years).

In the “Age-Dependent Reporting Susceptibility” model, once fixed the age-dependent susceptibility, we estimated the age-dependent reporting rates. The alternative choice is to fix a constant reporting rate and estimate the age-dependent susceptibility and is discussed in the SI. The results confirm a drop in susceptibility to H1N1 beyond school-years but it’s stronger than our initial assumption. Presumably reality lies in between, with drops both in susceptibility and reporting rate. For sure our analysis would benefit from the availability of more detailed information on the criteria adopted for the collection of the swabs (such as, for instance, the weekly percentages of swabs collected by sentinel GPs, the average delay between collection and testing of the samples etc.) and the eventual changes in the collection process, as the epidemic unfolded. It is indeed important to note that our analysis has been led under the assumption that the swabs selected for testing are a random sample of the ILI cases while a sizeable part has been collected outside the surveillance system (laboratories and hospitals operating within the national health service), presumably for clinical reasons. Unfortunately, available data do not distinguish among swabs of different sources. Finally, our model could be extended to include age-specific virological data, if available.

Our results show that the basic features of the epidemic are captured by the model, in particular the “Age-Dependent Reporting Susceptibility model” describes adequately the overall epidemic course and the age distribution of the cases. There are some minor systematic deviations of the data from the expected values of the predictions (for instance, the predictions regarding the starting weeks of the epidemic are systematically lower than the observed data in the 0-4 years age-class and systematically higher than the observed data in the 5-14 years age-class) so that it could be argued that our model misses some details of the infection and reporting process. It is possible that adding other factors such as changing behavioural patterns causing more complex variations of the reporting rates over time, heterogeneity in infectiousness, spatial and network substructuring for instance could improve the description of the virus spread. Determining which of these elements are needed to accurately describe the dynamics of virus spread in large populations is topic of ongoing research. Still, a simple model like the one we used appears adequate for an overall description of the epidemic course.

The methodology developed here can be applied to the analysis of the temporal spread of the A/H1N1 pandemic influenza in other countries, provided that epidemiological and virological data are available. Finally, our approach could be easily adopted to analyse existing or future emerging infectious diseases.

4.8 Supplementary Information

4.8.1 Data

Methodology adopted to compute the contact matrices

The methodology we adopted to compute the week and week-ends/holiday contact matrices mimics very closely the one used by Mossong et al. (2008). Starting from the raw data of the POLYMOD survey, we computed the equivalent matrices reported in (Mossong et al., 2008) with the distinction between working days and week-end/holiday contacts for Italy. The POLYMOD survey was conducted in Italy between May 17th 2006 and June 1st 2006, a period during which no official holidays occurred. For this reason we are able to distinguish only between the contacts occurring during the working days from those occurring during the week-ends. Since the age distribution of the survey population does not match the Italian population age distribution, we standardize the estimates as follows. First, we divide the total number of contacts had by the participants by the number of participants, thus obtaining the average number of contacts per respondent. Multiplying the average number of contacts per respondent times the size of the correspondent age class in the Italian population, we get the estimated number of contacts in the Italian population (i.e. the average number of contacts of an age class with the other age classes, in the Italian population). We symmetrize the obtained matrix substituting two symmetric off-diagonal elements with their arithmetic mean. After correction for reciprocity, we scale down to the individual level again thus obtaining Tables 4.5 and 4.6, which represent the symmetric contact matrices at the individual level.

	0-4	5-14	15-24	25-64	65+
0-4	5.2258065	1.4971592	0.5942825	11.6781801	1.1112577
5-14	0.7616100	14.4929577	1.7232274	13.1401577	1.0503617
15-24	0.2775376	1.5820011	13.9405941	9.1557241	0.8523744
25-64	0.9948124	2.2004017	1.6700517	11.3832487	1.9459243
65+	0.2618677	0.4865656	0.4300993	5.3830335	2.9318182

Table 4.5: Symmetrized contact matrix of all reported contacts (physical and non-physical) in Italy, consisting of the average number of contact persons recorded per working day per survey participant (Polymod 2008). Row index represents the age class of the participant, column index represents the age class of the contact.

	0-4	5-14	15-24	25-64	65+
0-4	1.6923077	1.4236201	0.2998418	6.6052098	0.7307692
5-14	0.7242004	7.8387097	1.4178535	9.3789261	1.4071014
15-24	0.1400299	1.3016539	10.4090909	10.0278330	0.2500000
25-64	0.5626686	1.5705599	1.8291289	9.0559006	1.7381579
65+	0.1722057	0.6518203	0.1261474	4.8082869	0.5714286

Table 4.6: Symmetrized contact matrix of all reported contacts (physical and non-physical) in Italy, consisting of the average number of contact persons recorded per holiday day per survey participant (Polymod 2008). Row index represents the age class of the participant, column index represents the age class of the contact.

4.8.2 Model formulation

Mathematical model

The equations of the age-structured SEIR model defined in the main text are

$$\left\{ \begin{array}{l} \dot{S}_i = -\lambda_i(t)S_i \\ \dot{E}_i^1 = \lambda_i(t)S_i - \nu E_i^1 \\ \dot{E}_i^2 = \nu(E_i^1 - E_i^2) \\ \dot{I}_i^1 = \nu E_i^2 - \gamma I_i^1 \\ \dot{I}_i^2 = \gamma(I_i^1 - I_i^2) \\ \dot{R}_i = \gamma I_i^2. \end{array} \right. \quad (4.16)$$

with $i, j = 1, \dots, 5$ corresponding to the five age-classes 0-4, 5-14, 15-24, 25-64, 65+ years. The rates of loss of latency ν and infectiousness γ are assumed not to

4.8. Supplementary Information

depend on the age class. The force of infection λ_i is given by

$$\lambda_i(t) = p\sigma_i \sum_{j=1}^5 c_{ij}(t) \left(h^1 \frac{I_j^1(t)}{N_j} + h^2 \frac{I_j^2(t)}{N_j} \right) \quad (4.17)$$

where σ_i represents the susceptibility of age-class i , $c_{ij}(t)$ indicates the mean number of contacts between an individual of age class i with individuals of age class j on day t (the time variable is here used just to distinguish the working days from the week-end days), N_j represents the (constant in time) size of age group j , with $i, j = 1, \dots, 5$, p is for the probability of getting infected upon a contact with an infectious individual and h^1 and h^2 represent the infectivity of the two infectious stages I^1 and I^2 respectively.

The mean number of new cases generated by an individual of age class j in age class i is given by

$$k_{ji} = p\sigma_i c_{ji} \int_0^{+\infty} A(\tau) d\tau \quad i, j = 1, \dots, 5 \quad (4.18)$$

where $A(\tau)$ denotes the infectivity function at time τ after infection. The entries given in (4.18) define the next generation matrix K and following Diekmann and Heesterbeek (2000) we define the reproduction number R_0 as the spectral radius $s(K)$ of the next generation matrix

$$R_0 = s(K) = ps(M) \int_0^{+\infty} A(\tau) d\tau \quad (4.19)$$

The reproduction number R_0 is clearly proportional to p , the probability of infection given an infectious contact. We used R_0 as a parameter and adjusted p accordingly. Matrix M on equation (4.19) is given by

$$m_{ji} = \sigma_i c_{ji}(t) \quad i, j = 1, \dots, 5 \quad (4.20)$$

For completeness, we define the infectivity function $A(\tau)$ at time τ after infection. Let $g(t)$ denote the probability density function of the variable T_E , the length of the latent period (i.e. the time spent in the classes E^1 and E^2). The probability of being in class I^1 at time τ (after infection) is given by

$$P(I^1, \tau) = \int_0^\tau g(t) e^{-\gamma(\tau-t)} dt \quad (4.21)$$

In a similar fashion, the probability of being in class I^2 at time τ (after infection) is given by

$$P(I^2, \tau) = \int_0^\tau g(t) \int_t^\tau \gamma e^{\gamma(u-t)} e^{-\gamma(\tau-u)} du dt \quad (4.22)$$

$A(\tau)$ is defined as follows

$$A(\tau) = h^1 P(I^1, \tau) + h^2 P(I^2, \tau) \quad (4.23)$$

Equation (4.19) needs the computation of

$$\begin{aligned} \int_0^{+\infty} A(\tau) d\tau &= \\ &= h_1 \int_0^{+\infty} \int_0^\tau g(t) e^{-\gamma(\tau-t)} dt d\tau + h_2 \int_0^{+\infty} \int_0^\tau g(t) \int_t^\tau \gamma e^{-\gamma(u-t)} e^{-\gamma(\tau-u)} du dt d\tau \\ &= h_1 \int_0^{+\infty} g(t) \int_t^{+\infty} e^{-\gamma(\tau-t)} d\tau dt + h_2 \int_0^{+\infty} \int_t^{+\infty} g(t) \int_t^\tau \gamma e^{-\gamma(u-t)} e^{-\gamma(\tau-u)} du d\tau dt \\ &= h_1 \int_0^{+\infty} g(t) \int_t^{+\infty} e^{-\gamma(\tau-t)} d\tau dt + h_2 \int_0^{+\infty} g(t) \int_t^{+\infty} \gamma e^{-\gamma(u-t)} \int_u^{+\infty} e^{-\gamma(\tau-u)} d\tau du dt \\ &= \frac{h_1}{\gamma} + \frac{h_2}{\gamma} \end{aligned} \quad (4.24)$$

Hence the basic reproduction number is given by

$$R_0 = ps(M) \frac{h_1 + h_2}{\gamma} \quad (4.25)$$

The mean generation time, defined as the mean duration between time of infection of a secondary infectee and the time of infection of its primary infector (Wallinga and Lipsitch, 2007), is given by

$$T_g = \frac{\int_0^{+\infty} \tau A(\tau) d\tau}{\int_0^{+\infty} A(\tau) d\tau} \quad (4.26)$$

In phase of parameterization, we fixed $\nu = 2.0$ (it corresponds to a latent period of 1.0 day) and tuned γ to obtain a generation time T_g of 2.6 days.

Statistical model

Fixed Z_t^i : computation of $P(ILL_t^i, P_t|T_t, \bar{Z}_t^i, \rho_i)$

We report below the computation of $P(ILL_t^i, P_t|T_t, \bar{Z}_t^i, \rho_i)$ that led to expression (10) in the main text. The computation is based on the mathematical definition of the Beta function B for two variables $x, y > 0$

$$B(x, y) = \int_0^1 t^{x-1}(1-t)^{y-1} dt \quad (4.27)$$

From the definition given in (4.27) it follows that

$$\begin{aligned} P(ILL_t^i, P_t|T_t, \bar{Z}_t^i, \rho_i) &= \\ &= \frac{\binom{T_t}{P_t}}{B(a, b)} \left((1 - \rho_i)^{Z_t^i} \int_0^1 \pi_t^{P_t + \alpha - 1} (1 - \pi_t)^{ILL_t^i + T_t - P_t + \beta - 1} d\pi_t + \right. \\ &+ \sum_{F_t^i=1}^{\min(ILL_t^i, Z_t^i)} \binom{ILL_t^i - 1}{F_t^i - 1} \binom{Z_t^i}{F_t^i} \rho_i^{F_t^i} (1 - \rho_i)^{Z_t^i - F_t^i} \cdot \\ &\quad \int_0^1 \pi_t^{F_t^i + P_t + \alpha - 1} (1 - \pi_t)^{ILL_t^i - F_t^i + T_t - P_t + \beta - 1} d\pi_t + \\ &+ \left. \binom{Z_t^i}{F_t^i} F_t^i \rho_i^{F_t^i} (1 - \rho_i)^{Z_t^i - F_t^i} \int_0^1 \pi_t^{P_t + \alpha - 1} (1 - \pi_t)^{ILL_t^i + T_t - P_t + \beta - 1} d\pi_t \right) \\ &= \frac{\binom{T_t}{P_t}}{B(a, b)} \left((1 - \rho_i)^{\bar{Z}_t^i} B(P_t + \alpha, ILL_t^i + T_t - P_t + \beta) + \right. \\ &+ \left. \sum_{F_t^i=1}^{\min(ILL_t^i, \bar{Z}_t^i)} \binom{ILL_t^i - 1}{F_t^i - 1} \binom{\bar{Z}_t^i}{F_t^i} \rho_i^{F_t^i} (1 - \rho_i)^{\bar{Z}_t^i - F_t^i} B(F_t^i + P_t + \alpha, ILL_t^i - F_t^i + T_t - P_t + \beta) \right) \end{aligned}$$

Random Z_t^i : computation of $P(F_t^i | \bar{Z}_t^i, \rho_i, r)$

In this section we prove that, given

$$Z_t^i \sim \text{NegBin} \left(r, \frac{\bar{Z}_t^i}{\bar{Z}_t^i + r} \right) \quad (4.28)$$

and

$$P(F_t^i | Z_t^i, \rho_i) = \binom{Z_t^i}{F_t^i} \rho_i^{F_t^i} (1 - \rho_i)^{Z_t^i - F_t^i} \quad (4.29)$$

it follows that

$$P(F_t^i | \bar{Z}_t^i, \rho_i, r) = \binom{F_t^i + r - 1}{r - 1} \left(\frac{\bar{Z}_t^i \rho_i}{\bar{Z}_t^i \rho_i + r} \right)^{F_t^i} \left(\frac{r}{\bar{Z}_t^i \rho_i + r} \right)^r \quad (4.30)$$

This fact can be shown using the probability generating function.

For simplicity of notation let's set

$$q_t^i = \frac{\bar{Z}_t^i}{\bar{Z}_t^i + r} \quad (4.31)$$

From (4.11) and (4.2) it follows that the probability generating function of Z_t^i is given by

$$G_{Z_t^i}(t) = \mathbb{E}[t^{Z_t^i}] = \frac{(1 - q_t^i)^r}{(1 - tq_t^i)^r} \quad (4.32)$$

and the probability generating function of F_t^i is given by

$$G_{F_t^i}(t) = \mathbb{E}[t^{F_t^i}] = (1 - \rho_i + t\rho_i)^{Z_t^i} \quad (4.33)$$

where \mathbb{E} denotes the expected value. Since

$$\mathbb{E}[t^{F_t^i}] = \mathbb{E}[\mathbb{E}[t^{F_t^i} | Z_t^i]] = \mathbb{E}[(1 - \rho_i + t\rho_i)^{Z_t^i}]$$

it follows that

$$\begin{aligned} G_{F_t^i}(t) &= \frac{(1 - q_t^i)^r}{[1 - (1 - \rho_i + t\rho_i)q_t^i]^r} \\ &= \frac{(1 - q_t^i)^r}{[1 - q_t^i(1 - \rho_i) - t\rho_i q_t^i]^r} \end{aligned} \quad (4.34)$$

$$= \frac{\left[1 - \frac{\rho_i q_t^i}{1 - q_t^i(1 - \rho_i)}\right]^r}{\left[1 - t \frac{\rho_i q_t^i}{1 - q_t^i(1 - \rho_i)}\right]^r}$$

and hence by (4.31) we may conclude (4.12).

4.8.3 Models definition and parametrization

Alternative models

In this work we define two age-dependent parameters: susceptibility (i.e. the probability of getting infected given a contact with an infectious individual) and reporting rate (i.e. the probability that an H1N1 case in the patients population reports ILI symptoms).

Due to identifiability issues, it is not possible to make inference on both parameters (susceptibility and reporting rate) at the same time and one has to fix one of the two and make inference on the other.

In order to estimate the age-specific reporting rates, we fix the age-specific susceptibility to the values extrapolated from the results of the cross-sectional serological study by Miller et al. (2010) as described below.

Notice first that the presence of cross-reactive antibody in the blood samples can be interpreted as conferring either partial or complete protection to infection by H1N1. In the first case, we assume that at the beginning of the epidemic the whole population is susceptible (i.e. no fraction of the population is immune) and assign an age-specific susceptibility to the different age-classes, thus defining what we call “Susceptibility” model. In the second case, we assume that at the beginning of the epidemic a fraction of the population in each age-class is immune (i.e. removed from the infection dynamics) and that the susceptible population is completely and equally susceptible to H1N1 ($\sigma_1 = \dots = \sigma_5 = 1.0$), thus defining what we call “Immunity” model.

As anticipated, we use the results given in the cross-sectional serological study by Miller et al. (2010) and average the percentages of samples showing microneutralization titre at or above the cut off value of 1:40 and haemagglutination inhibition at or above 1:32 and in the “Susceptibility” model fix the susceptibility of the age classes to the value obtained by subtraction of the obtained average percentage to 1. For example, if 2% is the average percentage of children in the 0-4 years age-class showing titres at or above the specified thresholds, we fix the susceptibility

of the youngest age-class to $\sigma_1 = 1 - 2\% = 98\%$. The values of susceptibility we fixed in the “Susceptibility model” are given in Table 1 of the main text. In the “Immunity model” we fix the fraction of immune population at the beginning of the epidemic to the value obtained by subtraction of the obtained average percentage to 1. Using the same example of above, we assume that the 98% of the population in age-class 0 – 4 years is in the susceptible class at the beginning of the epidemic (week 31). Similarly to what has been done for the “Susceptibility” model, we define a “Basic”, “Age-Dependent Reporting” and “Time-Varying Reporting” versions of the “Immunity” model without overdispersion. The estimates obtained by the “Immunity” model are given on Table 4.7.

	Basic Model	ADR Model	TVR Model
R_0	1.524 (1.518, 1.531)	1.574 (1.567, 1.581)	1.546 (1.531, 1.560)
I_0	119 (101, 138)	37 (32, 44)	66 (45, 95)
ρ_1	0.094 (0.092, 0.096)	0.205 (0.199, 0.212)	0.204 (0.156, 0.257)
ρ_2	0.094 (0.092, 0.096)	0.188 (0.184, 0.193)	0.182 (0.140, 0.227)
ρ_3	0.094 (0.092, 0.096)	0.062 (0.060, 0.64)	0.061 (0.047, 0.077)
ρ_4	0.094 (0.092, 0.096)	0.040 (0.037, 0.043)	0.040 (0.030, 0.051)
a			1.193 (0.878, 1.602)
b			0.607 (0.413, 0.864)

Table 4.7: Immunity model without overdispersion: mean and equal-tailed 95% credible interval of the marginal posterior distribution of the parameters for each specified model.

Note that the estimates of R_0 obtained with the “Immunity” model are higher than those obtained with the relative versions of the “Susceptibility” model. This fact is due to the definition of R_0 given in equation (4.19), which is theoretically adequate only for the “Susceptibility” model (i.e. in the case of the “Immunity” model, formula (4.19) does not account for the presence of an immune fraction of population at the early stages of the epidemic). For a proper comparison, a different definition of R_0 for the “Immunity” model would be needed.

In order to estimate the age-specific susceptibility, we fix the age-dependent reporting rate. Hence, by fixing the reporting rate as resulted from the “Basic” and “Age-Dependent Reporting” version of the “Susceptibility” model without overdispersion, we respectively define a “Basic” and “Age-Dependent Reporting” version of the “Fixed-Reporting” model. The susceptibility estimates obtained with the “Fixed-Reporting” model are given in Table 4.8.

4.8. Supplementary Information

	Basic Fixed-Reporting	rescaling	ADR Fixed-Reporting	rescaling
R_0	1.31 (1.29, 1.33)		1.30 (1.28, 1.32)	
I_0	645 (383, 1013)		651 (390, 1032)	
σ_1	3.02 (2.47, 3.62)	2.60 (2.12, 3.12)	1.20 (0.89, 1.42)	0.96 (0.77, 1.22)
σ_2	1.32 (1.13, 1.52)	1.14 (0.97, 1.31)	1.02 (0.83, 1.24)	0.88 (0.71, 1.07)
σ_3	fixed to 1.0	0.86	fixed to 1.0	0.86
σ_4	0.92 (0.67, 1.25)	0.79 (0.58, 1.08)	0.75 (0.60, 0.95)	0.65 (0.51, 0.82)

Table 4.8: Fixed reporting model with overdispersion ($r = 10$) in the “Basic” and “Age-Dependent Reporting” versions (i.e. having fixed the reporting rates as resulted respectively from the “Basic” and “Age-Dependent Reporting” versions of the “Susceptibility” model): mean and equal-tailed 95% credible interval of the marginal posterior distribution of the parameters; the susceptibility estimated have also been rescaled to the values fixed on Table 1 in the main text for the purpose of comparison.

Estimation of the infectivity parameters h^1 and h^2

The SEIR model used to estimate the infectivity values has one latency class and three infectious stages with a mean latency period of 1 day and a mean infectious period of 3 days. Let ω^1 , ω^2 and ω^3 denote the unknown infectivity parameters of the three infectious stages I^1 , I^2 and I^3 . Through the fit of the infectivity function (after infection) of the SEIIR model to the data reported in (Baccam et al., 2006) we obtain the following estimates

$$\omega^1 = 0.0 \quad \omega^2 = 16.1 \quad \omega^3 = 9.6 \quad (4.35)$$

The infectivity function (after infection) fitted to the data reported in (Baccam et al., 2006) is given in Figure 4.5.

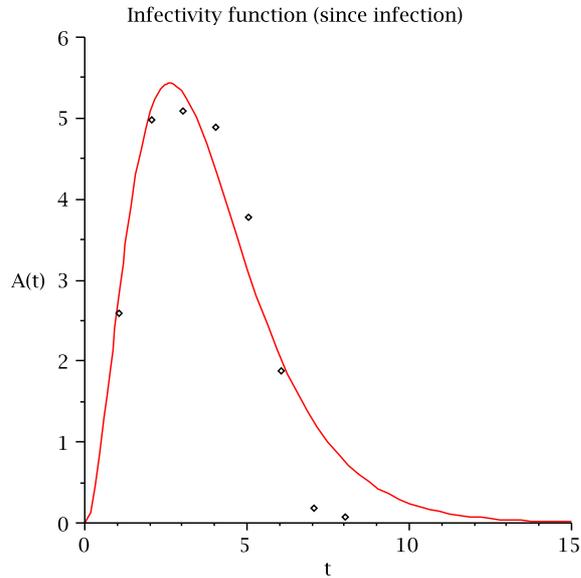


Figure 4.5: Plot of the fit of the infectivity function since infection $A(\tau)$ (defined in the SI by equation (4.23)) to the data reported by Baccam et al. (2006)

Since $\omega^1 = 0.0$ we interpreted the first infectious stage of the SEIIR model as being equivalent to a latent stage, thus obtaining a SEIR model with two latency and two infectious stages; this is the reason for which we took $h^1 = \omega^2$ and $h^2 = \omega^3$. We chose to consider the infectivity estimates as values that characterize the infectious stages, independently of the assumptions on the rates of latency and infectiousness upon which they were generated.

4.8.4 Parameter estimation

Sensitivity analysis

We performed sensitivity analysis on two assumptions: the distribution (at week 31) of the initial cases I_0 among the age-classes and the mean length of the latent period. Sensitivity analysis was performed on the “Age-Dependent Reporting” version of the “Susceptibility” model without overdispersion.

We considered the following distributions of the initial cases I_0 among the age-classes: (1%, 1%, 39%, 39%, 20%), (10%, 20%, 40%, 20%, 10%) and (20%, 20%, 20%, 20%, 20%) as alternatives to the original choice and the results (Table 4.9) show that the particular seeding does not affect the model output.

4.8. Supplementary Information

I_0	(1%, 1%, 39%, 39%, 20%)	(10%, 20%, 40%, 20%, 10%)	(20%, 20%, 20%, 20%, 20%)
R_0	1.412 (1.405, 1.418)	1.411 (1.405, 1.417)	1.411 (1.405, 1.418)
I_0	44 (38, 52)	37 (31, 44)	44 (37, 51)
ρ_1	0.189 (0.182, 0.195)	0.188 (0.182, 0.195)	0.189 (0.182, 0.195)
ρ_2	0.175 (0.171, 0.179)	0.175 (0.171, 0.179)	0.175 (0.171, 0.179)
ρ_3	0.055 (0.053, 0.057)	0.055 (0.054, 0.057)	0.055 (0.054, 0.057)
ρ_4	0.035 (0.033, 0.037)	0.035 (0.033, 0.037)	0.035 (0.033, 0.038)

Table 4.9: Sensitivity analysis on the distribution of the initial cases I_0 : mean and equal-tailed 95% credible interval of the marginal posterior distribution of the parameters for the “Age-Dependent Reporting Susceptibility” model without overdispersion.

Regarding the mean length of the latent period, the results given in the paper have been obtained under the assumption that the mean length of the latent period is of 1 day and the mean generation time of 2.6 days. Here we assume a mean length of the latent period of 1.3 days and the same mean generation time of 2.6 days. The results obtained under this assumption are given on Table 4.10. The estimates are robust also to the hypothesis on the length of the latent period.

	Basic model	ADR model	TVR model
R_0	1.357 (1.351, 1.362)	1.405 (1.398, 1.411)	1.377 (1.365, 1.390)
I_0	138 (118, 158)	39 (33, 46)	74 (51, 102)
ρ_1	0.086 (0.084, 0.0875)	0.191 (0.185, 0.198)	0.183 (0.140, 0.228)
ρ_2	0.086 (0.084, 0.0875)	0.177 (0.173, 0.181)	0.164 (0.127, 0.202)
ρ_3	0.086 (0.084, 0.0875)	0.056 (0.054, 0.058)	0.053 (0.041, 0.066)
ρ_4	0.086 (0.084, 0.0875)	0.036 (0.034, 0.038)	0.035 (0.027, 0.044)
a			1.255 (0.946, 1.684)
b			0.609 (0.416, 0.844)

Table 4.10: Susceptibility model: mean and equal-tailed 95% credible interval of the marginal posterior distribution of the parameters for the “Susceptibility” model without overdispersion assuming a mean length of the latent period of 1.3 days and a mean generation time T_g of 2.6 days.

4.8.5 How data are reproduced by the model

In order to validate the model, we compare the predicted age-structured weekly incidence of A/H1N1 cases with the data.

In the deterministic case $Z_t^i = \bar{Z}_t^i$ we draw 1000 (sets of) parameters from the joint posterior distribution and for each draw we numerically solve the SEIR model in the

Italian population. After rescaling into the patients population (and for each drawn set of parameters) we obtain a realization of \bar{Z}_t^i , the number of patients infected by the A/H1N1 virus. Given \bar{Z}_t^i and the selected reporting rate(s), we apply the Binomial model given in equation (3) of the main text and draw 100 realizations of F_t^i , the number of flu (H1N1) cases within the patients population. We can thus compute the incidence (per thousand) of H1N1 infections to be compared to the observed A/H1N1 incidence curve (red dots of Figure 2 in the main text).

When Z_t^i is taken as a random variable, for each of the 500 parameter draws from the joint posterior distribution we repeat the procedure described above and obtain \bar{Z}_t^i . For each \bar{Z}_t^i we draw 20 values of Z_t^i from the Negative Binomial distribution given on equation (12) in the main text having fixed the dispersion parameter r to the specified value and then for each of this Z_t^i generate 10 F_t^i as random draws from the Binomial model given in equation (13) of the main text.

The number of draws to be performed from the distributions defined in the fixed $Z_t^i = \bar{Z}_t^i$ case and in the random Z_t^i one has been chosen so that to obtain the same number of realizations of F_t^i .

Figure 4.6 shows the comparison between the incidence data and the predictions obtained by the “Immunity” model without overdispersion.

4.8. Supplementary Information

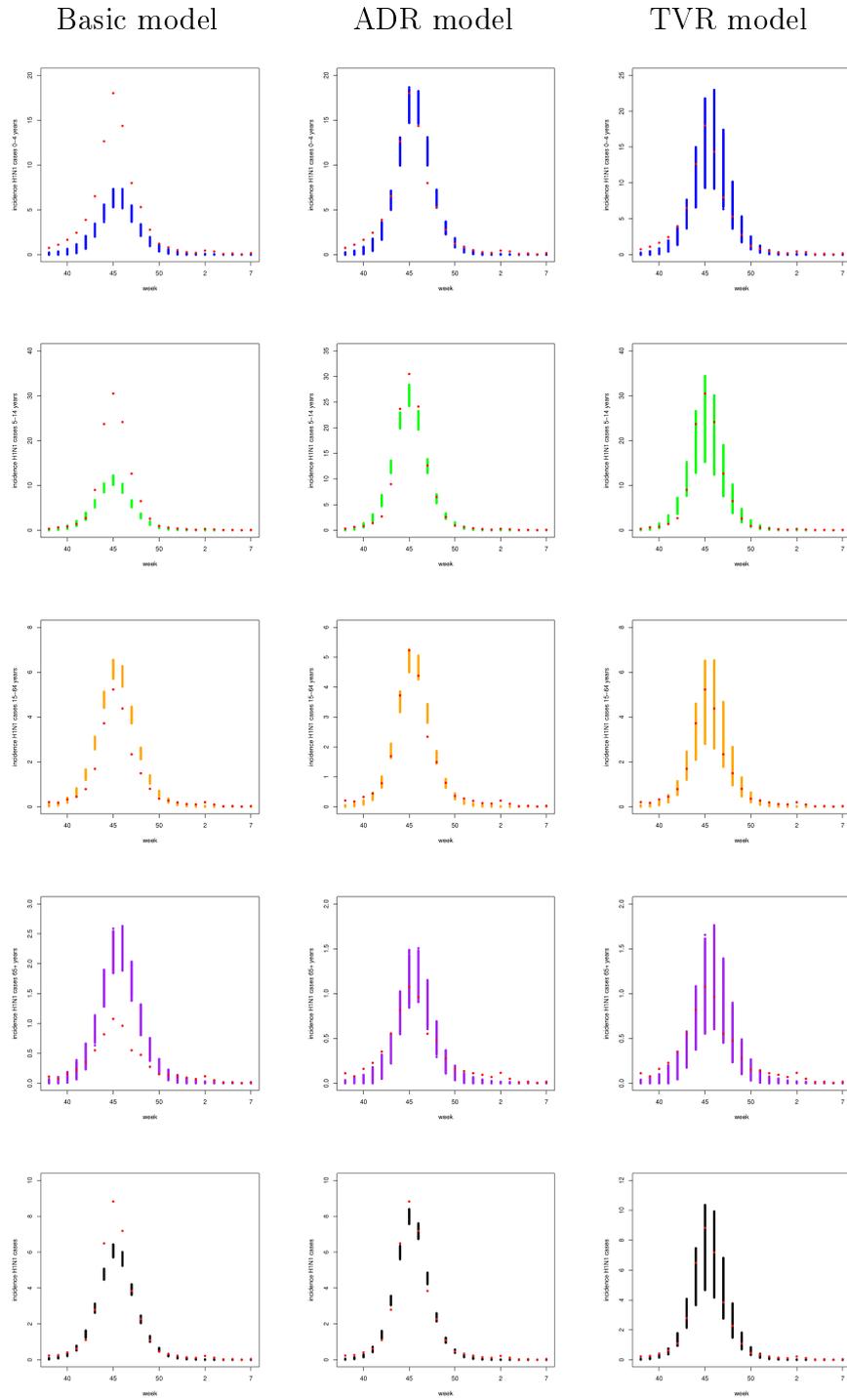


Figure 4.6: Immunity model without overdispersion: plot of the simulated weekly reported incidence (per thousand) of the new H1N1 cases in the 0 – 4 years age-class (blue), 5 – 14 years age-class (green), 15 – 64 years age-class (orange), 65+ years age-class (purple) and in the population as a whole (black) in comparison to the respective observed data (red).

Chapter 5

Estimation of R_0 from real and simulated school outbreaks

5.1 Introduction

The recent 2009-2010 pandemic influenza A/H1N1 virus mostly affected young people, in particular those in school-age years. The high number of cases observed among the youngest age-classes can be explained in terms of higher levels of susceptibility of children with respect to adults, something confirmed also by serological studies (Miller et al., 2010; Rizzo et al., 2010), and by the fact that transmission is favoured by the high contact rates occurring among children within schools.

There is evidence that schools play a crucial role in the transmission of infectious diseases such as influenza so that the impact of schools closure and the extent to which this non-pharmaceutical intervention can be used to reduce the total number of cases and slow the epidemic has been widely investigated and discussed (Ferguson et al., 2006; Cauchemez et al., 2008, 2009b; Wu, 2010).

Whenever a new infectious disease emerges, the estimation of the pathogen transmissibility is an urgent issue which is often addressed by modelling the infection spread using an infectious scheme of SIR type and by estimating the real-time growth rate on the number of syndromic or laboratory confirmed cases observed in the population (if available). Surveillance systems monitor the spread of an infection at a national level and surveillance data are often uninformative at the very early stages of disease transmission, due to the very low incidence (in terms of both number of syndromic cases and confirmed cases). On the contrary, clusters of cases are more easily monitored in small communities or specific social contexts and as

5.2. Estimation of exponential growth rate

a matter of fact the real-time estimation of infection transmissibility has been often performed on data collected in households (Yang et al., 2009b), schools (Yang et al., 2009b; Lessler et al., 2009) and small communities (Fraser et al., 2009).

The estimates for the reproduction number R_0 obtained in school settings (Yang et al., 2009b; Lessler et al., 2009; Nishiura et al., 2009) are generally higher than those obtained in community settings (Ghani et al., 2009; Fraser et al., 2009; Nishiura et al., 2010). It has been argued that the early estimation of the reproduction number R_0 obtained from the analysis of data collected in specific social contexts (such as schools) overestimates the transmissibility of infection at the level of the general community (Nishiura et al., 2010). Using an individual-based model recently developed to model the A/H1N1 influenza pandemic in Europe (Merler and Ajelli, 2010) we compare here the estimates of the reproduction number R_0 obtained from the curve of the cases observed in the general community and from the curve of the cases observed within selected samples of schools; this is compared also to the theoretical value of within school reproduction number, and to the mean number of cases actually generated by the index case in a “random” school. Model schools have been selected for the analysis either as among the schools with the highest number of cases (to simulate the ones that would presumably be chosen as target of outbreak analysis) or at random among all schools that had at least one case.

Finally, we present a first analysis of the data collected through a survey in two Italian primary schools after the influenza outbreak during the 2009-2010 H1N1 pandemic. These schools were chosen purely for convenience and should represent a “random” school from the point of view of influenza outbreaks.

5.2 Estimation of exponential growth rate

Infections emerging in large populations are characterized by the fact that the initial stochastic fluctuations in the number of cases is soon overcome by a clear exponential growth phase. Whenever an epidemic is observed in smaller contexts such as schools, the stochastic fluctuation in the number of cases cannot be neglected and an estimation of the growth-rate becomes more challenging.

In the literature, a few methods have been used to estimate the exponential growth rate of the number of cases and in this work we are going to consider three very simple of these techniques (Chowell et al., 2007b; Favier et al., 2006). The methods considered here are fast but not very refined and more accurate estimates can be

obtained with other and more complex techniques (Cauchemez et al., 2010).

Let us assume that, at the early stages of an epidemic, the number of cases grows exponentially. In other words, we assume that the incidence in the number of the new cases $i(t)$ grows exponentially with rate r in time t

$$i(t) = ke^{rt} \quad (5.1)$$

with $k \in \mathbb{R}$ a constant.

By definition, the cumulative function $c(t)$ is given by

$$c(t) = \int_0^t i(s)ds = \frac{k}{r}(e^{rt} - 1) \quad (5.2)$$

By applying the logarithmic function to $i(t)$, we explicitly find a linear relationship

$$\log_e(i(t)) = rt + \log_e k$$

By applying the logarithmic function to the cumulative $c(t)$, we obtain

$$\log_e(c(t)) = \log_e(e^{rt} - 1) + \log_e\left(\frac{k}{r}\right) = rt + \log_e(1 - e^{-rt}) + \log_e\left(\frac{k}{r}\right). \quad (5.3)$$

If we restrict ourselves to consider t large enough that $\log_e(1 - e^{-rt}) \approx 0$ (but small enough that the exponential phase is still going on), we see that r can be obtained as the slope of a line approximating the values of $\log_e(c(t))$. The plot of function $\log_e(c(t))$ versus time t (Figure 5.2) visually shows the approximately linear behaviour of (5.3) for t large enough

5.2. Estimation of exponential growth rate

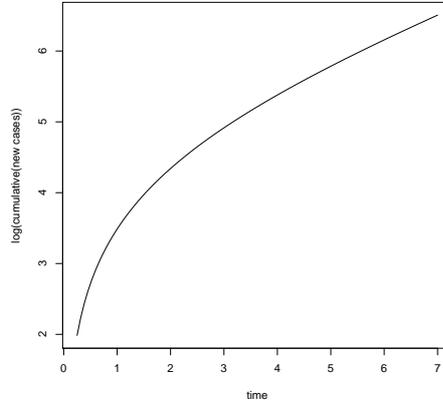


Figure 5.1: Plot of the logarithm of the cumulative function $c(t)$ defined in (5.2) versus time t .

Finally, notice the linear relationship existing between $i(t)$ and $c(t)$

$$i(t) = rc(t) + k$$

Here we estimate the exponential growth rate r through the least square fit of a linear model to the observed quantities just introduced and here summarized:

- L.1) to the logarithm of the incidence versus time [i.e. $\log_e(i(t))$ vs t];
- L.2) to the logarithm of the cumulative number of cases versus time (Chowell et al., 2007b) [i.e. $\log_e(c(t))$ vs t];
- L.3) to the incidence versus the cumulative number of cases (Favier et al., 2006) [i.e. $i(t)$ vs $c(t)$].

These methods should be applied to data coming from a phase of exponential growth in incidence. Furthermore, for method L.2, the temporal window should exclude times too close to 0, where $\log_e(1 - e^{-rt})$ is not negligible (Merler and Ajelli, 2010). Some care hence has to be taken in choosing a temporal window of data to which these methods should be applied. In order to decrease subjectivity in this choice, we proceeded through a semi-automatic procedure.

The linear models are fitted on a sequence of temporal windows $[t_i, t_i + \delta_t]$, where t_i denotes the time at which the i -th case has been detected in the school and δ_t assumes every integer value within $[t_{i+1} - t_i, t_{max} - t_i]$, with t_{max} denoting the

5.2. Estimation of exponential growth rate

time at which the last case is detected in the school. We take time t_1 (i.e. the time at which the index case is detected into a school) as a reference and set $t_1 = 1$. We adopted an iterative rule to draw a sample from the vector of the cumulative number of cases c_t : starting from $c_1 = 1$, we generate a sample of length n_{it} by iterating $n_{it} - 1$ times the following rule

$$c_{i+1} = c_i + \frac{\alpha}{100} N_s \quad (5.4)$$

with N_s denoting the school size and α to be chosen so that the sample covers a fair part of the observed epidemic. We generate the sequence of starting times of the temporal windows t_i by associating to each sampled cumulative number the time at which it was first observed.

From an epidemiological point of view, we assume to model the infection dynamics using a SEIR model with an exponentially distributed latent period of mean \bar{T}_E and an exponentially distributed infectious period of mean length \bar{T}_I . Therefore, if r denotes the exponential growth rate in the number of cases, the reproduction number R_0 is given by

$$R_0 = (1 + r\bar{T}_E)(1 + r\bar{T}_I) \quad (5.5)$$

Finally, let y_t denote the observed data at time t and $l_t = q + rt$ denote the value predicted by linear regression. We take R^2 , the fraction of the total squared error explained by the model

$$R^2 = 1 - \frac{\sum_t (y_t - l_t)^2}{\sum_t (y_t - \bar{y})^2} \quad (5.6)$$

as measure of goodness of fit, with \bar{y} representing the sample mean.

5.3 The individual-based model

This part of the work makes use of an extended version of the discrete-time, stochastic, spatially-explicit, individual-based SEIR simulation model recently developed by Merler and Ajelli (2010) to model the spatio-temporal spread of the pandemic H1N1 virus in Europe. Here we briefly review some key aspects of the model and refer to Merler and Ajelli (2010) (in particular to the relative Supplementary Material) for a detailed and exhaustive description of the sociodemographic and the epidemiological models.

The authors divide the study area into cells (average surface of about 77 kilometers) and on the defined spatial grid generate a synthetic population matching the census data. Schools and workplaces are distributed proportionally to the population which is grouped into households, schools and workplaces following the country specific sociodemographic data. In this model schools gather together individuals within a wide age range, from nursery school to university (a more realistic procedure of school assignment is being developed, but the current analysis refers to this version). Individuals are explicitly represented and characterized by household and school/workplace membership (if any). Transmission of infection occurs in households, schools, workplaces and by random contacts with infectious individuals in the global population (random contacts can be made within a radius of 1000 kilometers and according to the power-law distribution given in eq. 5.8 (Gonzalez et al., 2008)).

The original model (Merler and Ajelli, 2010) has been recently extended to explore the heterogeneity in the patterns of spread observed in the past 2009-2010 H1N1 pandemic in Europe (personal communication). Here we use this last version of the individual-based model, whose mainly difference from the original one consists in the introduction of a differential susceptibility: children (< 16 years) are assumed twice as susceptible to infection as adults. For the purpose of simplicity we further assume that all infected individuals are symptomatic and that sickness-induced absenteeism does not occur. Moreover, for the sake of computational speed, the model has been used on a single country (Italy) and workplace transmission (which is not relevant to the present analysis) has been not explicitly modelled but included within the general community transmission.

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. 5.7);

2. contacts with infectious individuals attending the same school (second term in eq. 5.7);
3. random contacts in the population (third term in eq. 5.7);

$$\begin{aligned}
 \lambda_i = & \sum_{\{k=1, \dots, N | H_k = H_i\}} \frac{I_k \beta_h}{n_i} \\
 & + \sum_{\{k=1, \dots, N | P_k = P_i\}} \frac{I_k \beta_s}{N_i} \\
 & + \sum_{\{k=1, \dots, N\}} \frac{I_k \beta_c f(d_{ik})}{\sum_{\{k=1, \dots, N\}} f(d_{ik})}
 \end{aligned} \tag{5.7}$$

The terms in equation (5.7) are defined as follows:

- H_i is the index of the household where individual i lives in;
- P_i is the index of the school where individual i studies (if i is a student);
- N is the size of the Italian population;
- n_i is the size of household H_i ;
- N_i is the size of school P_i ;
- $I_k = 1$ if individual k is infected, 0 otherwise;
- $f(d_{ik})$ is the function defined in in (Gonzalez et al., 2008) and here recalled

$$f(d_{ik}) = (d_{ik} + r_g^0)^{-\beta} e^{-\frac{r_g}{k}} \tag{5.8}$$

where $r_g^0 = 5.8\text{km}$, $\beta = 1.65$ and $k = 350\text{km}$. It makes the transmission of the epidemic in the general community explicitly dependent on patterns of human mobility, as described in (Gonzalez et al., 2008);

- β_h (expressed in day^{-1}) is the within-household transmission rate;
- β_s (in day^{-1}) is the within-school transmission rate;
- β_c (in day^{-1}) is the transmission rate in the general community.

5.3.1 Parameterization of the individual-based model

We adopt here the baseline parametrization proposed in (Merler and Ajelli, 2010): the discrete-time, stochastic SEIR model assumes an exponentially distributed latent period T_E of mean 1.5 days and an exponential distribution of the infectious period T_I of mean 1.6 days with constant infectiousness during the whole course of infection. Each epidemic is started by seeding 100 cases at random in the Italian population, and the time step Δ_t of the model has been fixed to 0.5 days. The “global” reproduction number R_0 is computed using formula (5.5) and the exponential growth rate r is estimated by fitting a linear model to the logarithm of the cumulative number of new cases generated in the global population in time (i.e. using method L.2). We run 100 simulations and compute the mean “global” R_0 and the mean number of cases generated within households, schools and by random contacts in the initial phase of disease transmission (i.e. on the first 1000 cases generated in the whole population) and at the end of the epidemic. The estimates are very stable among the model realizations so that we are going to report only the mean values computed on the 100 realizations. The adopted parametrization ($\beta_h = 0.691$, $\beta_s = 0.771$, $\beta_c = 0.506$) results in a mean “global” R_0 of 1.38. At the early stages of disease transmission the percentage of cases generated within households is on average 28%, the percentage of cases generated within schools is on average 37% and those generated in the general community amounts to 35%. At the end of the epidemic the proportion of cases generated in households, schools and in the general community are respectively of 31%, 30% and 39%.

5.3.2 Computation of the within school reproduction number

The model assumes homogeneous mixing within schools and households. If we assume that school s represents an isolated population, we can define the within school reproduction number R_s as the mean number of cases generated by a typical infectious individual within school s at the beginning of an epidemic.

Recall that β_s represent the within school transmission rate and let γ be the recovery rate (i.e. $\bar{T}_I \approx 1/\gamma$). Furthermore, N_s represents here the size of school s and remember that the simulation time-step has been fixed to $\Delta_t = 0.5$ days.

Let us assume that at the beginning of an epidemic one infectious case is seeded in school s and that the remaining school population is susceptible and large enough to assume that $N_s - 1 \approx N_s$. Any susceptible member i of school s is subjected

to a probability P_i of becoming infected within a time step by a given infectious individual, with P_i given by

$$P_i = 1 - e^{-\frac{\beta_s \Delta_t}{N_s}}. \quad (5.9)$$

If x infectives are present, the probability that a susceptible escapes infection for a time step will be $(1 - P_i)^x$.

On the basis of our assumptions (that imply that every infective is infectious at least for a period Δ_t), the infectious period has length $n\Delta_t$ with probability

$$P(T_I = n\Delta_t) = e^{-\gamma(n-1)\Delta_t}(1 - e^{-\gamma\Delta_t}) \quad \text{for } n = 1, 2, \dots$$

The mean length of the infectious period is hence given by

$$\begin{aligned} \bar{T}_I &= \sum_n n\Delta_t e^{-\gamma(n-1)\Delta_t}(1 - e^{-\gamma\Delta_t}) \\ &= (1 - e^{-\gamma\Delta_t})\Delta_t \sum_n n e^{-\gamma(n-1)\Delta_t} \\ &= (1 - e^{-\gamma\Delta_t}) \frac{\Delta_t}{(1 - e^{-\gamma\Delta_t})^2} \\ &= \frac{\Delta_t}{(1 - e^{-\gamma\Delta_t})} \end{aligned} \quad (5.10)$$

With our parametrization ($\gamma = 1/1.6$ and $\Delta_t = 0.5$) we find $\bar{T}_I = 1.86$ and notice that

$$\lim_{\Delta_t \rightarrow 0} \bar{T}_I = \frac{1}{\gamma}.$$

Similarly, the number T of temporal steps spent in the infectious stage is given by

$$P(T = n) = \rho(1 - \rho)^{n-1} \quad n \geq 1 \quad (5.11)$$

with $\rho = (1 - e^{-\gamma\Delta_t})$. Let p_s represent the fraction of children (< 16 years) attending school s and σ_a represent the susceptibility of adults with respect to children ($\sigma_a = 0.5$). Let C_k denote the number of cases generated by an infectious individual at time step k ; we assume that C_k are independently distributed according to a binomial model

$$C_k \sim \text{Bin}(N_s, \omega) \quad \text{with } \omega = (p_s + (1 - p_s)\sigma_a) \left(1 - e^{-\frac{\beta_s \Delta_t}{N_s}} \right)$$

5.3. The individual-based model

The number of cases H generated by an infectious individual at the beginning of an epidemic is a random variable defined by

$$H = \sum_{k=1}^T C_k \quad (5.12)$$

The mean number of cases generated by a typical infectious individual within school s at the beginning of the epidemic (i.e. assuming that all possible contacts are susceptible) is given by

$$R_s = \mathbb{E}[H] = \mathbb{E}[\mathbb{E}[H|T]] = \frac{N_s \omega}{\rho} = N_s (p_s + (1 - p_s) \sigma_a) \frac{(1 - e^{-\frac{\beta_s \Delta_t}{N_s}})}{(1 - e^{-\gamma \Delta_t})} \quad (5.13)$$

where \mathbb{E} denotes the expected value.

Notice that the reproduction number R_s depends on the length of the simulation step Δ_t , on the size of the school N_s and on the fraction of children p_s attending the school. Note furthermore that

$$\lim_{N_s \rightarrow \infty} R_s = \frac{\beta_s \Delta_t}{(1 - e^{-\gamma \Delta_t})} (p_s + (1 - p_s) \sigma_a)$$

and that (as can be seen from Figure 5.2) already for $N_s \approx 100$ the value of R_s is not very far from its limit value, that is almost reached for $N_s \geq 600$.

Considering instead the effect of time step (which is actually kept fixed in the analysis here), note that

$$\lim_{\Delta_t \rightarrow 0} R_s = \frac{\beta_s}{\gamma} (p_s + (1 - p_s) \sigma_a)$$

i.e. the usual value (independent of N_s) for continuous-time models. Furthermore note that R_s is an increasing function of Δ_t as long as $\gamma > \beta_s/N_s$ which is true in our examples, as well as in any reasonable scenario.

The variance of H (5.12) is given by

$$\begin{aligned} V(H) &= \mathbb{E}[H^2] - (\mathbb{E}[H])^2 \\ &= N_s^2 \omega^2 (1 - \rho) / \rho^2 + N_s \omega (1 - \omega) (2 - \rho) / \rho^2 \\ &= \frac{N_s \omega}{\rho^2} (N_s \omega (1 - \rho) + (1 - \omega) (2 - \rho)) \end{aligned} \quad (5.14)$$

We compute the mean school size $\bar{N}_s = 525$ (SD= 236) and the mean fraction of children attending a school $\bar{p}_s = 0.65$ (SD= 0.04) by averaging over 100 model

realizations, the mean values of the respective quantities computed on 100 randomly selected schools. Since the distributions of N_s and p_s are independent in the model, the average value of R_s can be obtained inserting \bar{N}_s and \bar{p}_s in (5.13) obtaining 1.184; with \bar{N}_s and \bar{p}_s the variance given in (5.14) results 8.66. The standard deviation (SD) reported within brackets has been obtained by averaging over the 100 model realization the standard deviations obtained on the 100 schools sampled for each model realization. On Figure 5.2 we plot R_s given in (5.13) as a function of the school size N_s having fixed $p_s = \bar{p}_s = 0.65$ and $\Delta_t = 0.5$.

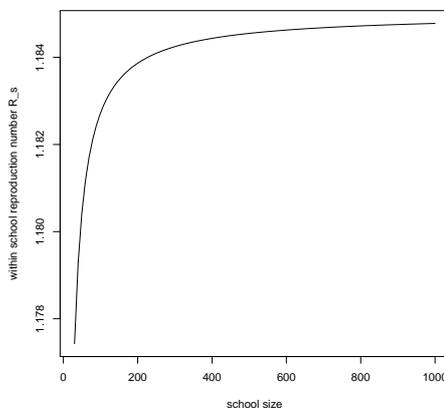


Figure 5.2: Plot of R_s given in (5.13) as a function of the school size N_s having fixed $p_s = \bar{p}_s = 0.65$ and $\Delta_t = 0.5$.

5.3.3 Analysis of simulated school epidemics

In this section we estimate the within school reproduction number R_s using different techniques.

Estimate of R_s using the infection tree

For the first 20000 cases of each model realizations we keep track of the infection tree (i.e. who infected whom). Following the definition of R_s , we count here the mean number of cases generated (within the school) by a “typical” infectious case at the beginning of the epidemic. For each of 100 model realizations we randomly draw 100 schools among those infected relatively early in time (i.e. among the schools that had at least one infectious case within the time of occurrence of the first 20000 cases in the global community) and for each school we count the number of cases

5.3. The individual-based model

generated by the index case. By averaging over the sampled 100 schools and over the 100 simulations we obtain an estimate of R_s amounting to 1.51 (SD= 1.83). Standard deviation (SD) has been computed by averaging over the 100 simulations the standard deviations computed for each simulation on the sampled schools. It has to be noted that the value obtained in this way is quite higher than the average value obtained from (5.13), although, given the large SD, the confidence interval one would obtain includes the theoretical value.

Estimate of R_s by the estimation of the exponential growth rate

The individual-based model allows us to keep track of the place where each case acquired infection and hence to distinguish a case infected within the school from a case infected anywhere and attending the school. For each simulated school epidemic s , we can hence distinguish between the curve of the cases *generated* within school s and *observed* in school s (but generated anywhere).

We propose here the analysis of 10 simulated school epidemics, which have been sampled as follows. We first randomly draw 10 out of 100 realizations of the model. For technical reasons (the infection-tree has been recorded up to the first 20000 cases in the general community), we selected the schools to be analysed among those infected relatively early in time (i.e. within the first 20 days since the start of the epidemic in the country). From realizations 1 to 5 we draw one school (for each realization) at random. From simulations 6 to 10 we choose the school that accounts for the largest number of cases among the schools infected during the initial phase of the epidemic. This last choice has been done to explore any dependency between the number of cases and the within school reproduction number R_s . Moreover, these last schools would presumably be chosen as target of outbreak analysis. For simplicity, the sampled schools have been numbered accordingly to the simulation from which they were drawn ($s = 1, \dots, 10$). Table 5.1 summarizes the characteristics of the selected schools (i.e. school size N_s and the fraction of young (< 16 years) population p_s), the “actual” value of R_s given by formula (5.13), the number of cases generated and observed within the respective schools and the corresponding attack rate (AR) computed on the number of observed cases until the time of occurrence of the 20000-th case.

s	N_s	p	R_s	generated cases	observed cases	AR
1	871	0.669	1.198	75	116	0.133
2	842	0.690	1.213	46	72	0.085
3	360	0.655	1.188	89	125	0.347
4	788	0.700	1.220	34	55	0.069
5	496	0.705	1.224	67	106	0.214
6	511	0.690	1.213	141	212	0.415
7	894	0.673	1.201	154	273	0.265
8	771	0.674	1.201	155	236	0.306
9	387	0.705	1.223	116	183	0.473
10	775	0.698	1.219	110	172	0.222

Table 5.1: Some basic statistics on the simulated school epidemics.

For each school $s = 1, \dots, 10$ we applied the iterative rule (5.4) with $\alpha = 0.3$ and $n_{it} = 10$ to generate the temporal-intervals to be used for the fit of a linear model on the logarithm of the number of new cases in time (method L.1), on the cumulative number of new cases in time (method L.2) and on the incidence versus the cumulative (method L.3). We required the estimated values of the exponential growth rate to satisfy the conditions given on Table 5.2; moreover, we discarded the estimates obtained on inappropriate temporal frames and those considered not informative enough, as specified next.

method	R^2
L.1	> 0.20
L.2	> 0.95
L.3	> 0.10

Table 5.2: Simulated school epidemics: threshold values for R^2 , for the different methods L.1, L.2 and L.3. The values reported on Table 5.1 satisfy the constraints here defined.

For each simulation, we discarded the estimates computed on the very initial generations using method L.2 and those computed after the exponential growth phase using methods L.1 and L.3. Given the random nature of the simulated school epidemics in terms of start and length of the exponential growth phase, we carefully selected the appropriate time-frames on which to perform linear regression, for every method and for every selected school epidemic on the basis of the behaviour of the school epidemics themselves. We discarded also the estimates obtained on

5.3. The individual-based model

time intervals shorter than 7 days (i.e. the estimates obtained for $\delta_t < 14$) and those computed on time intervals longer than 14 days (i.e. the estimates obtained for $\delta_t > 28$). Notice that for each school epidemic we estimate the exponential growth rate on two epidemic curves: those of the cases generated within the school and the curve of the observed cases. Table 5.3 summarizes the range for the estimates of R_s obtained with the three methods L.1, L.2 and L.3 on the curve of the cases generated and observed in the respective schools.

s	R_s	method L.1	method L.1	method L.2	method L.2	method L.3	method L.3
		generated	observed	generated	observed	generated	observed
1	1.198	1.15-1.23	1.15-1.41	1.13-1.21	1.12-1.18	1.13-1.53	1.14-1.81
2	1.213	1.11	1.17-1.28	1.08-1.24	1.07-1.23	1.16-1.40	1.19-1.30
3	1.188	1.14-1.39	1.17-1.40	1.19-1.27	1.17-1.24	1.13-1.69	1.16-1.67
4	1.220	1.18	1.15	1.14-1.21	1.15-1.21	1.25-1.35	1.16-1.32
5	1.224	1.11-1.51	1.24-2.13	1.11-1.15	1.13-1.16	1.24-1.46	1.24-1.64
6	1.213	1.15-1.67	1.19-1.83	1.17-1.20	1.15-1.18	1.11-1.76	1.14-1.69
7	1.201	1.13-1.23	1.15-1.50	1.17-1.21	1.17-1.19	1.12-1.43	1.13-1.56
8	1.201	1.14-1.36	1.15-1.48	1.19-1.40	1.22-1.41	1.12-1.63	1.13- 1.77
9	1.223	1.1-1.22	1.21-1.41	1.17-1.23	1.15-1.21	1.17-1.33	1.18-1.39
10	1.219	1.12-1.20	1.13-1.31	1.14-1.29	1.13-1.26	1.11-1.30	1.13-1.19

Table 5.3: Ranges (i.e. maximum and minimum value) of R_s estimated using methods L.1, L.2 and L.3 for the relative school epidemics and the theoretical value of R_s given by formula (5.13) on the basis of the data provided on Table 5.1. The selected estimates satisfy the constraints reported on Table 5.2 and the choice of the temporal intervals used to perform linear regression (in the three variants L.1, L.2 and L.3 and for each school epidemic) has been discussed in the text.

All methods L.1, L.2 and L.3 produce reasonable estimates of R_s in the appropriate time-intervals. Method L.2 revealed itself as the most stable method among the three. More precisely, the mean of the squared differences between the midpoints of the ranges and the relative values of R_s given by theory is smallest for method L.2; method L.3 tends to overestimate the “actual” value of R_s (i.e. all the midpoints of the ranges obtained with method L.3 overestimate the relative values of R_s given by theory, except for $s = 10$). Therefore, method L.2 seems to be preferable to the others, at least for data generated by the simulation model adopted here.

5.4 Real school outbreaks

5.4.1 The survey

In Italy the first (or primary) level of compulsory education starts at the age of 6 and ends at 11 years. During the 2009-2010 A/H1N1 influenza pandemic and precisely over the months of October-November 2009, two primary schools located in the villages of Povo and Villazzano (Trento, Italy) experienced a clear epidemic. The survey, conducted on December 2009 in the two schools in question, aimed at retrospectively reconstruct the outbreaks occurred over the previous months in each school. To each family of the schools we delivered a paper questionnaire composed by a first part, where we gave the definition of influenza-like-illness (ILI) and by a second part, where the parents on behalf of their children were asked to report the date(s) of onset of ILI symptoms in the members of the family. Table 5.4 summarizes some basic data and statistics collected at the time of the survey.

	school of Povo	school of Villazzano
school size	307	213
number of classes	14	10
number of responses	260	168
number of ILI cases	121	103
response rate	0.85	0.79
attack rate	0.46	0.61

Table 5.4: Some basic statistics on the survey led in the primary schools of Povo and Villazzano.

5.4.2 Analysis of real school epidemics

School of Povo

On the basis of the data collected through the survey, two cases seeded infection in the school of Povo ($t_1 = t_2 = 1$) and the last case showed the onset of ILI symptoms 56 days after the index cases ($t_{max} = 56$). Figure 5.3 shows the number of new cases (left panel) and the cumulative number of ILI cases (right panel) observed in the school over time.

5.4. Real school outbreaks

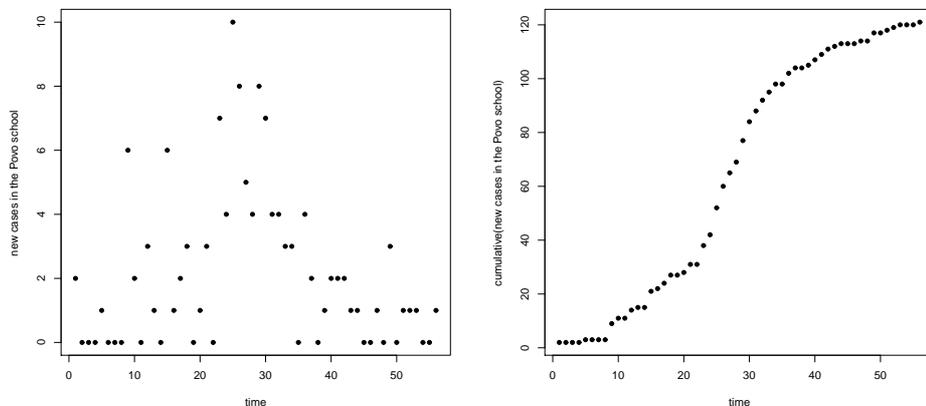


Figure 5.3: Plot of the number of new cases (left panel) and of the cumulative number of observed new cases (right panel) in the school of Povo in time, starting from the day of detection of the index case.

Using rule (5.4) with $\alpha = 2.5$ and $n_{it} = 15$ we define the temporal-windows on which to perform the linear regressions. In terms of goodness of fit R^2 we select the estimates that satisfy the constrains given on Table 5.5. The thresholds values given on Table 5.5 have been chosen on the basis of the relative average values of R^2 observed when fitting a linear model to this school epidemic. Indeed, the Povo school epidemic could be better explained (using a linear model) than the simulated school epidemics given that, on average, we obtained higher scores for the goodness of fit R^2 .

method	R^2
L.1	> 0.30
L.2	> 0.98
L.3	> 0.50

Table 5.5: Threshold values for R^2 for the estimates obtained for the school of Povo.

Tables 5.6, 5.7 and 5.8 summarize the estimated growth rate r and the corresponding values of R_0 and R^2 computed on the specified time intervals $[t_i, t_i + \delta_t]$ using linear regression on the three quantities L.1, L.2 and L.3. Figures 5.4(a), 5.4(b), 5.4(c) plot the linear fit marked with an asterisk in the respective tables and the filled dots in the plot represent the data on which linear regression has been performed.

L.1 - linear regression on $\log_e(i(t))$ vs t					
c_i	t_i	δ_t	r	R^2	R_s
1	1	29	0.0634	0.426	1.21
1	1	30	0.0611	0.430	1.20
1	1	31	0.0589	0.433	1.19
13	12	18	0.0798	0.327	1.26 *
13	12	19	0.0724	0.312	1.24
25	18	8	0.216	0.432	1.78
25	18	9	0.184	0.416	1.65
25	18	10	0.149	0.364	1.52
25	18	11	0.146	0.416	1.51
25	18	12	0.136	0.435	1.47

Table 5.6: School of Povo: summary of the estimated values of the exponential growth rate r obtained through the fit of $\log_e(i(t))$ vs t and the corresponding R_0 . Linear regression has been performed on the time intervals given by $[t_i, t_i + \delta_t]$. The marked (*) linear fit is plotted on Figure 5.4(a).

L.2 - linear regression on $\log_e(c(t))$ vs t					
c_i	t_i	δ_t	r	R^2	R_s
13	12	19	0.0993	0.982	1.33
13	12	20	0.0982	0.984	1.33
13	12	21	0.0966	0.984	1.32
13	12	22	0.0947	0.982	1.32 *
33	23	3	0.158	0.983	1.55
41	24	2	0.178	0.987	1.63
45	25	5	0.0916	0.988	1.3
45	25	6	0.0865	0.985	1.29

Table 5.7: School of Povo: summary of the estimated values of the exponential growth rate r obtained through the fit of $\log_e(c(t))$ vs t and the corresponding R_0 . Linear regression has been performed on the time intervals given by $[t_i, t_i + \delta_t]$. The marked (*) linear fit is plotted on Figure 5.4(b).

5.4. Real school outbreaks

L.3 -linear regression on $i(t)$ vs $c(t)$					
c_i	t_i	δ_t	r	R^2	R_s
1	1	8	0.973	0.973	6.29
1	1	9	0.49	0.572	3.09
1	1	28	0.0959	0.508	1.32
1	1	29	0.0913	0.537	1.30
13	12	3	0.667	0.651	4.13
13	12	14	0.171	0.529	1.60 *
17	15	11	0.195	0.521	1.70
21	15	11	0.195	0.521	1.70

Table 5.8: School of Povo: summary of the estimated values of the exponential growth rate r obtained through the fit of $i(t)$ vs $c(t)$ and the corresponding R_0 . Linear regression has been performed on the time intervals given by $[t_i, t_i + \delta_t]$. The marked (*) linear fit is plotted on Figure 5.4(c).

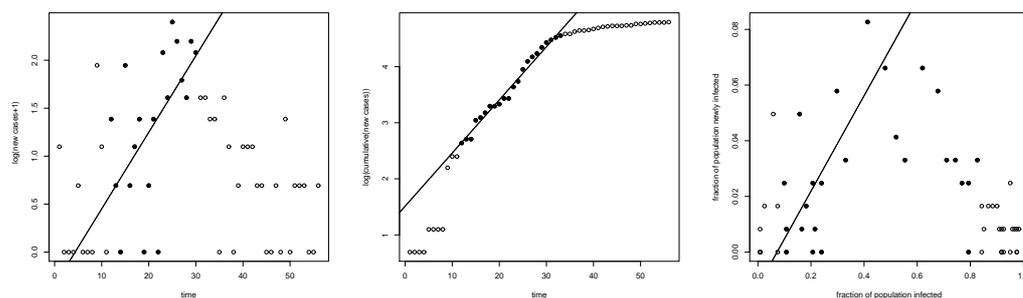


Figure 5.4: Plot of: (a) the incidence data collected in the Povo school (lin-log scale) and the best linear approximation obtained by linear least square fitting to the filled dots, (b) the cumulative data collected in the Povo school (lin-log scale) and the best linear approximation obtained by linear least square fitting to the filled dots, (c) the incidence as a function of the cumulative data collected in the Povo school and the best linear approximation obtained by linear least square fitting to the filled dots.

School of Villazzano

On the basis of the data collected in the survey, one case seeded the infection in the primary school located in Villazzano ($t_1 = 1$) and the last case showed the onset of ILI symptoms 64 days after the index cases ($t_{max} = 64$). Figure 5.5 plots the number of new cases (left panel) and the cumulative number of ILI cases (right panel) observed in the school of Villazzano over time.

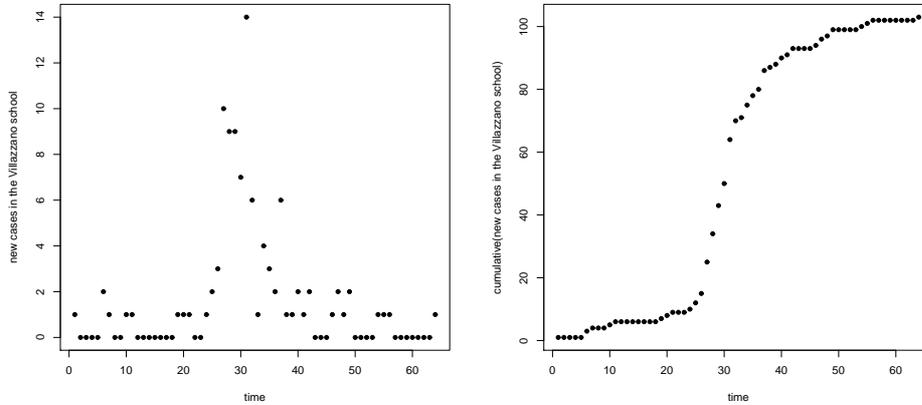


Figure 5.5: Plot of the number of new cases (left panel) and of the cumulative number of observed new cases (right panel) in the school of Villazzano in time, starting from the day of detection of the index case.

To define the temporal windows on which to perform the fit, we apply the iterative rule given in 5.4 with $n_{it} = 15$ and $\alpha = 2.5$ and select the estimates satisfying the constrains given on Table 5.9. The thresholds values given on Table 5.9 have been chosen on the basis of the relative average values of R^2 obtained through the fit of the linear models to this school epidemic. The selection would have been less accurate, if we had applied the threshold values given on Tables 5.2 and 5.5, and the estimates would have been much poorer, in the sense that we would have accepted wider ranges for R_s .

method	R^2
L.1	> 0.60
L.2	> 0.95
L.3	> 0.70

Table 5.9: Threshold values for R^2 for the estimates obtained for the school of Villazzano.

Tables 5.10, 5.11 and 5.12 summarize the estimated growth rate r and the corresponding values of R_0 and R^2 computed on the specified time intervals $[t_i, t_i + \delta_t]$ using linear regression on the three quantities L.1, L.2 and L.3. Figures 5.6(a), 5.6(b) and 5.6(c) plot the linear approximation marked with an asterisk in the respective tables and the filled dots in the plot show the time-interval on which linear regression has been performed.

5.4. Real school outbreaks

L.1 - linear regression on $\log_e(i(t))$ vs t					
c_i	t_i	δ_t	r	R^2	R_s
4	7	24	0.0923	0.558	1.31
4	7	25	0.0922	0.586	1.31
4	7	26	0.0814	0.511	1.27
4	7	27	0.0789	0.52	1.26
7	19	9	0.205	0.562	1.74
7	19	10	0.213	0.647	1.77
7	19	11	0.201	0.674	1.72
7	19	12	0.208	0.736	1.75 *
7	19	13	0.185	0.699	1.65
10	24	5	0.362	0.854	2.44
10	24	6	0.267	0.711	2.00
10	24	7	0.258	0.773	1.96
10	24	8	0.186	0.571	1.66

Table 5.10: School of Vilazzano: summary of the estimated values of the exponential growth rate r obtained through the fit of $\log_e(i(t))$ vs t and the corresponding R_0 . Linear regression has been performed on the time intervals given by $[t_i, t_i + \delta_t]$. The marked (*) linear fit is plotted on Figure 5.6(a).

L.2 - linear regression on $\log_e(c(t))$ vs t						
i	t_i	δ_t	r	R^2	R_s	
10	24	7	0.281	0.98	2.06	
10	24	8	0.263	0.974	1.98	*
10	24	9	0.24	0.954	1.88	
13	26	2	0.409	0.98	2.67	
13	26	3	0.347	0.968	2.36	
13	26	5	0.273	0.958	2.03	
16	27	2	0.271	0.994	2.02	
16	27	3	0.231	0.978	1.85	
16	27	4	0.227	0.988	1.83	
16	27	5	0.206	0.978	1.74	
28	28	2	0.193	0.984	1.69	
28	28	3	0.205	0.992	1.74	
28	28	4	0.184	0.983	1.65	

Table 5.11: School of Vilazzano: summary of the estimated values of the exponential growth rate r obtained through the fit of $\log_e(c(t))$ vs t and the corresponding R_0 . Linear regression has been performed on the time intervals given by $[t_i, t_i + \delta_t]$. The marked (*) linear fit is plotted on Figure 5.6(b).

5.4. Real school outbreaks

L.3 -linear regression on $i(t)$ vs $c(t)$					
i	t_i	δ_t	r	R^2	R_s
1	1	27	0.305	0.788	2.17
1	1	28	0.265	0.822	1.99
1	1	29	0.214	0.774	1.78
1	1	30	0.22	0.856	1.80
1	1	31	0.172	0.736	1.60
4	7	20	0.423	0.834	2.74
4	7	21	0.355	0.875	2.40
4	7	22	0.29	0.864	2.10
4	7	23	0.227	0.789	1.83
4	7	24	0.229	0.867	1.84
4	7	25	0.176	0.73	1.62
7	19	8	0.538	0.933	3.36
7	19	9	0.389	0.886	2.57
7	19	10	0.296	0.837	2.13
7	19	11	0.219	0.721	1.79
7	19	12	0.224	0.824	1.81

Table 5.12: School of Vilazzano: summary of the estimated values of the exponential growth rate r obtained through the fit of $i(t)$ vs $c(t)$ and the corresponding R_0 . Linear regression has been performed on the time intervals given by $[t_i, t_i + \delta_t]$. The marked (*) linear fit is plotted on Figure 5.6(c).

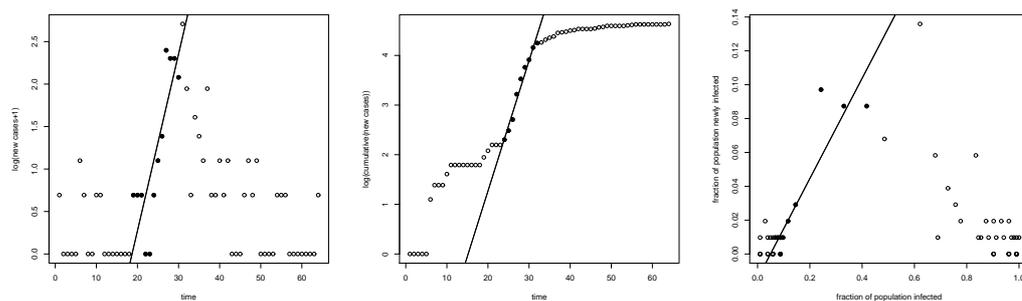


Figure 5.6: Plot of: (a) the incidence data collected in the Villazzano school (lin-log scale) and the best linear approximation obtained by linear least square fitting to the filled dots, (b) the cumulative data collected in the Villazzano school (lin-log scale) and the best linear approximation obtained by linear least square fitting to the filled dots, (c) the incidence as a function of the cumulative data collected in the Villazzano school and the best linear approximation obtained by linear least square fitting to the filled dots.

On the basis of the data collected through the survey conducted in the primary schools of Povo and Villazzano, we estimate that the within school reproduction R_s has been into the range 1.2 – 1.8 in the school of Povo and into the range 1.6 – 2.7 in the school of Villazzano. If we select the estimates obtained with methods L.1 and L.2 on the most reasonable time-intervals, the estimated within school reproduction number is into the range 1.25 – 1.35 for the Povo school and into the range 1.7 – 2.1 for the school of Villazzano. The ranges given here correspond to the values of R_s estimated on different exponential growth time-intervals, provided that we selected the most reasonable time-frames for the fit of a linear model to the data. Our analysis shows that the estimates of the reproduction number obtained by least square fit of a linear model to observed data are sensitive to the choice of the exponential growth phase selected for the fit.

The estimates of the within school reproduction number R_s computed in this work are smaller but comparable to the estimates of the reproduction number obtained from the analysis of a pandemic H1N1 outbreak in the St. Francis Preparatory school in New York [2.4 (95% CI: 1.8 – 3.2)] (Yang et al., 2009b) and to the estimates obtained from the analysis of a high school outbreak in Queens, New York [3.3 (95% CI: 3.0 – 3.6)] (Lessler et al., 2009).

Finally, notice that the estimates of R_s obtained for the school of Villazzano with method L.1 clearly reflect the two exponential growth phases that make this epidemic rather uncommon. The estimates obtained through the fit of the initial data (i.e. from day 7) are lower than the estimates obtained on the time intervals starting at day 19; these last seem much more reasonable to us. A more detailed analysis of the spread of infection among the school-classes could possibly clarify the uncommon behaviour of the epidemic curve observed in the school of Villazzano.

5.5 First results & discussion

The individual-based model developed by Merler and Ajelli (2010) has been recognized as a tool able to successfully reproduce the patterns of spread observed in the population-wide epidemic; here we explored its ability in reproducing school outbreaks. The model's heterogeneity, in terms of proportion of adults and children attending the same school and within school reproduction number, could be certainly improved. Despite this, the individual-based model is a valuable tool for the comparison of the estimates of the within school reproduction number with its

5.5. First results & discussion

“true” value given by theory. The estimates of the exponential growth rate depend on the choice of the exponential growth time-intervals adopted in the fit. Once fixed the most appropriate time-intervals, we find that the estimates of R_s obtained through the estimates of the exponential growth rate using linear regression closely reproduce the theoretical values of the within school reproduction number. The analysis of simulated data show that the least square fit of a linear model to the data performed on the cumulative number $c(t)$ of the cases versus time t (method L.2) is the most stable estimation method among the three exploited here. Moreover, the estimates of R_s computed on the cases observed within the schools do not significantly differ from the estimates obtained from the curve of the cases generated within the schools so that the importation of cases seems not to substantially affect the within school dynamics. The estimate of R_s obtained through the infection tree (i.e. counting the mean number of cases generated by a “typical” index case) amounts to 1.51 (standard deviation 1.83) and slightly overestimates the theoretical value of $R_s = 1.18$. On the basis of our estimates obtained by linear regression, the within school reproduction number R_s is lower than the “global” R_0 ; this fact is in conflict with some estimates obtained during the past 2009-2010 H1N1 pandemic (Lessler et al., 2009; Nishiura et al., 2009, 2010).

We compared the simulated results with two real within school outbreaks occurred in Italy during the 2009-2010 H1N1 pandemic. On the basis of the data collected through the survey conducted in the primary schools of Povo and Villazzano, the estimated within school reproduction number is into the range 1.25 – 1.35 for the Povo school and into the range 1.7–2.1 for the school of Villazzano. These estimates are lower than those obtained from the analysis conducted on the data collected in the St. Francis Preparatory school in New York [2.4 (95% CI: 1.8 – 3.2)] (Yang et al., 2009b) and in the Queens school, New York [3.3 (95% CI: 3.0 – 3.6)] (Lessler et al., 2009). The reasons that could explain this difference are manifold and could depend on the delay of the survey in respect to the timing of the epidemic, on some specific characteristics of the schools where the outbreaks have been monitored and on country-specific differences in terms of virus transmissibility, for instance.

The estimates of R_s obtained for the schools of Povo and Villazzano are slightly higher than those computed on the simulated Italian schools, which have been obtained for a “global” R_0 of 1.38. Despite the individual-based model could certainly be improved in a variety of different aspects (and indeed a more realistic procedure of school assignment is being developed), we find that even in the current version, the model can satisfactory reproduce within school outbreaks, so that it is

potentially a powerful tool for the simulation and analysis of disease transmission in specific social contexts.

5.5. *First results & discussion*

Bibliography

- Ajelli, M., Merler, S., Pugliese, A., and Rizzo, C. (2010). Model predictions and evaluation of possible control strategies for the 2009 A/H1N1v influenza pandemic in Italy. *Epidemiology and Infection*, First View:1–12 M3 – 10.1017/S0950268810001317.
- Ajelli, M., Goncalves, B., Balcan, D., Colizza, V., Hu, H., Ramasco, J., Merler, S., and Vespignani, A. (2010). Comparing large-scale computational approaches to epidemic modeling: Agent-based versus structured metapopulation models. *BMC Infectious Diseases*, 10(1):190–.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19:716–723.
- Alexander, N., Moyeed, R., and Stander, J. (2000). Spatial modelling of individual-level parasite counts using the negative binomial distribution. *Biostatistics*, 1(4):453–463.
- Anderson, R. and May, R. (1992). *Infectious Diseases in Humans*. Oxford Univ. Press, Oxford.
- Anderson, R. M. and May, R. M. (1991). *Infectious diseases of humans: dynamics and control*. Oxford Univ. Press, Oxford.
- Andreasen, V., Lin, J., and Levin, S. A. (1997-08-15). The dynamics of cocirculating influenza strains conferring partial cross-immunity. *Journal of Mathematical Biology*, 35(7):825–842.
- Andreasen, V. and Pugliese, A. (1995). Pathogen coexistence induced by density dependent host mortality. *J. theor. Biol.*, 177:159–165.

BIBLIOGRAPHY

- Austin, D., Kakehashi, M., and Anderson, R. (1997). The transmission dynamics of antibiotic-resistant bacteria: the relationship between resistance in commensal organisms and antibiotic consumption. *Proceedings of the Royal Society of London. Series B: Biological sciences*, 246(1388):1629–1638.
- Baccam, P., Beauchemin, C., Macken, C. A., Hayden, F. G., and Perelson, A. S. (2006). Kinetics of influenza a virus infection in humans. *J. Virol.*, 80(15):7590–7599.
- Baguelin, M., Hoek, A. J. V., Jit, M., Flasche, S., White, P. J., and Edmunds, W. J. (2010). Vaccination against pandemic influenza A/H1N1v in England: A real-time economic evaluation. *Vaccine*, 28(12):2370–2384.
- Bailey, N. T. J. (1975). *The mathematical theory of infectious diseases and its applications*. Griffin, London.
- Balcan, D., Colizza, V., Goncalves, B., Hu, H., Ramasco, J. J., and Vespignani, A. (2009). Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the National Academy of Sciences*, 106(51):21484–21489.
- Ball, F., Mollison, D., and Scalia-Tomba, G. (1997). Epidemics with two levels of mixing. *Annals Appl. Prob.*, 7:46–89.
- Ball, F. and Neal, P. (2002). A general model for stochastic sir epidemics with two levels of mixing. *Mathematical Biosciences*, 180(1-2):73–102.
- Becker, N. G. (1989). *Analysis of Infectious Disease Data*. Chapman and Hall, London, first edition.
- Becker, N. G. (1995). Statistical challenges of epidemic data. In Mollison, D., editor, *Epidemic Models: Their structure and Relation to Data*. Cambridge University Press.
- Blower, S. and McLean, A. (1994). Prophylactic vaccine, risk behavior change, and the probability of eradicating HIV in San Francisco. *Science*, 265- No. 5177:1451–1454.
- Boender, G. J., Hagenaars, T. J., Bouma, A., Nodelijk, G., Elbers, A. R. W., de Jong, M. C. M., and van Boven, M. (2007). Risk maps for the spread of highly pathogenic avian influenza in poultry. *Plos Computational Biology*, 3(4):704–712.

- Bouma, A., Elbers, A. R. W., Dekker, A., de Koeijer, A., Bartels, C., Vellema, P., van der Wal, P., van Rooij, E. M. A., Pluimers, F. H., and de Jong, M. C. M. (2003). The foot-and-mouth disease epidemic in the Netherlands in 2001. *Preventive Veterinary Medicine*, 57(3):155–166.
- Boelle, P., Bernillon, P., and Desenclos, J. (2009). A preliminary estimation of the reproduction ratio for new influenza A(H1N1) from the outbreak in Mexico, March-April 2009. *Euro Surveill.*, 14(19).
- Bremermann, H. and Thieme, H. (1989). A competitive exclusion principle for pathogen virulence. *J. Math. Biol.*, 27:179–190.
- Britton, N. (2003). *Essential Mathematical Biology*. Springer.
- Busani, L., Valsecchi, M., Rossi, E., Toson, M., Ferré, N., Dalla Pozza, M., and Marangon, S. (2009). Risk factors for highly pathogenic H7N1 avian influenza virus infection in poultry during the 1999-2000 epidemic in Italy. *The Veterinary Journal*, 181(2):171–177.
- Capua, I. and Marangon, S. (2000). The avian influenza epidemic in Italy, 1999-2000: a review. *Avian Pathology*, 29:289–294.
- Cauchemez, S., Donnelly, C. A., Reed, C., Ghani, A. C., Fraser, C., Kent, C. K., Finelli, L., and Ferguson, N. M. (2009a). Household transmission of 2009 pandemic influenza A (H1N1) virus in the United States. *New England Journal of Medicine*, 361(27):2619–2627.
- Cauchemez, S., Ferguson, N., Wachtel, C., Tegnell, A., Saour, G., Duncan, B., and Nicoll, A. (2009b). Closure of schools during an influenza pandemic. *Lancet Infectious Diseases*, 9:473–481.
- Cauchemez, S. and Ferguson, N. M. (2008). Likelihood-based estimation of continuous-time epidemic models from time-series data: application to measles transmission in London. *Journal of The Royal Society Interface*, 5(25):885–897.
- Cauchemez, S., Valleron, A. J., Boelle, P. Y., Flahault, A., and Ferguson, N. M. (2008). Estimating the impact of school closure on influenza transmission from sentinel data. *Nature*, 452.
- Cauchemez, S., Bhattarai, A., Marchbanks, T.L., Fagan, R.P., Ostroff, S., Ferguson, N.M. Swerdlow D. and the Pennsylvania H1N1 working group (2010). Role

BIBLIOGRAPHY

- of social networks in shaping disease transmission during a community outbreak of 2009 H1N1 pandemic influenza *PNAS Early Edition*, 1-15
- Chis Ster, I. and Ferguson, N. M. (2007). Transmission parameters of the 2001 foot and mouth epidemic in Great Britain. *PLoS ONE*, 2(6):e502–.
- Chowell, G., Ammon, C. E., Hengartner, N. W., and Hyman, J. M. (2006). Transmission dynamics of the great influenza pandemic of 1918 in Geneva, Switzerland: assessing the effects of hypothetical interventions. *J. Theor. Biol.*, 241:193–204.
- Chowell, G., Fenimore, P. W., Castillo-Garsow, M. A., and Castillo-Chavez, C. (2003). SARS outbreaks in Ontario, Hong Kong and Singapore: the role of diagnosis and isolation as a control mechanism. *Journal of Theoretical Biology*, 224(1):1–8.
- Chowell, G., Hengartner, N. W., Castillo-Chavez, C., Fenimore, P. W., and Hyman, J. M. (2004). The basic reproductive number of Ebola and the effects of public health measures: The cases of Congo and Uganda. *J. Theor. Biol.*, 229:119–126.
- Chowell, G., Miller, M. A., and Viboud, C. (2007a). Seasonal influenza in the United States, France, and Australia: transmission and prospects for control. *Epidemiology and Infection*, to appear.
- Chowell, G., Nishiura, H., and Bettencourt, L. (2007b). Comparative estimation of the reproduction number for pandemic influenza from daily case notification data. *Journal of the Royal Society Interface*, 22:155–166.
- Circolare del Ministero del lavoro, della salute e delle politiche sociali - 05 ottobre 2009 <http://www.normativasantitaria.it/normasan-pdf/0000/30400-1.pdf>
- Colizza, V., Barrat, A., Barthelemy, M., Valleron, A., and Vespignani, A. (2007). Modeling the worldwide spread of pandemic influenza: Baseline case and containment interventions. *PLoS Medicine*, 4:95–110.
- Cooper, B. S., Medley, G. F., and Scott, G. M. (1999). Preliminary analysis of the transmission dynamics of nosocomial infections: stochastic and management effects. *Journal of Hospital Infection*, 43(2):131–147.

- Daszak, P., Cunningham, A. A., and Hyatt, A. D. (2000). Emerging infectious diseases of wildlife— threats to biodiversity and human health. *Science*, 287(5452):443–449.
- Diekmann, O. and Heesterbeek, J. (2000). *Mathematical Epidemiology of Infectious Diseases*. Wiley, Chichester, New York.
- Dodd, P. and Ferguson, N. (2007). Approximate disease dynamics in household-structured populations. *Journal of The Royal Society Interface*, 4(17):1103–1106.
- Favier, C., Degallier, N., Rosa-Freitas, M. G., Boulanger, J. P., Lima, J. R. C., Luitgards-Moura, J. F., Menkes, C. E., Mondet, B., Oliveira, C., Weimann, E. T. S., and Tsouris, P. (2006). Early determination of the reproductive number for vector-borne diseases: the case of dengue in Brazil. *Tropical Medicine & International Health*, 11:332–340.
- Ferguson, N., Cummings, D., Fraser, C., Cajka, J., Cooley, P., and Burke, D. (2006). Strategies for mitigating an influenza pandemic. *Nature*, 442:448–452.
- Ferguson, N., Donnelly, C. A., and Anderson, R. M. (1999). Transmission dynamics and epidemiology of dengue: insights from age-stratified sero-prevalence surveys. *Philos Trans R Soc Lond B Biol Sci.*, 29; 354(1384)::757–768.
- Ferguson, N. M., Cummings, D. A. T., Cauchemez, S., Fraser, C., Riley, S., Meeyai, A., Iamsirithaworn, S., and Burke, D. S. (2005). Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature*, 437:209–214.
- Ferguson, N. M., Donnelly, C. A., and Anderson, R. M. (2001). The foot-and-mouth epidemic in great britain: Pattern of spread and impact of interventions. *Science*, 292(5519):1155–1160.
- Flahault, A., Vergu, E., and Boelle, P. (2009). Potential for a global dynamic of influenza A (H1N1). *BMC Infectious Diseases*, 9:129.
- Flahault, A., Vergu, E., Coudeville, L., and Grais, R. F. (2006). Strategies for containing a global influenza pandemic. *Vaccine*, 24:6751–6755.
- Fraser, C. (2007). Estimating individual and household reproduction numbers in an emerging epidemic. *PLoS ONE*, 2(8):e758–.

BIBLIOGRAPHY

- Fraser, C., Donnelly, C. A., Cauchemez, S., Hanage, W. P., Van Kerkhove, M. D., Hollingsworth, T. D., Griffin, J., Baggaley, R. F., Jenkins, H. E., Lyons, E. J., Jombart, T., Hinsley, W. R., Grassly, N. C., Balloux, F., Ghani, A. C., Ferguson, N. M., Rambaut, A., Pybus, O. G., Lopez-Gatell, H., Alpuche-Aranda, C. M., Chapela, I. B., Zavala, E. P., Guevara, D. M. E., Checchi, F., Garcia, E., Hugonnet, S., Roth, C., and Collaboration, T. W. R. P. A. (2009). Pandemic potential of a strain of influenza A (H1N1): Early findings. *Science*, 324(5934):1557–1561.
- Fraser, C., Riley, S., Anderson, R. M., and Ferguson, N. M. (2004). Factors that make an infectious disease outbreak controllable. *PNAS*, 101:6146–6151.
- Germann, T., Kadau, K., Longini, I. M., and Macken, C. A. (2006). Mitigation strategies for pandemic influenza in the United States. *PNAS*, 103:5935–5940.
- Ghani, A., Baguelin, M., Griffin, J., Flasche, S., Pebody, R., van Hoek, A., Cauchemez, S., Hall, I., Donnelly, C., Robertson, C., White, M., Barrass, I., Fraser, C., Bermingham, A., Truscott, J., Ellis, J., Jenkins, H., Kafatos, G., Garske, T., Harris, R., McMenamin, J., Hawkins, C., Phin, N., Charlett, A., Zambon, M., Edmonds, W., Catchpole, M., Leach, S., White, P., Ferguson, N., and Cooper, B. (2009). The early transmission dynamics of H1N1pdm influenza in the United Kingdom. *PLoS Curr Influenza*, RRN1130.
- Gilks, W., Richardson, S., and Spiegelhalter, D. (1996). *Markov Chain Monte Carlo in Practice*. Chapman & Hall/CRC.
- Gog, J. R. and Grenfell, B. T. (2002). Dynamics and selection of many-strain pathogens. *Proceedings of the National Academy of Sciences of the United States of America*, 99(26):17209–17214.
- Gonzalez, M. C., Hidalgo, C. A., and Barabasi, A.-L. (2008). Understanding individual human mobility patterns. *Nature*, 453(7196):779–782.
- Grais, R., Hugh Ellis, J., and Glass, G. (2003). Assessing the impact of airline travel on the geographic spread of pandemic influenza. *Eur J Empidemiol*, 18:1065–1072.
- Grassly, N. C., Fraser, C., Wenger, J., Deshpande, J. M., Sutter, R. W., Heymann, D. L., and Aylward, R. B. (2006). New strategies for the elimination of polio from India. *Science*, 314(5802):1150–1153.

- Green, D., Kiss, I., and Kao, R. (2006). Modelling the initial spread of foot-and-mouth disease through animal movements. *Proceedings of the Royal Society B: Biological Sciences*, 273(1602):2729–2735.
- Gupta, S., Ferguson, N., and Anderson, R. (1998). Chaos, persistence, and evolution of strain structures in antigenically diverse infectious agents. *Science.*, 280:912.
- Gupta, S., Hill, A. V., Kwiatkowski, D., Greenwood, A. M., Greenwood, B. M., and Day, K. P. (1994). Parasite virulence and disease patterns in plasmodium falciparum malaria. *Proceedings of the National Academy of Sciences of the United States of America*, 91(9):3715–3719.
- Hall, R. and Becker, N. G. (1996). Preventing epidemics in a community of households. *Epidemiology and Infection*, 117:443–455.
- Halloran, M., Haber, M., and Longini I.M., j. (1992). Interpretation and estimation of vaccine efficacy under heterogeneity. *American J. Epidem.*, 136:328–343.
- Halloran, M. E., Hayden, F. G., Yang, Y., Longini, I. M., and Monto, A. S. (2007). Antiviral effects on influenza viral transmission and pathogenicity: Observations from household-based trials. *American Journal of Epidemiology*, 165(2):212–221.
- Halvorson, D. and Karunakaran, D. (1980). Avian Influenza in caged laying chickens. *Avian Diseases*, 24:288–294.
- Hamer, W. H. (1906). Epidemic disease in England. *Lancet*, 1:733–739.
- Haydon, D. T., Chase-Topping, M., Shaw, D. J., Matthews, L., Friar, J. K., Wilesmith, J., and Woolhouse, M. E. J. (2003). The construction and analysis of epidemic trees with reference to the 2001 UK foot-and-mouth outbreak. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270(1511):121–127.
- Hens, N., Ayele, G., Goeyvaerts, N., Aerts, M., Mossong, J., Edmunds, J., and Beutels, P. (2009a). Estimating the impact of school closure on social mixing behaviour and the transmission of close contact infections in eight european countries. *BMC Infectious Diseases*, 9(1):187–.
- Hens, N., Goeyvaerts, N., Aerts, M., Shkedy, Z., Van Damme, P., and Beutels, P. (2009b). Mining social mixing patterns for infectious disease models based on a two-day population survey in belgium. *BMC Infectious Diseases*, 9(1):5–.

BIBLIOGRAPHY

- Henzler, D. J., Kradel, D. C., Davison, S., Ziegler, A. F., Singletary, D., DeBok, P., Castro, A. E., Lu, H., Eckroade, R., Swayne, D., Lagoda, W., Schmucker, B., and Nesselrodt, A. (2003). Epidemiology, Production Losses, and Control Measures Associated with an Outbreak of Avian Influenza Subtype H7N2 in Pennsylvania (1996-98). *Avian Diseases*, 47(s3):1022–1036.
- Hethcote, H. W. (2000). The mathematics of infectious diseases. *Siam Review*, 42(4):599–653.
- Hollingsworth, T. D., Ferguson, N. M., and Anderson, R. M. (2006). Will travel restrictions control the international spread of pandemic influenza? *Nat Med*, 12(5):497–499.
- House, T. and Keeling, M. (2009). Household structure and infectious disease transmission. *Epidemiology and Infection*, 137:654–661.
- Iannelli, M., Martcheva, M., and Li, X.-Z. (2005). Strain replacement in an epidemic model with super-infection and perfect vaccination. *Mathematical Biosciences*, 195(1):23–46.
- Kao, R., Danon, L., Green, D., and Kiss, I. (2006). Demographic structure and pathogen dynamics on the network of livestock movements in great britain. *Proceedings of the Royal Society B: Biological Sciences*, 273(1597):1999–2007.
- Keeling, M. and Rohani, P. (2008). *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press.
- Keeling, M. J. and Eames, K. T. (2005). Networks and epidemic models. *Journal of The Royal Society Interface*, 2(4):295–307.
- Keeling, M. J., Woolhouse, M. E. J., Shaw, D. J., Matthews, L., Chase-Topping, M., Haydon, D. T., Cornell, S. J., Kappey, J., Wilesmith, J., and Grenfell, B. T. (2001). Dynamics of the 2001 UK foot and mouth epidemic: Stochastic dispersal in a heterogeneous landscape. *Science*, 294(5543):813–817.
- Kermack, W. O. and McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London A*, 115:700–721.

- Le Menach, A., Vergu, E., Grais, R. F., Smith, D. L., and Flahault, A. (2006). Key strategies for reducing spread of avian influenza among commercial poultry holdings: lessons for transmission to humans. *Proc. Roy. Soc. B-Biological Sciences*, 273(1600):2467–2475.
- Lee, V., Lye, D., and Wilder-Smith, A. (2009). Combination strategies for pandemic influenza response - a systematic review of mathematical modeling studies. *BMC Medicine*, 7(1):76–.
- Lessler, J., Reich, N. G., and Cummings, D. A. (2009). Outbreak of 2009 pandemic influenza A (H1N1) at a New York City school. *New England Journal of Medicine*, 361(27):2628–2636.
- Lipsitch, M. (1999). Bacterial vaccines and serotype replacement: lessons from haemophilus influenzae and prospects for streptococcus pneumoniae. *Emerging Infectious Diseases*, 5- No. 3:336–345.
- Lipsitch, M., Cohen, T., Cooper, B., Robins, J. M., Ma, S., James, L., Gopalakrishna, G., Chew, S. K., Tan, C. C., Samore, M. H., Fisman, D., and Murray, M. (2003). Transmission dynamics and control of severe acute respiratory syndrome. *Science*, 300:1966–1970.
- Lloyd-Smith, J. O. (2007). Maximum likelihood estimation of the negative binomial dispersion parameter for highly overdispersed data, with applications to infectious diseases. *PLoS ONE*, 2(2):e180–.
- Lloyd-Smith, J. O., Schreiber, S. J., Kopp, P. E., and Getz, W. M. (2005). Super-spreading and the effect of individual variation on disease emergence. *NATURE*, 438(7066):355–359.
- Longini, I. M., Nizam, A., Xu, S., Ungchusak, K., Hanshaoworakul, W., Cummings, D. A. T., and Halloran, M. E. (2005). Containing pandemic influenza at the source. *Science*, 309:1083–1087.
- Mannelli, A., Busani, L., Toson, M., Bertolini, S., and Marangon, S. (2007). Transmission parameters of highly pathogenic avian influenza (H7N1) among industrial poultry farms in northern Italy in 1999-2000. *Preventive Veterinary Medicine*, 81(4):318–322.

BIBLIOGRAPHY

- Mannelli, A., Ferrè, N., and Marangon, S. (2006). Analysis of the 1999-2000 highly pathogenic avian influenza (H7N1) epidemic in the main poultry-production area in northern Italy. *Preventive Veterinary Medicine*, 73(4):273–285.
- Martcheva, M. (2006). On the mechanism of strain replacement in epidemic models with vaccination. In *Current Developments in mathematical biology. Proceedings of the Conference on Mathematical Biology and Dynamical Systems.* (eds. K. Mahadavi, R. Culshaw & J. Boucher), pp. 149-172. Hackensack, NJ: World Scientific Press.
- Martcheva, M., Bolker, B., and Holt, R. (2008). Vaccine-induced pathogen strain replacement: what are the mechanisms? *Journal of The Royal Society Interface*, 5- No. 18:3–13.
- Massad, E., Coutinho, F., Burattini, M., Lopez, L., and Struchiner, C. (2006). The impact of imperfect vaccines on the evolution of hiv virulence. *Medical Hypotheses*, 66(5):907–911.
- Mathews, J. D., McCaw, C. T., McVernon, J., McBryde, E. S., and McCaw, J. M. (2007). A biological model for influenza transmission: Pandemic planning implications of asymptomatic infection and immunity. *PLoS ONE*, 2(11):e1220–.
- Matthews, L., Haydon, D. T., Shaw, D. J., Chase-Topping, M. E., Keeling, M. J., and Woolhouse, M. E. J. (2003). Neighbourhood control policies and the spread of infectious diseases. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270(1525):1659–1666.
- Matthews, L., McKendrick, I., Ternent, H., Gunn, G., Synge, B., and Woolhouse, M. (2006). Super-shedding cattle and the transmission dynamics of *escherichia coli* O157. *Epidemiology and Infection*, 134(01):131–142 M3 – 10.1017/S0950268805004590.
- McLean, A. (1995). Vaccination, evolution and changes in the efficacy of vaccine - a theoretical framework. *Proceedings of the Royal Society of London Series B-Biological Sciences*, 261:389–393.
- Merler, S. and Ajelli, M. (2010). The role of population heterogeneity and human mobility in the spread of pandemic influenza. *Proceedings of the Royal Society B: Biological Sciences*, 277(1681):557–565.

- Miller, E., Hoschler, K., Hardelid, P., Stanford, E., Andrews, N., and Zambon, M. (2010). Incidence of 2009 pandemic influenza A H1N1 infection in England: a cross-sectional serological study. *The Lancet*, 375(9720):1100–1108.
- Minayev, P. and Ferguson, N. (2009). Improving the realism of deterministic multi-strain models: implications for modelling influenza A. *Journal of The Royal Society Interface*, 6(35):509–518.
- Morens, D. M., Folkers, G. K., and Fauci, A. S. (2004). The challenge of emerging and re-emerging infectious diseases. *Nature*, 430(6996):242–249.
- Morris, R. S., Stern, M. W., Stevenson, M. A., Wilesmith, J. W., and Sanson, R. L. (2001). Predictive spatial modelling of alternative control strategies for the foot-and-mouth disease epidemic in Great Britain, 2001. *Veterinary Record*, 149(5):137–144.
- Morse, S. (1995). Factors in the emergence of infectious diseases. *Emerg Infect Dis.*, 1(1):7–15.
- Mossong, J., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R., Massari, M., Salmaso, S., Tomba, G. S., Wallinga, J., Heijne, J., Sadkowska-Todys, M., Rosinska, M., and Edmunds, W. J. (2008). Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLOS Medicine*, 5:381–391.
- Murray, J. (2002). *Mathematical Biology I: An Introduction*. Springer-Verlag.
- Neumann, G., Noda, T., and Kawaoka, Y. (2009). Emergence and pandemic potential of swine-origin H1N1 influenza virus. *Nature*, 459(7249):931–939.
- Ninomiya, A., Takada, A., Okazaki, K., Shortridge, K. F., and Kida, H. (2002). Seroepidemiological evidence of avian H4, H5, and H9 influenza A virus transmission to pigs in southeastern China. *Veterinary Microbiology*, 88(2):107–114.
- Nishiura, H., Castillo-Chavez, C., Safan, M., and Chowell, G. (2009). Transmission potential of the new influenza A (H1N1) virus and its age-specificity in Japan. *Eurosurveillance*, 14(22).
- Nishiura, H., Chowell, G., Safan, M., and Castillo-Chavez, C. (2010). Pros and cons of estimating the reproduction number from early epidemic growth rate of influenza A (H1N1) 2009. *Theoretical Biology and Medical Modelling*, 7(1).

BIBLIOGRAPHY

- O'Neill, P. D. (2002). A tutorial introduction to bayesian inference for stochastic epidemic models using Markov chain Monte Carlo methods. *Math. Biosci.*, 180:103–114.
- Parham, P. E. and Ferguson, N. M. (2006). Space and contact networks: capturing the locality of disease transmission. *Journal of The Royal Society Interface*, 3(9):483–493.
- Parham, P. E., Singh, B. K., and Ferguson, N. M. (2008). Analytic approximation of spatial epidemic models of foot and mouth disease. *Theoretical Population Biology*, 73(3):349–368.
- Pellis, L., Ferguson, N. M., and Fraser, C. (2009). Threshold parameters for a model of epidemic spread among households and workplaces. *Journal of The Royal Society Interface*, 6(40):979–987.
- Pelupessy, I., Bonten, M. J. M., and Diekmann, O. (2002). How to assess the relative importance of different colonization routes of pathogens within hospital settings. *Proceedings of the National Academy of Sciences of the United States of America*, 99(8):5601–5605.
- Porco, T. C. and Blower, S. (1998). Designing HIV vaccination policies: Subtypes and cross-immunity. *Interfaces*, 23:167–190.
- Porco, T. C. and Blower, S. (2000). HIV vaccines: the effect of the mode of action on the coexistence of HIV subtypes. *Mathematical Population Studies*, 8(2):205–229.
- Press, W., Teukolsky, S., Vetterling, W., and Flannery, B. (1992). *Numerical Recipes in C. The art of scientific computing*. Cambridge University Press.
- Press, W., Teukolsky, S., Vetterling, W., and Flannery, B. (2002). *Numerical Recipes in C++, The Art of Scientific Computing*. Cambridge University Press.
- Pugliese, A. (1990). Population models for diseases with no recovery. *J. Math. Biol.*, 28:65–82.
- Racaniello, V. (2004). Emerging infectious diseases. *J Clin Invest.*, 113(6):796–798.
- Rice, J. A. (2004). *Mathematical Statistics and Data Analysis, Second Edition*. Duxbury Press.

- Riley, S., Fraser, C., Donnelly, C. A., Ghani, A. C., Abu-Raddad, L. J., Hedley, A. J., Leung, G. M., Ho, L., Lam, T., Thach, T. Q., Chau, P., Chan, K., Lo, S., Leung, P., Tsang, T., Ho, W., Lee, K., Lau, E. M. C., Ferguson, N. M., and Anderson, R. M. (2003). Transmission dynamics of the etiological agent of SARS in Hong Kong: Impact of public health interventions. *Science*, 300(5627):1961–1966.
- Rizzo, C., Declich, S., Bella, A., Caporali, M., Lana, S., Pompa, M., Vellucci, L., and Salmaso, S. (2009). Enhanced epidemiological surveillance of influenza a(h1n1)v in italy. *Euro Surveill.*, 9;14(27).
- Rizzo, C., Lunelli, A., Pugliese, A., Bella, A., Manfredi, P., Scalia Tomba, G. P., Iannelli, M., and Ciofi degli Atti, M. L. (2008). Scenarios of diffusion and control of an influenza pandemic in Italy. *Epidemiology and Infection*, 136:1650–1657.
- Rizzo, C., Rota, M. C., Bella, A., Alfonsi, V., Declich, S., Caporali, M. G., Ranghiasi, A., Lapini, G., Piccirella, S., Salmaso, S., and Montomoli, E. (2010). Cross-reactive antibody responses to the 2009 A/H1N1v influenza virus in the italian population in the pre-pandemic period. *Vaccine*, 28(20):3558–3562.
- Robert, C. (1996). *The Bayesian choice: a decision-theoretic motivation*. Springer, New York.
- Roberts, G. O., Gelman, A., and Gilks, W. R. (Feb., 1997). Weak convergence and optimal scaling of random walk metropolis algorithms. *The Annals of Applied Probability*, 7(1):110–120.
- Ross, R. (1916). An application of the theory of probabilities to the study of a priori pathometry. part i. *Proceedings of the Royal Society of London A*, 92:204–203.
- Ross, R. and Hudson, H. (1917a). An application of the theory of probabilities to the study of a priori pathometry. part ii. *Proceedings of the Royal Society of London A*, 93:212–225.
- Ross, R. and Hudson, H. (1917b). An application of the theory of probabilities to the study of a priori pathometry. part iii. *Proceedings of the Royal Society of London A*, 93:225–240.
- Ross, T., Zimmer, S., Burke, D., Crevar, C., Carter, D., Stark, J., Giles, B., Zimmerman, R., Ostroff, S., and Lee, B. (2010). Seroprevalence following the second wave of pandemic 2009 H1N1 influenza. *PLoS Curr Influenza*.

BIBLIOGRAPHY

- Rubin, G., Potts, H., and Michie, S. (2010; 14 (34): 183-266). The impact of communications about swine flu (influenza A H1N1v) on public responses to the outbreak: results from 36 national telephone surveys in the uk. *Health Technol. Assess.*
- Rvachev, L. A. and Longini, I. M. (1985). A mathematical model for the global spread of influenza. *Math. Biosci.*, 75:3–22.
- Sedlmaier, N., Hoppenheidt, K., Krist, H., Lehmann, S., Lang, H., and BÄttner, M. (2009). Generation of avian influenza virus (AIV) contaminated fecal fine particulate matter (PM2.5): Genome and infectivity detection and calculation of immission. *Veterinary Microbiology*, 139(1-2):156–164.
- Shortridge, K., Peiris, J., and Guan, Y. (2003). The next influenza pandemic: lessons from Hong Kong. *Journal of Applied Microbiology*, 94:70–79.
- Smith, P., Rodrigues, L., and Fine, P. (1984). Assessment of the protective efficacy of vaccines against common disease using case-control and cohort studies. *International Journal of Epidemiology*, 13:87–93.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal Of The Royal Statistical Society Series B*, 64(4):583–639.
- Stegeman, A. e. a. (2004). Avian Influenza A Virus (H7N1) Epidemic in the Netherlands in 2003: Course of the Epidemic and Effectiveness of Control Measures. *The Journal of Infectious Diseases*, 190:2088–2095.
- Thomas, M., Bouma, A., Ekker, H., Fonken, A., Stegeman, J., and Nielen, M. (2005). Risk factors for the introduction of high pathogenicity Avian Influenza virus into poultry farms during the epidemic in the Netherlands in 2003. *Preventive Veterinary Medicine*, 69(1-2):1–11.
- Tierney, L. (Dec., 1994). Markov chains for exploring posterior distributions. *The Annals of Statistics*, 22(4):1701–1728.
- Tildesley, M. J., Bessell, P. R., Keeling, M. J., and Woolhouse, M. E. J. (2009). The role of pre-emptive culling in the control of foot-and-mouth disease. *Proceedings of the Royal Society B: Biological Sciences*, 276(1671):3239–3248.

- Tumpey, T., Kapczynski, D., and Swyne, D. (2004). Comparative susceptibility of chickens and turkeys to avian influenza a H7N1 virus infection and protective efficacy of a commercial avian influenza H7N2 virus vaccine. *Avian Diseases*, 48:167–176.
- Turchin, P. (2003). *Complex population dynamics. A theoretical empirical synthesis*. Princeton University Press.
- Van der Goot, J., De Jong, M., Koch, G., and Van Boven, M. (2003). Comparison of the transmission characteristics of low and high pathogenicity avian influenza a virus (H5N2). *Epidemiology and Infection*, 131:1003–1013.
- Wallinga, J. and Lipsitch, M. (2007). How generation intervals shape the relationship between growth rates and reproductive numbers. *Proceedings of the Royal Society B: Biological Sciences*, 274(1609):599–604.
- Walsh, B. (2004). Markov chain monte carlo and gibbs sampling. Lecture Notes for EEB 581.
- Wang, W. and Ruan, S. (2004). Simulating the sars outbreak in Beijing with limited data. *Journal of Theoretical Biology*, 227(3):369–379.
- Webster, R. G. (2001). Virology: Enhanced: A molecular whodunit. *Science*, 293(5536):1773–1775.
- Webster, R. G., Bean, W. J., Gorman, O. T., Chambers, T. M., and Kawaoka, Y. (1992). Evolution and ecology of influenza A viruses. *Microbiol. Mol. Biol. Rev.*, 56(1):152–179.
- Woolhouse, M. and Gowtage-Sequeria, S. (2005). Host range and emerging and reemerging pathogens. *Emerging Infectious Diseases*, 11(12):1842–1847.
- Woolhouse, M. E., Haydon, D. T., and Antia, R. (2005). Emerging pathogens: the epidemiology and evolution of species jumps. *Trends in Ecology & Evolution*, 20(5):238–244.
- Woolhouse, M. E. J. (2002). Population biology of emerging and re-emerging pathogens. *Trends in Microbiology*, 10(10):s3–s7.
- Wu, J. e. a. (2010). School closure and mitigation of pandemic (H1N1) 2009, Hong Kong. *Emerging Infectious Diseases*, 16:538–541.

BIBLIOGRAPHY

- Wu, J. T., Riley, S., Fraser, C., and Leung, G. M. (2006). Reducing the impact of the next influenza pandemic using household-based public health interventions. *PLoS Med*, 3(9):e361–.
- Yang, Y., Halloran, M. E., and Longini, Ira M., J. (2009a). A bayesian model for evaluating influenza antiviral efficacy in household studies with asymptomatic infections. *Biostat*, 10(2):390–403.
- Yang, Y., Longini, I. M., and Halloran, M. E. (2006). Design and evaluation of prophylactic interventions using infectious disease incidence data from close contact groups. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 55(3):317–330.
- Yang, Y., Sugimoto, J. D., Halloran, M. E., Basta, N. E., Chao, D. L., Matrajt, L., Potter, G., Kenah, E., and Longini, I. M. (2009b). The transmissibility and control of pandemic influenza A (H1N1) virus. *Science*, 326(5953):729–733.
- Zacks, S. (1971). *The theory of statistical inference*. Wiley, New York.