

The role of population heterogeneity and human mobility in the spread of pandemic influenza

Stefano Merler^{1,*} and Marco Ajelli^{1,2}

¹Fondazione Bruno Kessler, Trento, Italy

²Information Engineering and Computer Science Department, University of Trento, Italy

Little is known on how different levels of population heterogeneity and different patterns of human mobility affect the course of pandemic influenza in terms of timing and impact. By employing a large-scale spatially explicit individual-based model, founded on a highly detailed model of the European populations and on a careful analysis of air and railway transportation data, we provide quantitative measures of the influence of such factors at the European scale. Our results show that Europe has to be prepared to face a rapid diffusion of a pandemic influenza, because of the high mobility of the population, resulting in the early importation of the first cases from abroad and highly synchronized local epidemics. The impact of the epidemic in European countries is highly variable because of the marked differences in the sociodemographic structure of European populations. R_0 , cumulative attack rate and peak daily attack rate depend heavily on sociodemographic parameters, such as the size of household groups and the fraction of workers and students in the population.

Keywords: mathematical modelling; infectious diseases; pandemic influenza

1. INTRODUCTION

The spread of an infectious disease epidemic is driven by the interplay of two factors: the transmissibility of the pathogen responsible for the infection and the characteristics of the host population. When the role of host is played by a human population, predicting the spread of an epidemic is difficult owing to the complexity of modern human societies. It is well established that the spatial structure of the population has an impact on the diffusion of an epidemic: measles waves in England and Wales, spreading from large cities to small towns, are determined by the spatial hierarchy of the host population structure (Grenfell *et al.* 2001), and the spatial distribution of farms influences the regional variability of foot-and-mouth outbreaks in United Kingdom (Keeling *et al.* 2001). The heterogeneity of the population itself can play an important role in the spread of an epidemic (Dushoff & Levin 1995). It is also well known that human mobility patterns affect the spatiotemporal dynamics of an epidemic: the role played by the airline transportation network has been analysed in Colizza *et al.* (2006), and it has been shown that the high degree of predictability of the worldwide spread of infectious diseases is caused by the strong heterogeneity of the transport network (Hufnagel *et al.* 2004).

Large-scale individual-based spatially explicit transmission models of infectious diseases (Riley 2007) have become an important tool to evaluate the intervention options for containing (Ferguson *et al.* 2005; Longini *et al.* 2005) or mitigating (Longini *et al.* 2004; Ferguson *et al.* 2006; Germann *et al.* 2006; Ciofi degli Atti *et al.*

2008; Halloran *et al.* 2008; Merler *et al.* 2009) an influenza pandemic. Because of their complexity, these models have been developed only at the country level, also including some European countries (Ferguson *et al.* 2006; Ciofi degli Atti *et al.* 2008).

However, Europe has never been analysed as a whole and thus it is still uncertain how pandemic influenza could spread in Europe. Europe comprises countries characterized by completely different social and economical backgrounds that result in different levels of population heterogeneity, in terms of both socio-demographic structure and mobility. As it is reasonable to assume that the epidemiological characteristics of the virus do not vary among the European countries, we expect that the high variability in the sociodemographic structure of the European countries results in a high variability in the impact of a pandemic influenza in the different European countries. This is the first key issue we want to address. To what extent the heterogeneity of mobility patterns affects the spread in Europe is the second key issue we want to address.

2. MATERIAL AND METHODS

(a) *The model*

We developed a stochastic, spatially explicit, individual-based simulation model of the spread of influenza pandemic in 37 European countries (515 million individuals). In each country, individuals are explicitly represented in the model and can transmit the infection to household members, to school/work colleagues (on the basis of the employment) and in the general population (where the force of infection is assumed to depend explicitly on the geographical distance). The epidemic can spread from one country to another through cross-borders diffusion (individuals living close to the borders can transmit the infection outside the

* Author for correspondence (merler@fbk.eu).

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rspb.2009.1605> or via <http://rspb.royalsocietypublishing.org>.

country by random contacts with individuals living in other countries) and because of international travel (taken by individuals). The infection is continuously sustained in the study area by the importation of cases from countries outside Europe. Sociodemographic data were used to generate highly detailed synthetic populations of individuals in Europe, explicitly grouped in households, schools and workplaces. Data on air and railway transportation data were used to simulate long-distance travel across the countries of the study area and to simulate importation of cases. Details are given in the electronic supplementary material.

(b) Population heterogeneity

By analysing data on the sociodemographic structure of 37 European countries (see figure 1a) provided by the Statistical Office of the European Communities (Eurostat) and integrated with data provided by the National Statistical Offices for countries not covered by Eurostat, we found that the frequencies of household type and size (figure 1b,c), the age structure (figure 1e), the schools size (figure 1d), the rates of school attendance and employment by age (figure 1f) are highly variable across Europe. The age structure of countries like Ireland, which is one of the youngest European countries (with 31% of the population aged less than or equal to 20 years), differs drastically from that of countries like Germany and Italy (where only 22% and 20.5% of the population is aged less than or equal to 20 years, respectively), which are characterized by very low fertility rates (United Nations 2007). This results in largely different frequencies of household type and size. The fraction of households with children ranges from 0.3 in Denmark to 0.6 in Sweden and the average household size ranges from 2.1 in Denmark to three in Cyprus. By restricting our attention to households with children, a large variability in the number of children per household is also observable (see electronic supplementary material, figure S1), with countries such as Ireland and Cyprus, where households have several children, opposite to countries like Germany and Bulgaria. We have also observed a large difference in terms of employment rates in the population aged over 15 years (the legal working age in Europe is 15 or 16, with some exceptions): the fraction of workers ranges from 0.39 in Bulgaria to 0.67 in Lichtenstein. The fraction of students in the population aged over 15 years ranges from 0.04 in Denmark to 0.12 in Cyprus. According to the PIRLS 2001 and PISA 2000 and 2003 international surveys, as elaborated in European Commission (2005), the average size of primary schools ranges from 200 to 750 and the average size of secondary schools ranges from 270 to 1000. We used an independent dataset providing information on all the Italian schools to validate the surveys data (see electronic supplementary material, figure S2). Data concerning workplaces in Italy and United Kingdom do not highlight significant differences in the size of workplaces (see electronic supplementary material, figure S3). We used the sociodemographic data described above to generate a highly detailed synthetic population of individuals, explicitly grouped in households, schools and workplaces, for simulating the populations in the different countries of the study area. Details on the analysis of the European sociodemographic structure are given in the electronic supplementary material.

(c) Human mobility

We analysed air and railway transportation data as provided by Eurostat. We found that in 2007, more than 360 million passengers took international trips across EU27

(see figure 1a), 323 million of whom travelled by airplane and 37 million travelled by train. In the same year, more than 135 million passengers entered EU27 from countries outside EU27. The great majority of these trips are from and to the Western part of Europe (see figure 2a,b), namely United Kingdom, Germany, France, Italy and Spain (about 85% of the trips were from and to these countries). The probability density function of travel distances is shown in figure 2c. As shown in Matsumoto (2004) and Grosche *et al.* (2007), international travel flows are related to economic factors. By considering only the travels across EU27, we found that the flow from country i to country j can be explained by a gravity model depending on the gross domestic product (GDP) per capita, the population and the distance: $F_{ij} = \theta(g_i^{\tau_f} g_j^{\tau_r}) / (d_{ij}^{\rho})$, where g_i is a normalized GDP of country i ($g_i = p_i G_i / G^*$ where G_i is the GDP per capita of country i , G^* is the average GDP per capita of EU27 and p_i is the population of country i) and d_{ij} is the distance between the two countries. τ_f and τ_r tune the dependence of dispersal on donor and recipient sizes, ρ tunes the dependence on the distance and θ is a proportionality constant. To show this, we generated a synthetic population of travellers travelling according to a gravity model whose masses are given by the normalized GDPs (model A), by the population sizes, as in Viboud *et al.* (2006) (model B), and travelling by choosing a random destination (model C). We found that model A explains the origin–destination matrix (figure 2d) and the distance distribution (figure 2e) better than models B and C (see electronic supplementary material, figure S5). This considered, we used model A for simulating long-distance trips across the study area. As for the internal commuting, i.e. daily trips to school and workplace, we adopted the following procedure. First, schools and workplaces of the proper size were spatially distributed proportionally to the population (see electronic supplementary material, figure S2). Afterward, students and workers were randomly assigned to a school or a workplace, in such a way that the resulting distance to school/work distribution complies with a truncated power-law distribution (see figure 2f), as proposed in Gonzalez *et al.* (2008) for the radius of gyration of mobile phone users, extending the precursor work presented in Brockmann *et al.* (2006) on the circulation of banknotes in the United States. Electronic supplementary material, figure S6 shows how well the proposed model of internal commuting compares with a gravity model previously developed on Italian commuting data. Details on the analysis of human mobility patterns are given in the electronic supplementary material.

3. RESULTS AND DISCUSSION

The transmission rates, defined as the product of the contacts rate times the probability of transmitting the infection, of a new influenza pandemic are unknown. By looking at past pandemics, we can only make assumptions on its transmissibility potential, which can be summarized by the reproductive number R_0 (essentially, the number of secondary infections that results from a single infectious individual in a fully susceptible population (Anderson & May 1992)). Therefore, according to the recent estimates of the reproductive number for recent influenza pandemics (Mills *et al.* 2004; Ferguson *et al.* 2005; Chowell & Nishiura 2008), plausible transmissibility scenarios on R_0 are drawn: the investigated values range from 1.6 to 2.4.

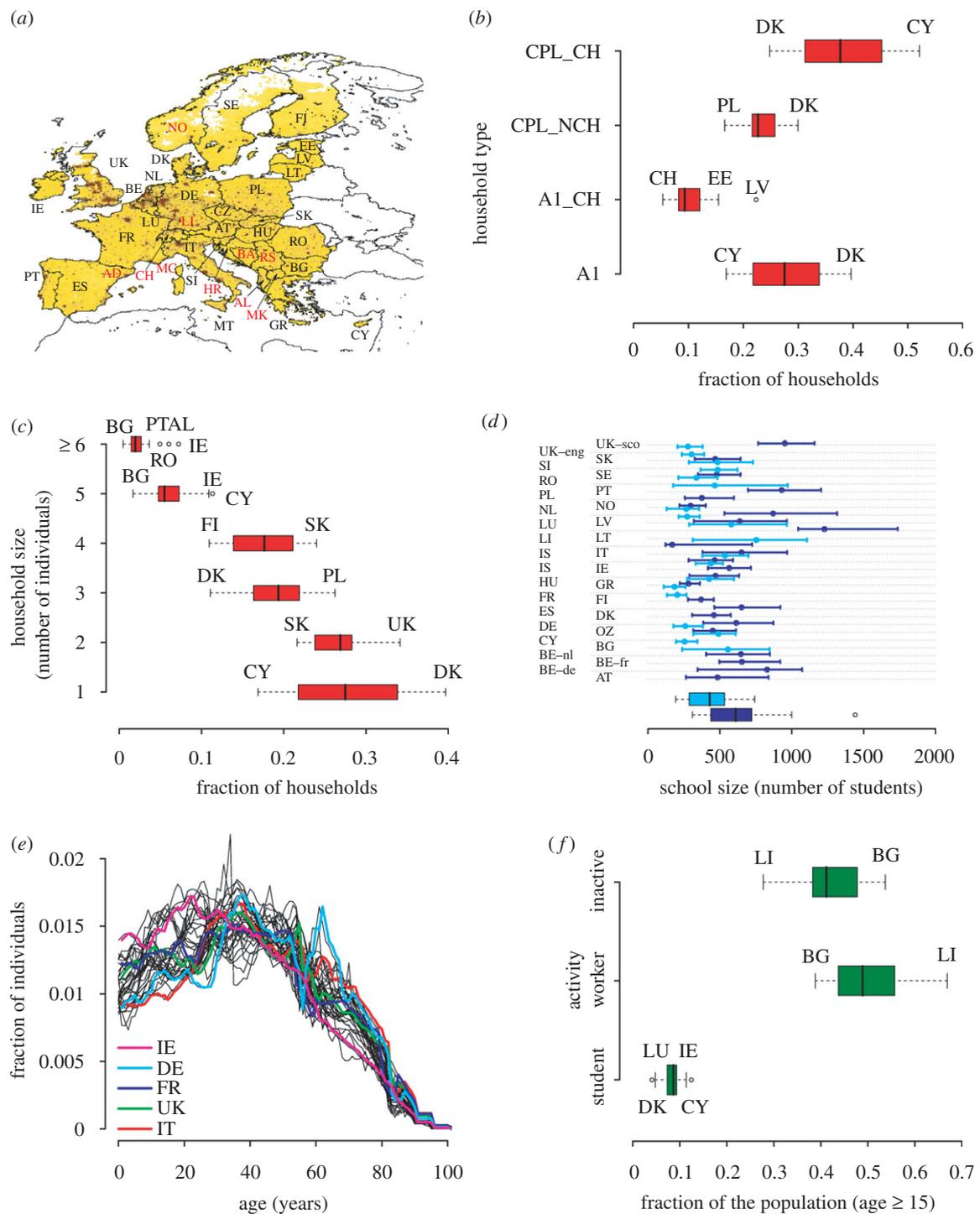


Figure 1. Sociodemographic structures. (a) The study area includes 37 countries and accounts for about 515 million individuals (details are provided in table S1, electronic supplementary material). Colours from yellow to brown indicate the increasing density of population. Black labels refer to countries belonging to EU27 while red labels refer to countries that do not belong to EU27. (b) Variability in the frequencies of household type at European level. A1 represents single persons, A1_CH single parents with children, CPL_NCH couples without children, CPL_CH couples with children. More than 95 per cent of European households are structured as one of the four abovementioned types. (c) Variability in the frequencies of household size. (d) Variability in schools size (primary schools in cyan, secondary schools in blue). Horizontal lines identify the percentiles 25 and 75, the points represent the median values. The two boxplots represent the distributions of the average school size. (e) Variability in the age structure. (f) Variability in the employment and school attendance rates. Only individuals aged over 15 years are considered. In the model, individuals aged 15 years or younger are assumed to attend schools.

Moreover, following Ferguson *et al.* (2006), the model is parametrized so that in the United Kingdom 30 per cent of transmission occurs in households, 37 per cent in schools and workplaces and 33 per cent in the general community. As the contact rates are determined by the

sociodemographic structure of the population, we are somehow setting the probability of transmitting the infection in the different social contexts. After having parametrized the model in the United Kingdom (i.e. by simulating an epidemic spreading in the United Kingdom

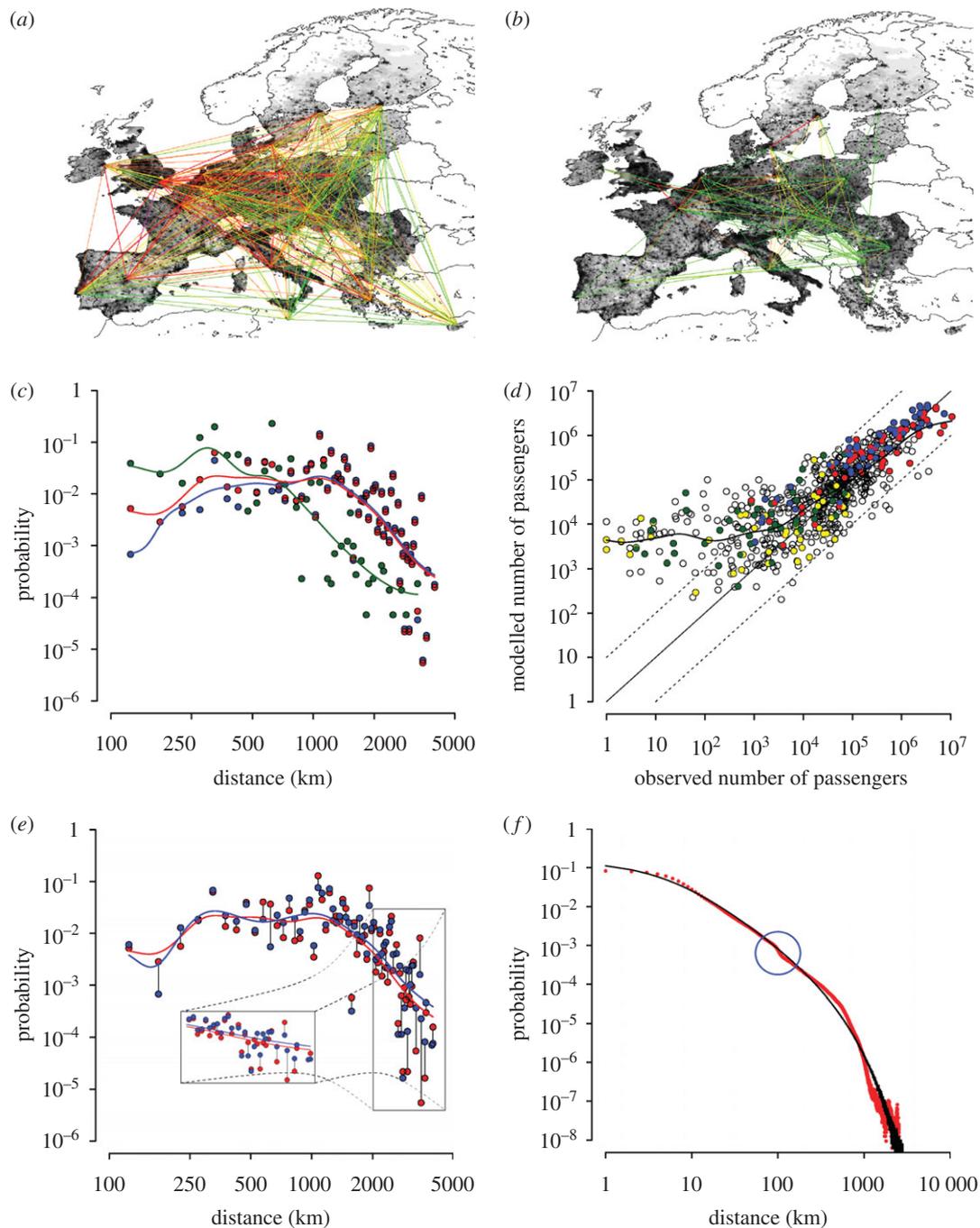


Figure 2. Population movement patterns. (a) Network of yearly airplane travellers across Europe (colours are defined as follows: green, less than 10 000 travellers; yellow, 10 000–100 000 travellers; orange, 100 000–1 000 000 travellers; red, 1 000 000–10 000 000 travellers; purple, more than 10 000 000 travellers). Each link between two countries is identified by an arc connecting the two capitals. (b) as (a) but for train travellers. (c) Probability density function of travel distances by train (green points), by airplane (blue points) and total (red points). Solid lines represent the smooth interpolations of data. (d) Model A (described in the main text; parameters: $\tau_f = 0.53$, $\tau_r = 1.18$ and $\rho = 0.13$): comparison between the observed and the modelled origin–destination matrix. Points compare the generic entries of the two matrix and the solid black line represents a smooth interpolation. The model tends to overestimate the number of travellers when the actual yearly number of travellers is less than 1000; it is in good agreement with the data on the most important links. Red circle, UK; blue circle, DE; green circle, SI; and yellow circle, MA. (e) Model A: resulting probability density function of travel distances compared with that resulting from the analysis of the observed data (shown in c) (red points) and blue points show the model. (f) Internal commuting: probability density function of travel distances to school/workplace (in the model, red points), compared with that proposed in Gonzalez *et al.* (2008) for the radius of gyration of mobile phone users (black points). In the model, students are assumed to attend schools no more than 100 km from home. This results in a change in the slope of the probability curve (blue circle). Solid line, $p(r_g) = (r_g + r_g^0) \exp(-r_g/k)$.

without considering connections with the rest of the world), the same transmission rates are assumed in the rest of the study area. From now on, unless otherwise specified, we assume $R_0 = 2$ (as discussed, it means

$R_0 = 2$ in the United Kingdom and it represents the in-country estimate of R_0). We assume that the latent period is 1.5 days and the infectious period is 2 days. Infected individuals are assumed to have a

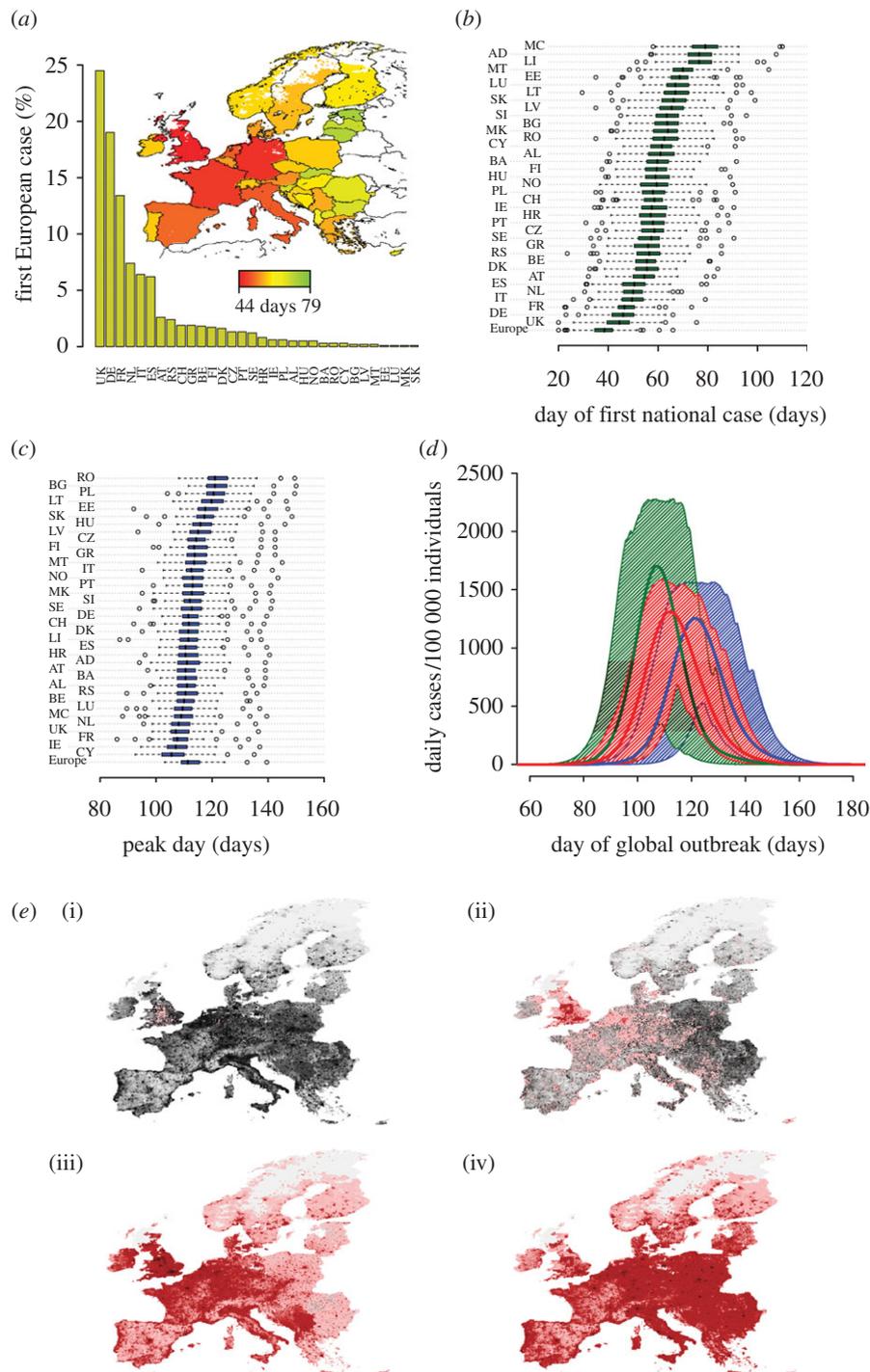


Figure 3. Spatiotemporal dynamics of a new pandemic influenza ($R_0 = 2$). (a) Probable destination of the first case imported in Europe and cartographic representation of the timing of the first national case. (b) Distributions of the day of the first national case (days since the first world case). (c) Distributions of the peak day (days since the first world case). (d) Expected number of daily cases per 100 000 individuals in time in Europe (red line) and 95% confidence intervals (shaded area). Green and blue lines (and shaded areas) refer to the expected number of daily cases per 100 000 individuals in Ireland and Bulgaria, respectively. These two countries are among those where the impact of the epidemic is expected to be the highest and the lowest, respectively. (e) Time sequence (in days) of a simulated epidemic. A single simulation with the first European case in the United Kingdom is shown. Colours from pink to dark red indicate an increasing number of daily cases (dark red indicates more than 10 000 daily cases). (i) Day 60, (ii) day 80, (iii) day 100 and (iv) day 120.

probability of 0.5 of developing clinical symptoms. In the electronic supplementary material, we provide sensitivity analysis on all the abovementioned modelling choices.

We found that the probability of importing the first case is higher in Western countries (the first case is imported in United Kingdom or Germany in almost

50% of simulations, see figure 3a). The distributions of the timing of the first case differ largely from country to country (see figure 3b, its cartographic representation is shown in figure 3a): on average, the first case occurs 44 (95% CI 30–60) and 79 (95% CI 66–92) days after the first world case in the United Kingdom and the Principality of Monaco, respectively. By ignoring

the less populous countries, a west–east gradient is clearly observable. The variability in the peak day in the different countries (see figure 3c) is less remarkable (on average, it ranges from 106, 95% CI 96–119, days in Cyprus to 122, 95% CI 112–135, in Romania) as long-distance travels tend to synchronize the national epidemics and these are much faster in less populous countries. In general, we have observed that the high mobility inside the countries (internal commuting) and the long-distance travels tend to synchronize, thus fastening, the epidemic. The average peak day in a country is positively correlated with the longitude of the country (Spearman test, $\rho = 0.55$, $p = 0.0003$), as confirmed by the clear spatial trend observable in the time sequence of the simulated epidemic shown in figure 3e (see also the electronic supplementary material, video M2). This finding is supported by the results presented in Paget *et al.* (2007), where a spatial analysis revealed a significant west–east pattern in the timing of peak influenza activity across Europe in four winters from 1999 to 2007. The average peak day in a country is negatively correlated with the yearly number of passengers entering the country from other countries in the study area (Spearman test, $\rho = -0.59$, $p = 0.001$), supporting the hypothesis that the observed pattern of epidemic spread is related to the patterns of human mobility. When the first European case is observed in the United Kingdom, the delay in epidemic onset in other countries is much shorter in the Western part of Europe, compared to the Eastern and small countries (see figure 4a, other scenarios are shown in electronic supplementary material, figure S14). At the national scale, spatial heterogeneity in population density results in a relevant delay in epidemic onset between urban and rural areas (see figure 4b). These last results allow the identification of the top risk areas during the initial phase of the epidemic. The expected pattern of spread in Europe is shown in figure 3d. The epidemic peaks some 110 days after the first world case and the epidemic lasts about 3 months. We remark that figure 3d reports the expected number of new cases in time. As the national epidemics are not synchronized in time, the actual peak incidence will be much higher than the value corresponding to the peak day as reported in the figure (it will be closer to the upper 95% confidence limit, see also figure 5c).

On average, the cumulative attack rate in the different countries ranges from 31.2 per cent (95% CI 31.2–31.2) in Bulgaria to 37.8 per cent (95% CI 37.6–37.9) in Cyprus (see figure 5a, its cartographic representation is shown in figure 5b). By looking at the study area as a whole, the average cumulative attack rate is 33.7 per cent (95% CI 33.7–33.7). Among the most populous countries, the cumulative attack rate is expected to be 31.7 per cent (95% CI 31.7–31.7) in Germany, 33.5 per cent (95% CI 33.5–33.5) in the United Kingdom, 33.5 per cent (95% CI 33.5–33.6) in Italy, 34.6 per cent (95% CI 34.5–34.6) in France and 35.5 per cent (95% CI 35.5–35.6) in Spain. It is worth noting that the value obtained for the United Kingdom is very similar to that obtained in Ferguson *et al.* (2006). The standard deviations of the distributions of the national cumulative attack rates are very small, except for the less populous countries. The average cumulative attack rate in a country is positively correlated

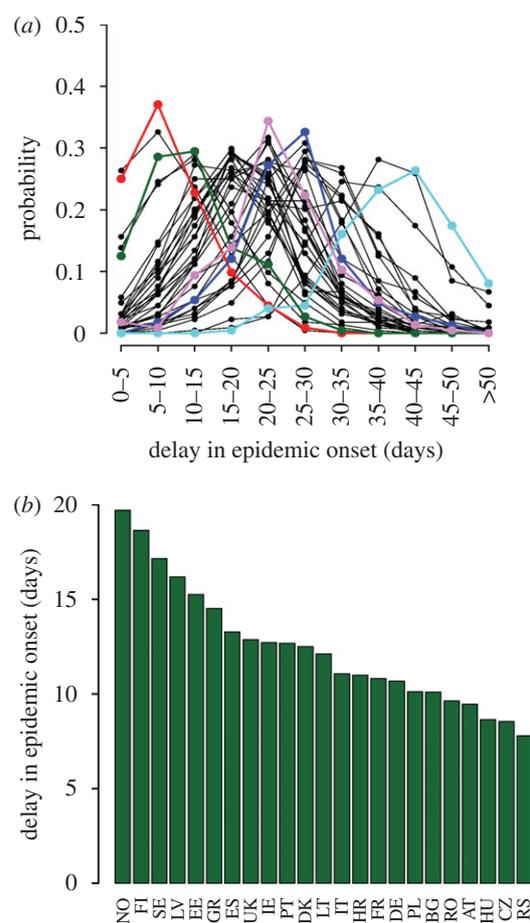


Figure 4. Spatiotemporal dynamics of a new pandemic influenza ($R_0 = 2$). (a) Delay (days since the first European case) in epidemic onset when the first European case is imported in the United Kingdom. Red line, DE; green line, IT; pink line, RO; dark blue line, BG; light blue line, MC. (b) Average delay of epidemic onset in rural areas (population density $< 100 \text{ km}^{-2}$, accounting for 22.6% of the population) with respect to the urban areas (population density $> 400 \text{ km}^{-2}$, accounting for 50.2% of the population) in European countries with area of at least $50\,000 \text{ km}^{-2}$.

with the average household size (Spearman test, $\rho = 0.77$, $p < 0.0001$) (see figure 5d) and with the fraction of students in the population (Spearman test, $\rho = 0.77$, $p < 0.0001$), and negatively correlated with the fraction of inactive individuals in the population (Spearman test, $\rho = -0.38$, $p = 0.02$) (see figure 5h). It is worth noticing that a simple linear regression model whose independent variables are the average household size, the fraction of students and the fraction of inactive individuals in the population predicts very well the average cumulative attack rate in the different countries (coefficient of determination $R^2 = 0.985$, root mean square error (r.m.s.e.) = 0.17). The cumulative attack rate is positively correlated with the basic reproduction number R_0^E (Spearman test, $\rho = 0.8$, $p < 0.0001$, see the inset in figure 5e). Thus, we also observed a high variability in R_0^E , ranging from 1.9 (95% CI 1.63–2.23) in the Principality of Monaco to 2.27 (95% CI 1.99–2.47) in Croatia (see figure 5e). We used the apex E since the above estimates of R_0^E were obtained by analysing the national epidemics as obtained by simulating the spread at European level, and thus additionally considering cases generated through

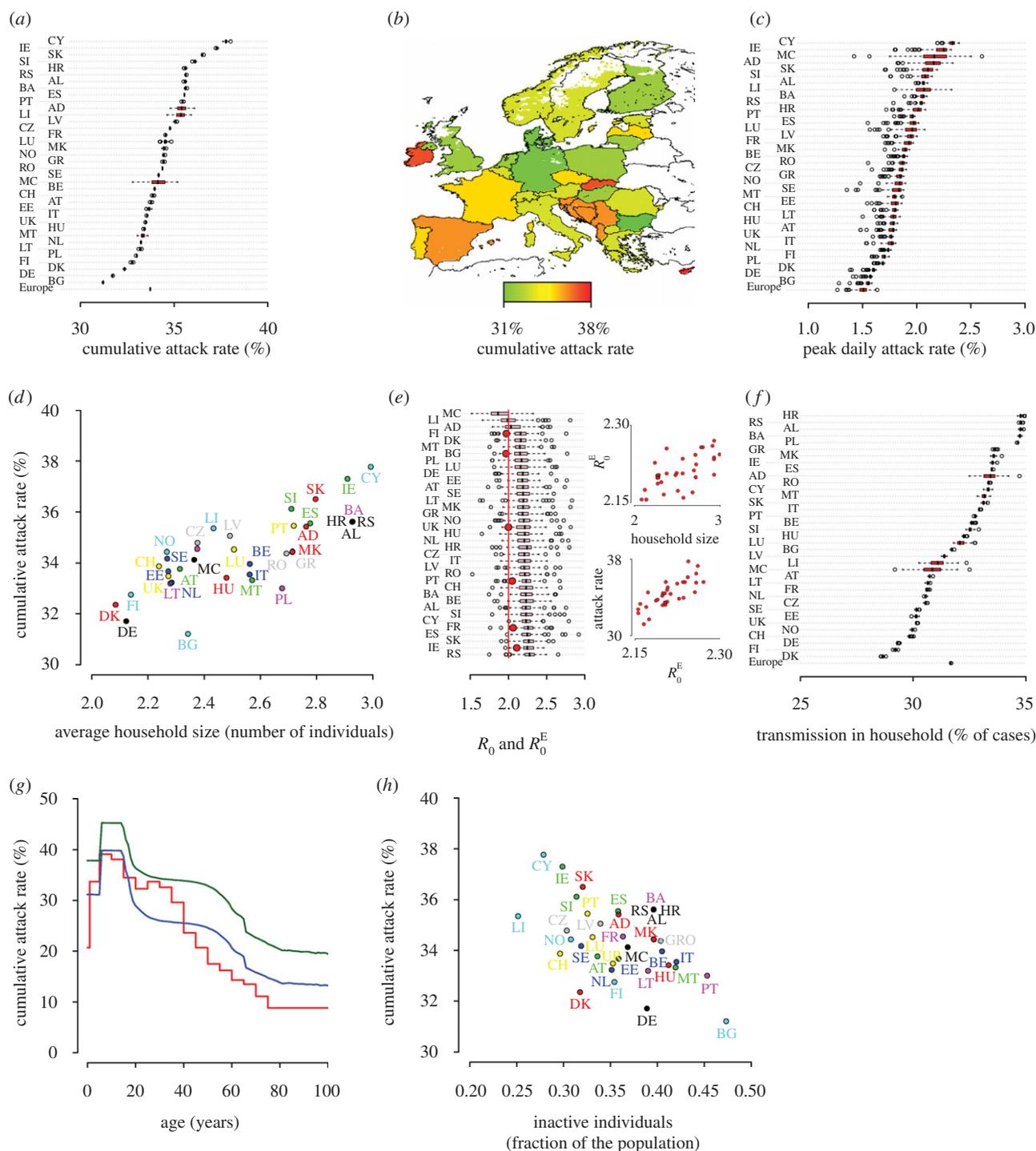


Figure 5. Impact of a new pandemic influenza ($R_0 = 2$). (a) Distributions of the cumulative attack rate. (b) Cartographic representation of the cumulative attack rate. (c) Distributions of the peak daily attack rate. (d) Average cumulative attack rate as a function of the average household size. (e) Distribution of R_0^E (pink boxplots), compared to the in-country estimates of R_0 (red points). The two insets show the relationship between R_0^E , household size and cumulative attack rate. (f) Distributions of the percentage of cases owing to transmission among household members. (g) Cumulative attack rates by age as observed during the 1918–1919 Spanish influenza in the United States (red line) (Glezen 1996) and the average European values as predicted by the model with $R_0 = 1.6$ (blue line) and $R_0 = 2$ (green line). (h) Average cumulative attack rate as a function of the fraction of inactive (neither students nor workers) individuals.

cross-border diffusion and long-distance travels and subsequent secondary infections. Therefore, R_0^E may differ from the in-country estimates (see the electronic supplementary material for details on the computation of the reproduction number). It is worth noticing that we obtained the estimates of R_0^E systematically larger than 2, United Kingdom included. This can be explained by

considering that imported cases give rise to chains of infections not synchronized in time that contribute cumulatively to increase the estimate of R_0^E compared to the in-country estimate (this effect is less relevant in small countries where the number of imported cases is very low). In-country estimates of R_0 are much closer to 2, though variable from country to country, as the

importation of cases is not considered at all (see figure 5e). The peak daily attack rate in the different countries is also highly variable. It ranges from 1.5 per cent (95% CI 1.5–1.6) in Bulgaria to 2.3 per cent (95% CI 2.2–2.4) in Cyprus (see figure 5c). As the national epidemics are not synchronized, the average peak daily attack rate of the whole study is similar to the value observed in Bulgaria, namely 1.5 per cent (95% CI 1.4–1.6). In the United Kingdom we obtained a lower peak daily attack rate, 1.8 per cent (95% CI 1.7–1.8) with respect to the 2.1 per cent as reported in Ferguson *et al.* (2006). This is due to different modelling choices for the infective period. We assume an exponential distribution for both latent and infectious period (as in the classical mathematical models of infectious diseases) and infectiousness is assumed to be constant during the infectious period (2 days). In Ferguson *et al.* (2006) individuals transmit more at the very beginning of the infectious period, giving rise to faster simulated epidemics and to higher peak daily attack rates. While these different modelling choices can affect the evaluation of some containment strategies (e.g. antiviral treatment) and can lead to differences in the timing of the simulated epidemics, they do not affect the results presented in this work. Electronic supplementary material, figure S12 shows the dependence of the timing of the epidemic and of the peak daily incidence on the length of the infectious period: given R_0 , the shorter the infectious period is, the faster the epidemic spreads and the higher the peak daily attack rate is. The cumulative attack rate does not depend on the length of the infectious period. Values of peak day and peak daily attack rate in the United Kingdom similar to that reported in Ferguson *et al.* (2006) were obtained by assuming an infectious period of 1.5 days (see the electronic supplementary material). The effects of varying the probability of developing symptoms is discussed in the electronic supplementary material (see figure S13). We found that the peak daily attack rate in a country is positively correlated with the average household size (Spearman test, $\rho = 0.72$, $p < 0.0001$) and with the fraction of students in the total population (Spearman test, $\rho = 0.79$, $p < 0.0001$), and slightly negatively correlated with the fraction of inactive individuals in the total population (Spearman test, $\rho = -0.31$, $p = 0.06$). Moreover, it is negatively correlated with the number of inhabitants (Spearman test, $\rho = -0.51$, $p = 0.001$) as national epidemics tend to be less spatially synchronized in larger countries. It is also relevant to analyse where transmission occurs. We found that the transmission in households in the different countries ranges from 28.7 per cent (95% CI 28.6–28.7) in Denmark to 34.8 per cent (95% CI 34.7–34.9) in Croatia (see figure 5f). Our assumptions on the level of transmission in the different social contexts, which result in the attack rates by age as shown in figure 5g, are discussed in the electronic supplementary material (see figures S10 and S11).

By examining the results for the values of R_0 in the range 1.6–2.4, we did not find significant qualitative differences (of course, the timing and the impact of the epidemics are drastically different; see figures S8, S9 and the electronic supplementary material, videos M1 and M3). The model was validated by comparing its

predictions to outcomes seen in the ongoing A(H1N1) pandemic (see electronic supplementary material, figure S15) and by comparing the attack rates by age with those observed during the 1918–1919 Spanish influenza in the United States (Glezen 1996) (see electronic supplementary material, figure S16).

4. CONCLUSIONS

Because of the high mobility of the population, resulting in an early importation of the first cases from abroad and highly synchronized local epidemics, European countries have to be prepared to face a rapid diffusion of a pandemic influenza. Importation of the first cases is much more likely to occur in the Western part of Europe, and a stable west–east pattern in the timing of peak influenza activity across Europe can be predicted which is independent of R_0 . The in-country timing of influenza activity is affected by the spatial heterogeneity of population density—epidemic reaches urban areas much earlier than rural areas. Countries characterized by large household groups and by a large fraction of students in the population (notably, Ireland) would face more severe epidemics than countries such as Germany and Bulgaria. In fact, R_0 (whose estimates depend on the importation of cases), cumulative attack rate and peak daily attack rate depend heavily on sociodemographic parameters, such as the size of household groups and the fraction of workers and students in the population, mostly because they affect the contact rates in the population. The proposed model confirms the widely held theories; remarkably, it also gives insight into the quantitative aspects never analysed before at this level of detail (differences and similarities between the proposed and the previously published models are discussed in the electronic supplementary material). Finally, let us remark that by assuming $R_0 = 1.6$, the predicted attack rates by age comply with those observed during the 1918–1919 Spanish influenza in the United States, and that by assuming $R_0 = 1.2$ the predicted time course of the epidemic reflects that of the current A(H1N1) pandemic in Europe.

We thank three anonymous reviewers for their useful comments that certainly contributed to improve this manuscript. We thank the European Union FP7 FLUMODCONT and EPIWORK projects for research funding.

REFERENCES

- Anderson, R. M. & May, R. M. 1992 *Infectious diseases of humans: dynamics and control*. Oxford, UK: Oxford University Press.
- Brockmann, D., Hufnagel, L. & Geisel, T. 2006 The scaling laws of human travel. *Nature* **439**, 462–465. (doi:10.1038/nature04292)
- Chowell, G. & Nishiura, H. 2008 Quantifying the transmission potential of pandemic influenza. *Phys. Life Rev.* **5**, 50–77. (doi:10.1016/j.plev.2007.12.001)
- Ciofi degli Atti, M. L., Merler, S., Rizzo, C., Ajelli, M., Massari, M., Manfredi, P., Furlanello, C., Scalia Tomba, G. & Iannelli, M. 2008 Mitigation measures for pandemic influenza in Italy: an individual based model considering different scenarios. *PLoS ONE* **3**, e1790. (doi:10.1371/journal.pone.0001790)

- Colizza, V., Barrat, A., Barthélemy, M. & Vespignani, A. 2006 The role of the airline transportation network in the prediction and predictability of global epidemics. *Proc. Natl Acad. Sci. USA* **103**, 2015–2020. (doi:10.1073/pnas.0510525103)
- Dushoff, J. & Levin, S. 1995 The effects of population heterogeneity on disease invasion. *Math. Biosci.* **128**, 25–40. (doi:10.1016/0025-5564(94)00065-8)
- European Commission 2005 *Key data on education in Europe 2005*. Luxembourg: Office for Official Publications of the European Communities.
- Ferguson, N. M., Cummings, D. A., Cauchemez, S., Fraser, C., Riley, S., Meeyai, A., Iamsirithaworn, S. & Burke, D. S. 2005 Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature* **437**, 209–214. (doi:10.1038/nature04017)
- Ferguson, N. M., Cummings, D. A., Fraser, C., Cajka, J. C. & Cooley, P. C. 2006 Strategies for mitigating an influenza pandemic. *Nature* **442**, 448–452. (doi:10.1038/nature04795)
- Germann, T. C., Kadau, K., Longini, I. M. J. & Macken, C. A. 2006 Mitigation strategies for pandemic influenza in the United States. *Proc. Natl Acad. Sci. USA* **103**, 5935–5940. (doi:10.1073/pnas.0601266103)
- Glezen, W. 1996 Emerging infections: pandemic influenza. *Epidemiol. Rev.* **18**, 64–76.
- Gonzalez, M., Hidalgo, C. & Barabasi, A.-L. 2008 Understanding individual human mobility patterns. *Nature* **453**, 779–782. (doi:10.1038/nature06958)
- Grenfell, B. T., Bjornstad, O. N. & Kappey, J. 2001 Traveling waves and spatial hierarchies in measles epidemics. *Nature* **414**, 716–723. (doi:10.1038/414716a)
- Grosche, T., Rothlauf, F. & Heinzl, A. 2007 Gravity models for airline passenger volume estimation. *J. Air Transp. Manag.* **13**, 175–183. (doi:10.1016/j.jairtraman.2007.02.001)
- Halloran, M. E. *et al.* 2008 Modeling targeted layered containment of an influenza pandemic in the United States. *Proc. Natl Acad. Sci. USA* **105**, 4639. (doi:10.1073/pnas.0706849105)
- Hufnagel, L., Brockmann, D. & Geisel, T. 2004 Forecast and control of epidemics in a globalized world. *Proc. Natl Acad. Sci. USA* **101**, 15 124–15 129. (doi:10.1073/pnas.0308344101)
- Keeling, M. *et al.* 2001 Dynamics of the 2001 UK foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape. *Science* **294**, 813–817. (doi:10.1126/science.1065973)
- Longini, I. M. J., Halloran, M. E., Nizam, A. & Yang, Y. 2004 Containing pandemic influenza with antiviral agents. *Am. J. Epidemiol.* **159**, 623–633. (doi:10.1093/aje/kwh092)
- Longini, I. M. J., Nizam, A., Xu, S., Ungchusak, K., Hanshaoworakul, W., Cummings, D. A. & Halloran, M. E. 2005 Containing pandemic influenza at the source. *Science* **309**, 1083–1087. (doi:10.1126/science.1115717)
- Matsumoto, H. 2004 International urban systems and air passenger and cargo flows: some calculations. *J. Air Transp. Manag.* **10**, 239–247. (doi:10.1016/j.jairtraman.2004.02.003)
- Merler, S., Ajelli, M. & Rizzo, C. 2009 Age-prioritized use of antivirals during an influenza pandemic. *BMC Infect. Dis.* **9**, 117. (doi:10.1186/1471-2334-9-117)
- Mills, C., Robins, J. & Lipsitch, M. 2004 Transmissibility of 1918 pandemic influenza. *Nature* **432**, 904–906. (doi:10.1038/nature03063)
- Paget, J., Marquet, R., Meijer, A. & van der Velden, K. 2007 Influenza activity in Europe during eight seasons (1997–2007): an evaluation of the indicators used to measure activity and an assessment of the timing, length and course of peak activity (spread) across Europe. *BMC Infect. Dis.* **7**, 141. (doi:10.1186/1471-2334-7-141)
- Riley, S. 2007 Large-scale spatial-transmission models of infectious disease. *Science* **316**, 1298–1301. (doi:10.1126/science.1134695)
- United Nations 2007 *World population prospects—the 2006 revision*. New York: United Nations Publication.
- Viboud, C., Bjornstad, O. N., Smith, D. L., Simonsen, L., Miller, M. A. & Grenfell, B. T. 2006 Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science* **312**, 447–451. (doi:10.1126/science.1125237)