



## Spatiotemporal dynamics of viral hepatitis A in Italy

Marco Ajelli<sup>a,\*</sup>, Laura Fumanelli<sup>a,b</sup>, Piero Manfredi<sup>c</sup>, Stefano Merler<sup>a</sup>

<sup>a</sup> Predictive Models for Biomedicine & Environment, Bruno Kessler Foundation, Italy

<sup>b</sup> Department of Mathematics, University of Trento, Italy

<sup>c</sup> Department of Statistics and Mathematics Applied to Economics, University of Pisa, Italy

### ARTICLE INFO

#### Article history:

Received 1 July 2009

Available online 29 September 2010

#### Keywords:

HAV  
Spatial contact matrix  
Vaccination  
Equilibria  
Stability

### ABSTRACT

Viral hepatitis A is still common in Italy, especially in Southern regions. In this study, a metapopulation model for hepatitis A virus (HAV) transmission is proposed and analyzed. Analytical results on the asymptotic and transient behaviors of the system are carried out. Based on the available Italian movement data, a national spatial contact matrix at the regional level, which could be used for new studies on the transmission dynamics of other infectious diseases, is derived for modeling fluxes of individuals. Despite the small number of fitted parameters, model simulations are in good agreement with the observed average HAV incidence in all regions. Our results suggest that the mass vaccination program introduced in one Italian region only (Puglia, the one with the highest endemicity level) could have played a role in the decline of HAV incidence in the country as a whole. The only notable exception is represented by Campania, a Southern region showing a high endemicity level, which is not substantially affected by HAV dynamics in Puglia. Finally, our results highlight that the continuation of the vaccination campaign in Puglia would have a relevant impact in decreasing long-term HAV prevalence, especially in Southern Italy.

© 2010 Elsevier Inc. All rights reserved.

### 1. Introduction

Viral hepatitis A is an acute liver infectious disease caused by the hepatitis A virus (HAV). The virus can be transmitted from person to person by the oral-fecal route, by ingestion of contaminated food or drink or by the use of intravenous drugs (Stapleton and Lemon, 1994). Hepatitis A is one of the most common infectious diseases worldwide (Das, 2003; Murray and Lopez, 1997), both in developing and developed countries (Lucioni et al., 1998). An effective vaccine is available (Averhoff et al., 2001) and many countries recommend vaccination of children (e.g., the United States). For such reasons, various mathematical modeling studies of HAV have been carried out for evaluating the effectiveness of different control strategies (Sattenspiel and Simon, 1988; Samandari et al., 2004; Srinivasa Rao et al., 2006; Bauch et al., 2007; Ajelli et al., 2008; Ajelli and Merler, 2009).

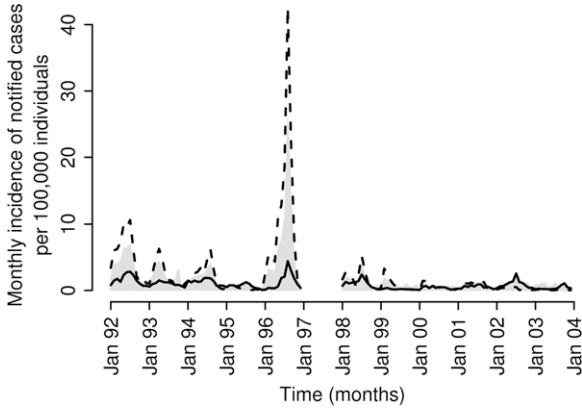
In developed countries the two main sources of HAV infection are represented by direct contacts between individuals and consumption of infected food/water (Fiore, 2004). In Italy, seafood (such as shellfish and mussels consumed raw) represents the

main source of HAV infections (Mele et al., 2006). For this reason, circulation of HAV in the “environment” was explicitly modeled in two recent studies, focused on the transmission dynamics of hepatitis A in Southern Italy (Ajelli et al., 2008) and on the evaluation of different strategies for its control (Ajelli and Merler, 2009). In Italy, HAV seroprevalence (Ansaldi et al., 2008) and incidence of notified cases are highly variable between regions (see Fig. 1). This study focuses on a metapopulation model (whose classes represent different geographical areas) accounting for both person-to-person transmission and ingestion of contaminated food in order to investigate spatiotemporal dynamics of hepatitis A in Italy.

Metapopulation models have been largely used for studying endemic (e.g., Hethcote (1978), Koopman et al. (2002) and Amariet et al. (2008)) and epidemic diseases (e.g., Rvachev and Longini (1985), Sattenspiel and Dietz (1995), Colizza et al. (2007) and Balcan et al. (2009)). A major problem when dealing with this kind of model is to obtain reliable estimates of the mixing level between classes. Here we introduce a spatial contact matrix based on real data and accounting for both occasional long-distance travels (e.g., tourism) and daily short-distance (e.g., commuting between place of residence and of work/study) travels. Such a matrix could also be useful for studying the transmission of other infectious diseases in Italy through metapopulation models. In this study the central role of Puglia in determining hepatitis A dynamics is highlighted as well as the effects on other Italian regions of a mass vaccination program implemented in Puglia.

\* Corresponding address: Predictive Models for Biomedicine & Environment, Bruno Kessler Foundation, Via Sommarive 18, I-38123 Trento Povo, Italy.

E-mail addresses: [ajelli@fbk.eu](mailto:ajelli@fbk.eu) (M. Ajelli), [lfumanelli@fbk.eu](mailto:lfumanelli@fbk.eu) (L. Fumanelli), [manfredi@ec.unipi.it](mailto:manfredi@ec.unipi.it) (P. Manfredi), [merler@fbk.eu](mailto:merler@fbk.eu) (S. Merler).



**Fig. 1.** Monthly incidence (notified cases per 100,000 individuals) in Puglia (dashed black line) and Campania (solid black line). The gray area represents the 95% CI of the monthly incidence in Italy.

## 2. The model

We start by considering the classical SIR model:

$$\begin{cases} S'(t) = -\lambda(t)S(t) - \mu S(t) + bN(t) \\ I'(t) = \lambda(t)S(t) - (\gamma + \mu)I(t) \\ R'(t) = \gamma I(t) - \mu R(t), \end{cases}$$

where  $N(t) = S(t) + I(t) + R(t)$  represents the total population, composed of susceptible ( $S$ ), infectious ( $I$ ) and removed ( $R$ ) individuals;  $\mu$  is the mortality rate,  $b$  is the fertility rate,  $1/\gamma$  is the average infectious period;  $\lambda(t)$  is the force of infection.

Since the Italian population is experiencing a situation of (about) zero growth (ISTAT, 2007), it is reasonable to assume that the mortality and fertility rates coincide (i.e.,  $b = \mu$ ) and, consequently, the total population remains constant (i.e.,  $N(t) = N$ ). However, this modeling choice is not completely realistic since we are not considering immigration and emigration phenomena which could be relevant in this context (Iannelli and Manfredi, 2007).

The two main sources of HAV infection in Italy are direct contacts between individuals and consumption of infected food/water (from now on, we will call the latter the “indirect” contacts component). Therefore, we may write the overall force of infection at time  $t$  as:

$$\lambda(t) = \lambda_1(t) + \lambda_2(t)$$

where  $\lambda_1$  and  $\lambda_2$  account for direct and indirect transmission respectively. As in classical epidemic models,  $\lambda_1(t)$  is assumed to be of the form  $\beta \frac{I(t)}{N}$  (where  $\beta$  is the transmission rate). As regards  $\lambda_2(t)$ , we take the expression  $\lambda_2(t) = \tilde{\beta} \int_{-\infty}^t I(s)G(t-s)ds$  where the delaying kernel  $G$  represents the survival probability of hepatitis A virus in seafood. Assuming that this quantity decays exponentially over time with decay rate  $\delta$ , i.e.  $G(x) = \delta e^{-\delta x}$  with  $x > 0$ ,  $\delta > 0$ , as in Ajelli et al. (2008), we derive the following system:

$$\begin{cases} S'(t) = -\lambda(t)S(t) - \mu S(t) + \mu N \\ I'(t) = \lambda(t)S(t) - (\gamma + \mu)I(t) \\ R'(t) = \gamma I(t) - \mu R(t) \\ A'(t) = \delta [I(t) - A(t)] \end{cases} \quad (1)$$

where  $\lambda(t) = \beta \frac{I(t)}{N} + \tilde{\beta} A(t)$  and  $A$  represents, in a suitable unit, a proxy for the amount of virus circulating in the environment. A full discussion of the modeling assumptions can be found in Ajelli et al. (2008).

Model (1) can be easily extended to  $n$  classes, each of them representing a geographic area (e.g., in the simulations we will

consider the twenty Italian regions). For simplicity we assume that the environment of a region can be contaminated only by the individuals living in the same region. This is not too restrictive, since the average time spent traveling to other regions is short compared to the incubation period of HAV (e.g., on average in Italy 4.3 days a year are spent away from home for touristic purposes, ISTAT (2002), while the HAV incubation period is 2–4 weeks, Stapleton and Lemon (1994) and CDC (2007)).

Let  $\beta_{kj} \geq 0$  be the transmission rate for direct contacts between regions  $k$  and  $j$ , for  $k, j = 1, \dots, n$ ; we assume that at least one of the  $\beta_{kj}$  is strictly positive. Likewise, let  $\tilde{\beta}_{kj} \geq 0$  be the transmission rate for indirect contacts between regions  $k$  and  $j$ , for  $k, j = 1, \dots, n$  (with at least one of the  $\tilde{\beta}_{kj}$  strictly positive). The elements  $\tilde{\beta}_{kj}$  with  $k \neq j$  account for the transmission due to the consumption of infected food during travels outside the region of residence.

Furthermore, we assume that the values of the mortality/fertility rate, the average duration of infection of an individual and the decay rate of the virus in the environment may vary from region to region: thus we take  $\mu_k$ ,  $\gamma_k$  and  $\delta_k$  with  $k = 1, \dots, n$ . This general formulation of the model would allow considering a heterogeneous population, in terms of both individuals and environments. Interestingly,  $\delta_k$  would allow taking into account different HAV survival probabilities in different areas, which could be variable from region to region (e.g., they could depend on the distance between the fishing areas and the pipe of the sewage system or the harbors). However, because of the lack of reliable data, the theoretical analysis of the model will be carried out in the general case. On the other hand, in the simulations we will assume  $\mu_k = \mu$  and  $\gamma_k = \gamma$ , for all  $k$ .

Based on these considerations, the extension of system (1) to  $n$  classes is the following  $4n$ -dimensional system:

$$\begin{cases} S'_k(t) = -\Lambda_k(t)S_k(t) - \mu_k S_k(t) + \mu_k N_k \\ I'_k(t) = \Lambda_k(t)S_k(t) - (\gamma_k + \mu_k)I_k(t) \\ R'_k(t) = \gamma_k I_k(t) - \mu_k R_k(t) \\ A'_k(t) = \delta_k [I_k(t) - A_k(t)] \end{cases} \quad (2)$$

for  $k = 1, \dots, n$ , where  $\Lambda_k(t) = \sum_{j=1}^n \beta_{kj} \frac{I_j(t)}{N_j} + \sum_{j=1}^n \tilde{\beta}_{kj} A_j(t)$ .

In each region the total population size is assumed to be constant, that is  $N_k = S_k + I_k + R_k$  for all  $k$ . System (2) can be reformulated in terms of relative frequencies: for  $k = 1, \dots, n$ , we define  $s_k(t) = \frac{S_k(t)}{N_k}$ ,  $i_k(t) = \frac{I_k(t)}{N_k}$ ,  $r_k(t) = \frac{R_k(t)}{N_k}$ . In addition, for  $k = 1, \dots, n$ , let us define the rescaled environment-related variable  $a_k(t) = \frac{A_k(t)}{N_k}$  and, for  $k, j = 1, \dots, n$ , the rescaled indirect transmission rate  $\hat{\beta}_{kj} = N_k \tilde{\beta}_{kj}$ . System (2) can now be rewritten as follows:

$$\begin{cases} i'_k(t) = \lambda_k(t) \left[ 1 - i_k(t) - r_k(t) \right] - (\gamma_k + \mu_k) i_k(t) \\ r'_k(t) = \gamma_k i_k(t) - \mu_k r_k(t) \\ a'_k(t) = \delta_k \left[ i_k(t) - a_k(t) \right] \end{cases} \quad (3)$$

where

$$\lambda_k(t) = \sum_{j=1}^n \beta_{kj} i_j(t) + \sum_{j=1}^n \hat{\beta}_{kj} a_j(t) \quad (4)$$

and the fraction of susceptible individuals can be computed as  $s_k(t) = 1 - i_k(t) - r_k(t)$ , for  $k = 1, \dots, n$ .

## 3. Analysis

### 3.1. Equilibria

System (3) trivially admits the disease-free equilibrium:  $s_k = 1$ ,  $i_k = a_k = r_k = 0$ , for  $k = 1, \dots, n$ . Next we investigate

the existence and uniqueness of the endemic equilibrium. The following relations must be satisfied:

$$i_k^* = a_k^* \\ r_k^* = \frac{\gamma_k}{\mu_k} i_k^* \tag{5}$$

From the first equation in (3), using (5) and setting  $\lambda_k^* = \sum_{j=1}^n (\beta_{kj} + \hat{\beta}_{kj}) i_j^*$ , we get the following expression for  $i_k^*$ :

$$i_k^* = \frac{\lambda_k^*}{\gamma_k + \mu_k + \frac{\gamma_k + \mu_k}{\mu_k} \lambda_k^*} = \frac{\sum_{j=1}^n \alpha_{kj} i_j^*}{1 + \vartheta_k \sum_{j=1}^n \alpha_{kj} i_j^*} \tag{6}$$

This is a closed system in the variables  $i_k^*$ , equivalent to (3) evaluated at the equilibrium, where we have defined  $\vartheta_k = \frac{\gamma_k + \mu_k}{\mu_k}$  and  $\alpha_{kj} = \frac{\beta_{kj} + \hat{\beta}_{kj}}{\gamma_k + \mu_k}$ .

Now we consider the application  $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ ,  $i \mapsto F(i)$  where  $i = (i_1, \dots, i_n)$ ,  $F(i) = (F_1(i), \dots, F_n(i))$  and  $F_k(i) = \frac{\sum_{j=1}^n \alpha_{kj} i_j}{1 + \vartheta_k \sum_{j=1}^n \alpha_{kj} i_j}$ .

Provided that the matrix of direct and indirect transmission rates  $(\beta_{kj} + \hat{\beta}_{kj})$  is irreducible (which implies that  $F'(0)$  is irreducible), map  $F$  satisfies the hypotheses of the fixed point theorem (see Hethcote and Thieme (1985) for its statement and proof).

Thus, if we define  $R_0$  as the spectral radius of  $F'(0) = (\alpha_{kj})$ , i.e.,

$$R_0 = \rho \left\{ \left( \frac{\beta_{kj} + \hat{\beta}_{kj}}{\gamma_k + \mu_k} \right) \right\}, \tag{7}$$

by applying the fixed point theorem to  $F$ , we have that a unique endemic equilibrium exists whenever  $R_0 > 1$ . Furthermore, we can apply

**Lemma 3.1** (Thieme, 2003). *If the contact matrix is irreducible, then any endemic equilibrium is strongly endemic, i.e., every component of the endemic equilibrium is strictly positive.*

Thus we obtain the following result:

**Proposition 3.2.** *If  $R_0 > 1$ , then system (3) admits a unique strongly endemic equilibrium.*

If the matrix of transmission rates is not irreducible then the endemic equilibrium is not strongly endemic, i.e.,  $i_k$  may be equal to zero for some  $k$ .

### 3.2. Stability of steady states

In order to prove stability properties of the equilibria, we linearize system (3) in a neighborhood of the steady state, for both the disease free and the strongly endemic equilibrium.

First we study the stability of the disease free equilibrium, and in particular we look for conditions such that it is locally unstable, so that the infection is endemic. The following result holds:

**Proposition 3.3.** *The disease free equilibrium of system (3) is locally asymptotically stable for  $R_0 < 1$ , unstable for  $R_0 > 1$ .*

As regards the stability of the strongly endemic equilibrium, whose existence and uniqueness are guaranteed when  $R_0 > 1$ , we are able to state that

**Proposition 3.4.** *Whenever it exists, the strongly endemic equilibrium (5)–(6) is locally asymptotically stable.*

Both propositions are proved in Appendix.

### 3.3. Transient behavior

In the previous section we proved that the asymptotic behavior of system (2) is similar to a “classical” (i.e., without indirect transmission) SIR model with the same metapopulation structure. This means that the inclusion of indirect transmission in the form of exponentially delayed contacts with people who had been infectious in the past does not modify the basic qualitative behaviors of the model. On the other hand, indirect transmission might significantly change the quantitative outcomes of the system.

If the strongly endemic equilibrium exists, then we know that it is locally asymptotically stable. In the previous section we have not proved anything about how the trajectories of the system reach this equilibrium. Here we show, both analytically and numerically, that the transient behavior of this system is quite different from the classical SIR model.

For illustrative purposes, let us consider a one-region model, so that we can compute the equilibrium explicitly; in particular we have  $i^* = \frac{\mu}{\gamma + \mu} \left( 1 - \frac{1}{R_0} \right)$ , where  $R_0 = \frac{\beta + \hat{\beta}}{\gamma + \mu}$  (see Eq. (7)),  $r^* = \frac{\gamma}{\mu} i^*$ ,  $\lambda^* = (\beta + \hat{\beta}) i^*$ .

It is well known (Thieme, 2003) that the period of transient oscillations around an equilibrium is related to the value of the imaginary part of the eigenvalues defining the spectral bound (i.e., the eigenvalues having the largest real part). In more detail, if such an eigenvalue is  $\xi = \rho \pm i\omega \in \mathbb{C}$ , then the length of the inter-epidemic period can be suitably approximated by the quasi-period  $T = 2\pi / \omega$ .

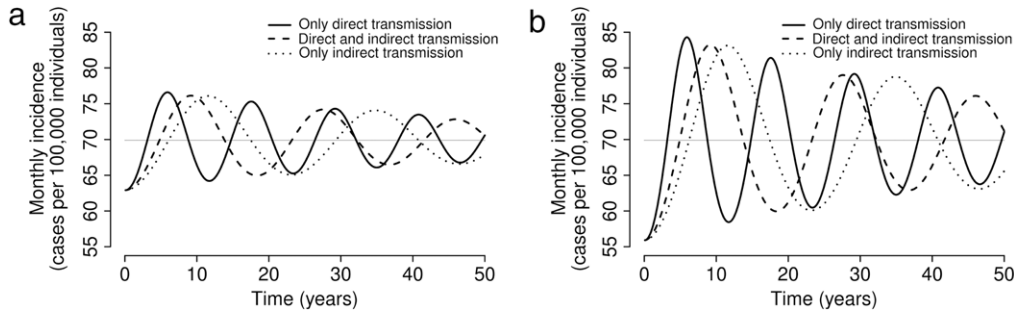
Therefore, we compute the characteristic polynomial of the Jacobian matrix evaluated at the endemic equilibrium, which turns out to be

$$-\xi^3 + (s^* \beta - \lambda^* - \gamma - \delta - 2\mu) \xi^2 + \left[ \delta s^* (\beta + \hat{\beta}) + \mu s^* \beta - \lambda^* (\gamma + \delta + \mu) - (\gamma + \mu) (\delta + \mu) - \delta \mu \right] \times \xi + \delta \left[ \mu s^* (\beta + \hat{\beta}) - \lambda^* (\gamma + \mu) - \mu (\gamma + \mu) \right]$$

(where  $s^* = 1 - i^* - r^*$ ), and find its roots.

As an example, we take fixed parameter values as in Ajelli et al. (2008) (and in particular we assume  $R_0 = 2.9$ ) and let the proportion between the two sources of infection vary. In particular we consider three situations: transmission can occur only by direct contact ( $\hat{\beta} = 0$ ), only by indirect contact ( $\beta = 0$ ) or equally by both types of contact ( $\beta = \hat{\beta}$ ). By computing the roots of the characteristic polynomial we are able to find the period of transient oscillations for the three cases:  $T = 11.6$  years for direct transmission only,  $T = 23.3$  years for indirect transmission only and  $T = 18.4$  years for  $\beta = \hat{\beta}$ . These results are also confirmed by numerical simulations, whose outcomes are shown in Fig. 2 for two different initial conditions. This figure highlights that the presence of indirect transmission increases the wave period, while the initial condition affects only the amplitude of waves without altering their period.

Therefore we are able to conclude that indirect transmission increases the inter-epidemic period, and also the duration of a single epidemic episode. This seems to be consistent with the very long duration of the HAV epidemic episode observed in Puglia during 1996–1997. This makes HAV dynamics in the presence of indirect transmission potentially very different from other endemic infections characterized by recurrent epidemics (e.g., measles, see Grenfell et al. (2001)). The role of the environmental reservoir in increasing the inter-epidemic period emerges in other works on diseases such as cholera (Codeço, 2001), influenza (Rohani et al., 2009; Roche et al., 2009) and scrapie in sheep flocks (Woolhouse et al., 1998). For example, the persistence of bubonic plague in humans, with long inter-epidemic periods, has been suggested to be possible thanks to the reservoir of infection represented by the rodent population (Keeling and Gilligan, 2000).



**Fig. 2.** (a) Different inter-epidemic periods as obtained by varying the proportion between direct and indirect transmission. The solid line refers to  $\hat{\beta} = 0$  (i.e., transmission occurs only by direct contacts), the dotted line refers to  $\beta = 0$  (i.e., transmission occurs only by indirect contacts), and the dashed line represents the intermediate situation in which the transmission risk is equally distributed between direct and indirect contacts ( $\beta = \hat{\beta}$ ). The gray horizontal line represents the endemic equilibrium. Parameters used:  $R_0 = 2.9$ ,  $\gamma = 1.0 \text{ month}^{-1}$ ,  $\mu = 0.001068 \text{ month}^{-1}$ ,  $\delta = 0.333333 \text{ month}^{-1}$ ,  $s(0) = s^*$ ,  $i(0) = 0.9 i^*$  and  $a(0) = i(0)$ . (b) As in (a) except for  $i(0) = 0.8 i^*$ . Source: Parameter values are taken from Ajelli et al. (2008).

#### 4. Simulations

Here we provide a novel approach for estimating spatial contact matrices for metapopulation models, whose classes represent different geographic areas. Moreover, we present a comparison between model simulations and the hepatitis A notified incidence (provided in ISTAT (2006) and available at a regional scale) as well as predictions on the effects of a mass vaccination in Puglia.

##### 4.1. Italian regions

The Italian population (about 60,000,000 individuals, ISTAT (2009b)) is administratively subdivided into twenty regions. The number of individuals per region ranges from about 130,000 (Valle d'Aosta) to about 9,740,000 (Lombardia). Fifteen of the twenty regions border the Mediterranean Sea, but there exists evidence of the presence of infected seafood in the marine environment only in Southern Italy (Mele et al., 2006). Thus, the geographic location of regions is of a certain interest in this study. Social and economic conditions greatly vary from region to region, especially between the North and the South of the country. Fig. 3(a) shows geographic location, population size and enumeration (adopted throughout this manuscript) of the Italian regions.

##### 4.2. Spatial contact matrix

In a metapopulation model, the force of infection depends on the level of mixing between classes and on the per capita transmission rate. Therefore, it is possible to write Eq. (4) as follows:

$$\lambda_k(t) = \sum_{j=1}^n p_k c_{kj} i_j(t) + \sum_{j=1}^n \hat{p}_k c_{kj} a_j(t) \quad (8)$$

where  $p_k$  represents the transmission rate via direct contact for individuals living in region  $k$ ;  $\hat{p}_k$  represents the contagion rate via indirect contact for individuals living in region  $k$  and  $(c_{kj})$  is the contact matrix describing the level of mixing between the Italian regions. For the sake of simplicity, we assume  $\hat{p}_k = \nu p_k$ , where  $\nu$  is a scaling factor which does not vary between regions. The latter assumption is a simplification which allows us to consider the data on hepatitis A risk factors described in Mele et al. (2006).

Both short-distance and long-distance travels may play a central role in driving the spatiotemporal dynamics of hepatitis A in Italy. One of the aims of this study is to quantify the relative impact of spatial mobility on hepatitis A dynamics and the effects of a vaccination program adopted in a single region on the others. It is of particular interest to evaluate if the mass vaccination program started in Puglia after the great 1996–1997 epidemic could have played a role in the decline of HAV incidence observed almost

everywhere in Italy in the period 1997–2003. Such an effect has been already observed in North America, where the hepatitis A vaccination program performed in the US has been shown to play a crucial role in the decline of HAV incidence in Canada (Amarie et al., 2008).

Specifically, for defining the spatial contact matrix, we consider two types of movements: short-distance travel, accounting for daily commuting between place of residence and place of work/study, and long-distance travel, related to occasional touristic travels.

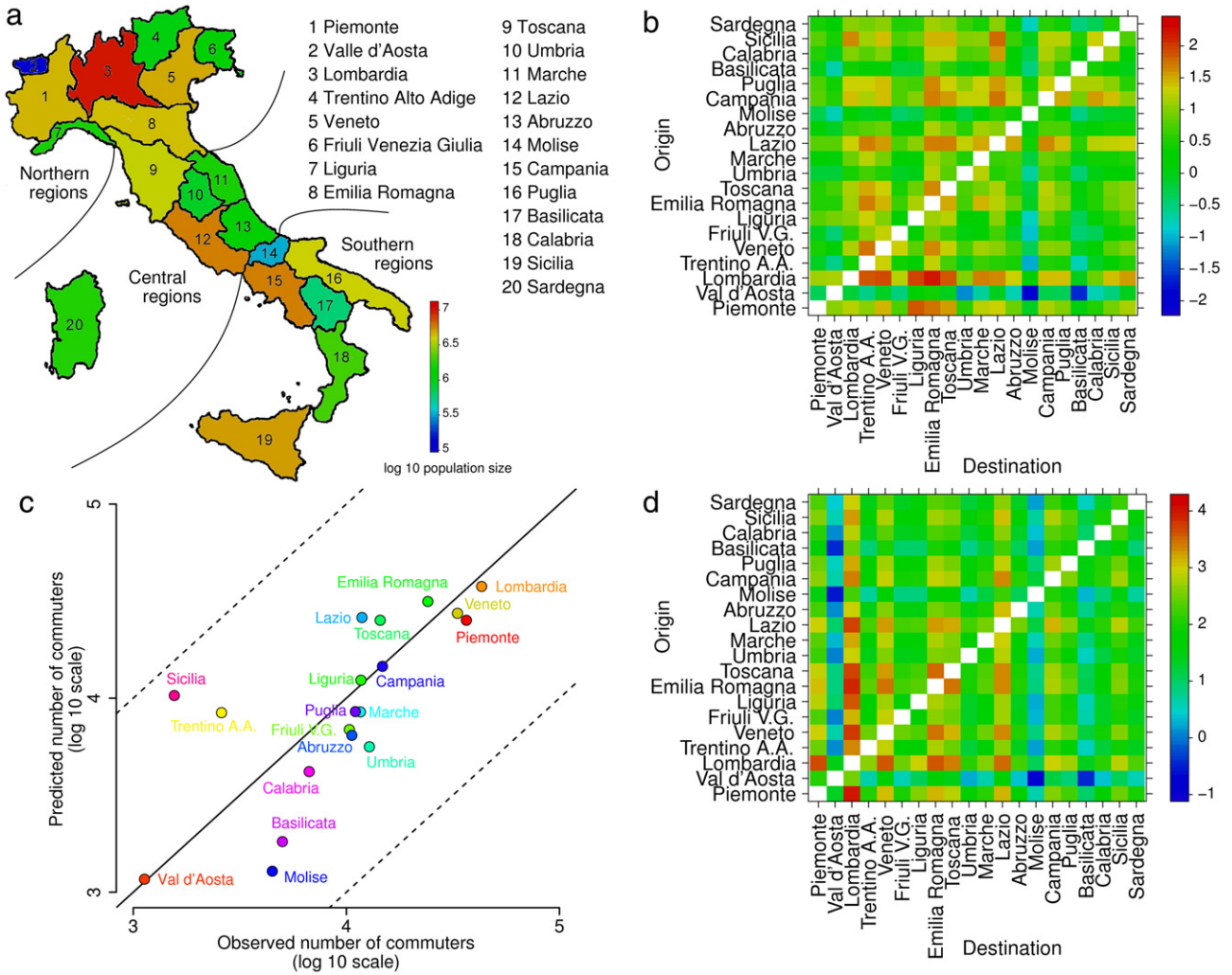
As regards touristic flows, the Italian Institute of Statistics provides the full origin–destination matrix at the regional level. In Fig. 3(b), the daily number of tourists traveling from region  $i$  to region  $j$  is shown, as derived by ISTAT (2002). As one may expect, the destination regions are more related to touristic attractions (e.g., art cities, seaside and mountain resorts) rather than to the geographical distance between origin and destination regions.

As regards commuting, unfortunately the origin–destination matrix for daily travels between the place of residence and the place of work/study is not available to us. Therefore we use a gravity model for simulating such fluxes. This model is based on the population size, on the Gross Domestic Product (GDP) of the regions and on the distances between them. Differently from Ajelli and Merler (2008) (where commuting destinations are supposed to depend only on the distance between place of residence and work/school) and also from Viboud et al. (2006) and Ciofi degli Atti et al. (2008) (where destinations are supposed to depend on distance and population size), in Merler and Ajelli (2010) it is shown that the three aforementioned factors should all be taken into account for better reproducing the origin and destination of travels at the European level. Specifically, the gravity model  $H_{kj}$ , describing the flux of commuters from region  $k$  to region  $j$ , can be formalized as follows:

$$H_{kj} = \theta \frac{g_k^{\tau_d} g_j^{\tau_r}}{d_{kj}^\rho} \quad (9)$$

where  $g_k$  is the normalized GDP of region  $k$  ( $g_k = \frac{G_k}{G^*} \varphi_k$  where  $G_k$  is the per capita GDP of region  $k$ ,  $G^*$  is the average Italian per capita GDP and  $\varphi_k$  is the population of region  $k$ ) and  $d_{kj}$  is the geodetic distance between the two capitals (often corresponding to the most populous cities) of the regions.  $\tau_d$  and  $\tau_r$  tune the dependence of population movements on donor and recipient regions,  $\rho$  tunes the dependence on the distance and  $\theta$  determines the overall number of commuters. The gravity model depends on four parameters (namely,  $\tau_d$ ,  $\tau_r$ ,  $\rho$ ,  $\theta$ ) which must be optimized in order to fit the available data.

In particular, for each region, the Italian Institute of Statistics (ISTAT, 2001) provides the number of commuters (here a commuter is defined as an individual working/attending school in a



**Fig. 3.** (a) Geographic location of the Italian regions. Colors account for the size of the population (in log 10 scale). (b) Origin–destination matrix of touristic flows between regions as obtained from the data provided by the Italian Institute of Statistics (ISTAT, 2002). Colors account for the number of travelers in log 10 scale. (c) Comparison between the observed number of commuters (here defined as individuals working/attending school in a region other than that of residence) and the number predicted by the commuting model. Sardegna island is not shown since it has no commuters. (d) Simulated spatial matrix accounting for both touristic and commuting flows. Colors account for the number of travelers on a log 10 scale.

region other than the one of residence), without specifying the region of destination. Therefore the objective function is defined as follows

$$\Gamma(\tau_d, \tau_r, \rho, \theta) = \sum_{k=1}^{20} \left( \theta \sum_{j=1}^{20} \frac{g_k^{\tau_d} g_j^{\tau_r}}{d_{kj}^{\rho}} - M_k \right)^2$$

where  $M_k$  is the number of commuters from region  $k$ .

By analytically solving the first order condition on  $\theta$ ,  $\frac{\partial}{\partial \theta} \Gamma(\tau_d, \tau_r, \rho, \theta) = 0$ , we obtain the following constraint:

$$\theta = \frac{\sum_{k=1}^{20} M_k}{\sum_{k=1}^{20} \sum_{j=1}^{20} \frac{g_k^{\tau_d} g_j^{\tau_r}}{d_{kj}^{\rho}}} \quad (10)$$

By imposing condition (10), model (9) is forced to produce  $\sum_{k=1}^{20} M_k$  commuters and thus  $H_{kj}$  can now be interpreted as the probability of commuting from region  $k$  to region  $j$ . Therefore, only three parameters need to be optimized and this can be done by simple algebraic calculation. The estimated values for the parameters are:  $\tau_d = 0.938$  (95% CI: 0.804–1.373),  $\tau_r = 1.674$  (95% CI: 0.582–1.994),  $\rho = 0.716$  (95% CI: 0.631–1.31), where

the confidence intervals have been computed by a leave-one-out procedure (Hastie et al., 2001). Gravity model predictions (where optimal parameters values are considered) and observed data are presented in Fig. 3(c). Predictions and observations are in good agreement and all predicted values differ by much less than one order of magnitude from the observed ones.

Finally, the matrices for long-distance and short-distance travels are combined to obtain a spatial contact matrix describing the mixing level between Italian regions. As in Sattenspiel and Dietz (1995), the level of mixing between geographic areas is based on the assumption that the time spent by an individual in a certain area plays a central role for determining the received force of infection. Therefore, in order to combine the two matrices, we sum the daily touristic flows matrix with the daily commuting matrix multiplied by a factor 0.6 (i.e., 220/365, in order to take into account that there are 200 scholastic days and about 240–260 working days per year). The resulting spatial contact matrix is shown in Fig. 3(d). The matrix shows a remarkable number of travels to regions having large GDP and population, either absolute (as the case of Lombardia and Lazio) or relative to neighboring regions (as the case of Veneto, Emilia Romagna, Toscana and Piemonte). Populous regions with a small GDP, instead, do not show large numbers of incoming travelers (e.g., Campania). Also, small regions with high

**Table 1**  
Model parameters kept fixed in the simulations.

Parameter	Description	Measure unit	Value	Reference
$N_k$	Number of individuals in region $k$	Dimensionless	<sup>a</sup>	(ISTAT, 2009b)
$1/\mu$	Average life expectancy	Years	78	(ISTAT, 2009a)
$1/\gamma$	Average duration of the infectivity period	Months	1	(CDC, 2007; Stapleton and Lemon, 1994)
$1/\delta_k$	Average survival period of HAV in the environment for region $k$	Months	$3, 0^b$	(Abad et al., 1994; Biziagos et al., 1988; Mbithi et al., 1991; Ajelli et al., 2008)
$\epsilon$	Fraction of notified cases	Percentage	$3\text{--}8^c$	(Ajelli and Merler, 2009)
$\nu$	Scaling factor for the two sources of infection	Dimensionless	2.5	(Mele et al., 2006)
$V_{16}^c$	Vaccination coverage at birth	Percentage	$20^d$	(Lopalco et al., 2005; Ajelli et al., 2008)
$V_{16}^y$	Vaccination rate of 12-years-old individuals	Months <sup>-1</sup>	$0.0009^d$	(Lopalco et al., 2005; Ajelli et al., 2008)

<sup>a</sup> Values reported in Fig. 3(a).  
<sup>b</sup>  $1/\delta_k = 3$  for  $14 \leq k \leq 19$  (i.e., for Southern Italy); 0 otherwise.  
<sup>c</sup> Uniformly distributed.  
<sup>d</sup> Puglia is the only Italian region which has experienced a vaccination program; for the other regions this parameter is set to 0.

touristic flows do not necessarily have a high number of incoming travelers (e.g., Trentino Alto Adige). These results suggest that, for defining the spatial contact matrix at national level for epidemic modeling, commuting flows are more relevant than touristic flows.

Finally, it is worth noting that the estimated contact matrix ( $C_{kj}$ ) is irreducible by construction. Therefore, from the theory above, a unique, strongly endemic equilibrium exists and is stable for the empirical specification of the model.

4.3. Model parameterization, fitting procedure and vaccination

Model parameterization is based on basic demographic information on the Italian population, hepatitis A natural history and model fit to average notification data. Specifically, the population size of each region and life expectancy at national level are taken from the Italian Institute of Statistics (ISTAT, 2009b,a). The average infectivity period for hepatitis A is fixed to 1 month (CDC, 2007; Stapleton and Lemon, 1994). The life expectancy of HAV in the environment is set to three months (Abad et al., 1994; Biziagos et al., 1988; Mbithi et al., 1991; Ajelli et al., 2008). As regards the reporting factor  $\epsilon$ , we assume a unique value across the regions. This is a simplifying assumption since we expect that regions characterized by different levels of endemicity, as is the case of Italy, could have very different proportions of symptomatic cases (due to differences in the age at which infection is acquired) and reporting rates. We also explore the case of two different reporting rates for Northern and Southern Italy. However, given the uncertainty on the reporting factor, we use a range of plausible values for  $\epsilon$  taken from previous literature on HAV dynamics in Italy (Ajelli and Merler, 2009) rather than fix it to one specific value. Therefore, conditionally on the estimated contact matrix, the uncertainty in parameter estimates (and thus also in  $R_0$ ) will mainly reflect the uncertainty in the reporting factor. Model parameter values are summarized in Table 1.

Transmission rates are estimated by fitting the model at its endemic equilibrium to the actual average HAV incidence of notified cases in each region, computed over the pre-vaccination period January 1992–December 1996. Specifically, the optimization procedure works as follows. First of all, a value of  $\epsilon$  (the reporting factor, which accounts for both the symptomaticity of HAV infection and the accuracy of the reporting system) is randomly chosen from the uniform distribution on the range 0.03–0.08 according to previous investigations on hepatitis A in Italy (Ajelli and Merler, 2009); then we compute the corresponding mean square error (MSE) between the number of notified cases simulated by the model  $\epsilon i_k^*$  and the actual number of notified cases as follows:

$$MSE(p_k) = \sum_{k=1}^{20} (\bar{\Phi}_k - \epsilon i_k^*)^2$$

where  $i_k^*$  is the HAV incidence at the endemic equilibrium in region  $k$ , which depends on the estimated value of the transmission rates

$p_k$ , and  $\bar{\Phi}_k$  is the average number of monthly notified cases over the 5-years period from 1992 to 1996 in region  $k$  under the assumption they reflect the endemic equilibrium. In general, case reports are a better match for incidence rather than for prevalence but, since the average length of the infectivity period is assumed to be 1 month (i.e.,  $\gamma_k = 1$  for  $k = 1, \dots, n$ ), at the endemic equilibrium monthly incidence and prevalence coincide. For Puglia,  $\bar{\Phi}_{16}$  refers to the period from January 1992 to December 1995 in order to exclude the large 1996 epidemic that could hardly be considered close to the equilibrium value.

Finally, for each choice of the reporting factor, the best transmission rates  $p_k$  are defined as the ones minimizing the resulting MSE. The MSE is optimized by using a random walk stochastic local search algorithm (Hoos and Stutzle, 2005). In order to keep our model as parsimonious as possible we assume that there are only four distinct  $p_k$  values (see Eq. (8)), i.e., one for Puglia, one for Campania (the two regions which mostly contribute to HAV endemicity in Italy), one for the remaining Southern regions, and one for Northern and Central Italy and Sardegna, where HAV may be close to subcriticality. Therefore, we consider

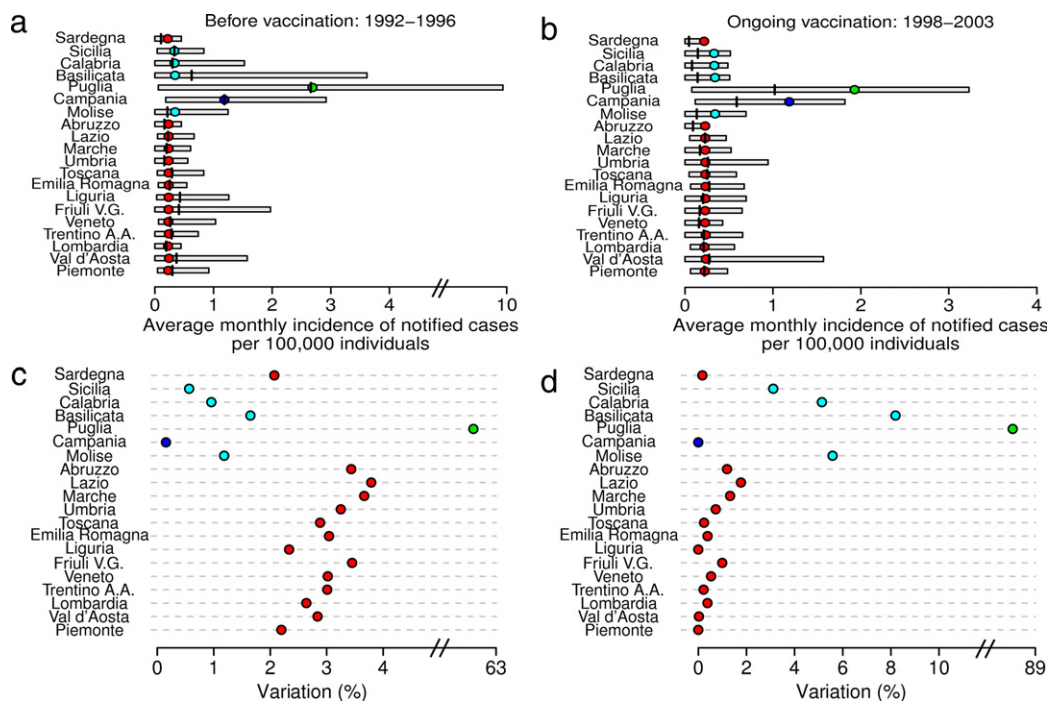
$$p_k = \begin{cases} p^{(1)} & \text{if } 1 \leq k \leq 13 \text{ or } k = 20 \text{ (i.e., Northern and Central Italy and Sardegna island)} \\ p^{(2)} & \text{if } k = 15 \text{ (i.e., Campania)} \\ p^{(3)} & \text{if } k = 16 \text{ (i.e., Puglia)} \\ p^{(4)} & \text{otherwise (i.e., the rest of Southern Italy).} \end{cases}$$

The proportionality factor between indirect and direct transmission rates  $\nu = \hat{p}_k/p_k$  is obtained by survey data (Mele et al., 2006) and it is set to 2.5. Since there does not exist evidence of native infected seafood in the Northern, Central and Sardegna marine environments, we assume that  $\delta_k$  and  $a_k(0)$  are both identically zero for  $1 \leq k \leq 13$  or  $k = 20$ . Finally, by sampling the distribution of the reporting factor 1000 times, and repeating the optimization procedure, the related distribution of the estimates of the transmission rates  $p_k$  and of the basic reproduction numbers were found.

For comparing model predictions with notification data on the period January 1998–December 2003, the vaccination program (involving both newborns and 12-year-old individuals) started in Puglia in 1997 has to be taken into account. Therefore, we introduce the following system:

$$\begin{cases} S'_k(t) = -\Lambda_k(t)S_k(t) - (\mu_k + V_k^y)S_k(t) + (1 - V_k^c)\mu_k N_k \\ I'_k(t) = \Lambda_k(t)S_k(t) - (\gamma_k + \mu_k)I_k(t) \\ R'_k(t) = \gamma_k I_k(t) - \mu_k R_k(t) + V_k^c \mu_k N_k + V_k^y S_k(t) \\ A'_k(t) = \delta_k [I_k(t) - A_k(t)] \end{cases}$$

where  $V_k^c$  represents the vaccination coverage of newborns in region  $k$ , and  $V_k^y$  represents the vaccination rate of adolescent individuals in region  $k$ .



**Fig. 4.** (a) Average monthly incidence of notified cases per 100,000 individuals in the twenty Italian regions during the period January 1992–December 1996 (black line), 95% CI (gray rectangles) and model equilibrium (colored dots). Red dots refer to Northern and Central regions and Sardegna; the green dot refers to Puglia; the blue dot refers to Campania; cyan dots refer to the other Southern regions. For Puglia, the average monthly incidence is computed over the period from January 1992 to December 1995. (b) As in (a), but for the period January 1998–December 2003. Here, a vaccination program involving 20% of newborns and 80% of 12-years-old adolescents is also simulated. (c) Variations (in percentage) between the monthly incidence evaluated at the endemic equilibrium (not considering any vaccination program) and the average monthly incidence after 5 years from the beginning of the vaccination program in Puglia. Colors as in (a). (d) Variations (in percentage) between the monthly incidence evaluated at the endemic equilibrium (not considering any vaccination program) and the monthly incidence evaluated at the new endemic equilibrium, as resulting by simulating the vaccination program in Puglia. Colors as in (a).

The vaccination coverage at birth  $V_{16}^c$  is kept fixed at 20% and the monthly vaccination rate of young individuals  $V_{16}^y$  is kept fixed at 0.0009, which corresponds to 80% of the fraction of 12-year-old individuals ( $\approx 0.013$  of the population) divided by 12 (in order to obtain a monthly rate). Such values are chosen in order to mimic the vaccination campaign performed in Puglia (Lopalco et al., 2005).

#### 4.4. Model predictions and empirical epidemiological data

The simulated equilibrium is in good agreement with the average monthly hepatitis A incidence of notified cases computed over the period from January 1992 to December 1996 (we recall that for Puglia the average refers to the period January 1992–December 1995). As shown in Fig. 4(a), the model fits all data very well except for a few regions (namely Basilicata, Liguria and Friuli Venezia Giulia).

The fitting procedure allows also computing the basic reproductive number, as given by Eq. (7). In particular, it is possible to estimate  $R_0$  for different Italian areas by restricting the computation of Eq. (7) to the corresponding rows and columns. In Puglia the estimated  $R_0$  is 2.43 (95% CI: 1.47–5.48), where the observed variability mainly depends on the uncertainty in the HAV reporting factor. This value is between the estimate given in Martinelli et al. (2010), namely 2.01, and the one given in Ajelli et al. (2008), namely 2.9, which were both computed from the average age at infection. In Campania we obtain  $R_0 = 1.31$  (95% CI: 1.17–1.56), which is noticeably lower than the value estimated in Ajelli et al. (2008), namely 2.2. Therefore, both estimates given here, based on equilibrium model fit, are lower than the corresponding ones based on age specific data. As regards the other Italian areas, the estimated  $R_0$  values are  $R_0 = 1.07$  (95% CI: 1.04–1.11) for Southern Italy (Puglia and Campania excluded) and  $R_0 = 1.04$  (95% CI:

1.02–1.07) for Northern and Central Italy and Sardegna. The latter estimate, close to 1, suggests that hepatitis A is in a situation of extremely low endemicity in Northern and Central Italy. As discussed above, the reporting factor could be highly variable between the Italian regions: in particular its average value of 5.5% used here seems a reasonable approximation for Puglia, while the reporting factor is probably noticeably higher in other Italian regions (e.g., in Northern Italy) and probably lower in other Southern regions, especially in Campania: a survey conducted in Campania has highlighted that only 39% of the interviewed people would visit a medical doctor in case of icteric onset (Salamina and D'Argenio, 1998).

In order to deepen our understanding of the actual hepatitis A situation in Italy, we relaxed the assumption of a single reporting factor for the whole study area by considering a specific value for Northern and Central regions and Sardegna, namely  $\epsilon_1$ , and another one for Southern regions (including Puglia and Campania), namely  $\epsilon_2$ . Since the average age at which infection is acquired is notably higher in Northern and Central regions than in Southern regions (ISTAT, 2006) and HAV symptoms increase with age (Stapleton and Lemon, 1994; Armstrong and Bell, 2002), we fix  $\epsilon_1 = 0.2$  according to the estimate given in Armstrong and Bell (2002) for the US population aged more than 14 years. On the other hand, we continue to sample  $\epsilon_2$  from the uniform distribution of range 0.03–0.08. In this case, the best estimate of the reproductive number for Northern and Central Italy and Sardegna results to be  $R_0 = 1.002$ , with a corresponding 95% confidence interval of 0.992–1.006, while the estimates for the other areas are basically the same as obtained by assuming a single reporting factor. This confirms our result that in Northern and Central Italy, hepatitis A is in a situation of extremely low endemicity and it could possibly be even non-endemic. From now on results will refer to the case of a single reporting factor  $\epsilon$  for the whole country.

By keeping the estimated transmission rates fixed, it is possible to simulate the vaccination program in Puglia and study its effect on the overall hepatitis A dynamics in Italy. Our predictions are in good agreement with observed notification data for the whole Northern Italy, while the model overestimates the number of cases in Southern regions, especially in Puglia and Campania (see Fig. 4(b)). These discrepancies may be due to the simplifying assumption that the proportion of notified cases does not vary over time. In fact, it is well known by epidemiologists that the high perceived risk during the 1996 epidemic in Puglia and Campania induced a significant increase in the notification rate during the course of the epidemic (Ajelli et al., 2008). Since this is not taken into account in the model fit, the value of the respective fitted transmission rates may be somewhat overestimated and thus model predictions for the period 1998–2003 are higher than the observed data, especially in Puglia and Campania. As regards the impact of the vaccination program in Puglia, the model predicts a notable decline in HAV incidence in Puglia, namely about 60% after 5 years from the beginning of vaccination. Moreover, a slight decline is observable in all other regions, except Campania (see Fig. 4(c)). The difference in the number of notified cases during the pre-vaccination period is larger in Northern and Central Italy than in Southern regions; this can be explained by the lower  $R_0$  value in the Northern-Central part of the country. Moreover, the limited effect on hepatitis A dynamics in Campania strongly suggests that this disease is endemic in that region even without contacts with Puglia. Fig. 4(d) shows the effects of the vaccination program in Puglia (by assuming that the vaccination rates do not vary over time) on the new long-term endemic equilibrium. It is interesting to note that the vaccination program has a significant effect on hepatitis A dynamics in all Southern regions (Campania excluded) and a limited effect on Northern-Central Italian regions. This is a consequence of the spatial contact matrix, which shows higher travel fluxes between Puglia and the other Southern regions rather than between Puglia and the rest of Italy (see Fig. 3(d)).

## 5. Discussion and conclusions

This work extends the model for HAV transmission dynamics with multiple sources of infection presented in Ajelli et al. (2008). To the best of our knowledge the present manuscript represents the first instance of a metapopulation model including direct and indirect transmission, of which HAV is perhaps the most notable example. Our analysis proves that the general form of the proposed model shows the qualitative behavior of the classical SIR model (considering only the person-to-person source of infection). However, from a quantitative point of view, it shows a different pattern accounting for longer epidemic and inter-epidemic periods. A spatial contact matrix for Italy at the regional level has been derived. This matrix takes into account the most relevant human movements: daily short-distance travels (e.g., for commuting from the place of residence to the place of study/work) and occasional long-distance travels (e.g., for touristic reasons). A gravity model for commuting flows has been developed and specifically parameterized for the Italian situation. The resulting spatial contact matrix represents a useful starting point for developing Italian metapopulation models, with regional spatial resolution, focused on other infectious diseases such as, for instance, influenza or measles.

The recent history of hepatitis A in Italy is characterized by a large heterogeneity, in that although HAV might still be considered endemic in the nation as a whole, this seems to be true only for a few areas (mainly located in Southern Italy) which are strongly affected by HAV, while the others show only few, sporadic cases during the year. This suggests that in the latter regions HAV has gone near to, or even below, the critical threshold. This fact seems to be appropriately captured by our model which indicates

values of the reproductive numbers slightly greater than one in Northern and Central Italy or even possibly slightly below one if heterogeneity in reporting factors, as empirically documented, is considered. This is what can be said by a model like the present one, i.e., one which depicts a strongly endemic situation governed by constant transmission rates. Further inquiry on the actual patterns of HAV in low endemic Italian areas would require more specific approaches. For example the pattern in low endemic Italian areas could also be consistent with “source/sink” dynamics driven by seasonal variation in within- or between-subpopulation transmission, or with a low endemic equilibrium in a discrete population, in presence of a low reporting rate. We also feel it would be important to better investigate the role of international importation, quite substantial in Italy, under situations of sub-criticality.

Model simulations are in good agreement with average notification values in both the considered periods for most of Italy. The simulations highlight that the vaccination program in Puglia could have played a role in the observed decline of hepatitis A incidence in several Italian areas. Moreover, the effects of a vaccination program on the long-term endemic equilibrium is shown and it has been underlined that the strong connection between Puglia and other Southern regions is responsible for a large decrease of HAV prevalence. On the contrary, Campania, which shows a high level of endemicity by itself, is not substantially affected by HAV dynamics in Puglia.

The aim of this study is not to evaluate the effectiveness of specific control measures or to suggest which kind of vaccination program would be more effective. For answering these questions a more structured model is preferable; e.g., an age-structured or an individual-based model. The latter should be useful also for testing individually targeted interventions such as closure of day care centers or kindergartens which are considered as the main source of person-to-person transmission for many childhood diseases (Galil et al., 2002), and hepatitis A as well (Chitambar et al., 1996). A specific study on this topic can be found in Ajelli and Merler (2009).

The model proposed here does not take into account either the seasonality of HAV in the marine environment, which may be an explanation for the seasonal pattern observed in Puglia and Campania, as previously suggested in Ajelli et al. (2008), or any stochastic fluctuation in the transmission rates (e.g., due to human behavior, such as different levels of seafood consumption during the year, or to climatic changes in the marine environment which in turn would be responsible for different seafood availability). Therefore, a comparison limited to the average notification data over two distinct periods, as carried out here, rather than to the monthly time series seems to represent a better option. Moreover, this choice is also supported by the lack of dynamical patterns observed in the time series of notification data (with some notable exceptions). As regards the reporting factor, we assumed in most of our work a unique value for Italy as a whole. This however certainly represents an oversimplification of a complex phenomenon involving the endemicity level, the average age at which individuals acquire infection and the accuracy of the reporting system. Indeed, simulations assuming two different reporting factors have been performed and seem to better capture the behavior of hepatitis A but at the cost of introducing a further parameter whose estimate is uncertain.

Finally, it would certainly be important to improve tools for the evaluation of the global uncertainty embedded in the main model parameter estimates. In this preliminary effort uncertainty has been primarily related to the HAV reporting factor, which is a well-documented source of uncertainty in HAV notification data. However, many other uncertainty sources exist, e.g. in the estimation of transmission rates, in the parameter estimates of the mobility model tuning the spatial contact matrix, and so on, which

would be important to appropriately take into account in future work.

Despite some limitations, the proposed model has shown some capability in explaining some observed hepatitis A patterns both in pre-vaccination and ongoing vaccination settings and therefore could represent a reliable starting point for future further analyses. Combining the use of notification, seroprevalence and age-specific data would allow further improvements in understanding hepatitis A dynamics.

**Acknowledgments**

The authors thank Benjamin M. Bolker (University of Florida), Lorenzo Pellis (Imperial College London), Mimmo Iannelli (University of Trento) and Andrea Pugliese (University of Trento) and an anonymous reviewer for their useful comments that certainly have contributed to improve this manuscript. MA thanks Alessandro Vespignani (Indiana University) for having hosted him at the Indiana University School of Informatics and Computing during his visiting period, when part of the present work has been carried out. MA, LF and SM thank the European Union FP7 EPIWORK project for research funding. MA and PM additionally thank the EPICO project for research funding.

**Appendix. Details on the stability analysis**

**Proposition 3.3.** *The disease-free equilibrium of system (3) is locally asymptotically stable for  $R_0 < 1$ , unstable for  $R_0 > 1$ .*

We linearize system (3) in a neighborhood of the disease-free equilibrium  $(s, i, a) = (1, 0, 0)$ :

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} J_{11} & J_{12} & J_{13} \\ 0 & J_{22} & J_{23} \\ 0 & J_{32} & J_{33} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

where  $x, y, z$  are  $n$ -dimensional vectors and the  $n \times n$  submatrices are defined as follows:

$$\begin{aligned} J_{11} &= \text{diag}(-\mu_k)_{k=1,\dots,n}, & J_{12} &= (-\beta_{kj})_{k,j=1,\dots,n}, \\ J_{13} &= (-\hat{\beta}_{kj})_{k,j=1,\dots,n}, \\ J_{22} &= \begin{pmatrix} \beta_{11} - (\gamma_1 + \mu_1) & \beta_{12} & \dots & \beta_{1n} \\ \beta_{21} & \beta_{22} - (\gamma_2 + \mu_2) & \dots & \beta_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{n1} & \beta_{n2} & \dots & \beta_{nn} - (\gamma_n + \mu_n) \end{pmatrix} \\ J_{23} &= (\hat{\beta}_{kj})_{k,j=1,\dots,n}, & J_{32} &= \text{diag}(\delta_k)_{k=1,\dots,n}, \\ J_{33} &= \text{diag}(-\delta_k)_{k=1,\dots,n}. \end{aligned}$$

We notice that the Jacobian matrix is decomposable; therefore its spectrum is given by the eigenvalues of  $J_{11}$  and of the matrix

$$L = \begin{pmatrix} J_{22} & J_{23} \\ J_{32} & J_{33} \end{pmatrix}.$$

Since the eigenvalues of  $J_{11}$  all have negative real parts, the stability or instability of the disease-free equilibrium comes from  $L$ ; so from now on we focus on this matrix. The components of the linearized system we consider are therefore those related to infectious individuals and the amount of virus in the environment:

$$\begin{cases} y'_k(t) = \lambda_k(t) - (\gamma_k + \mu_k)y_k(t) \\ z'_k(t) = \delta_k[y_k(t) - z_k(t)]. \end{cases} \quad (\text{A.1})$$

First, we prove that the disease-free equilibrium is locally asymptotically stable for  $R_0 < 1$ . We set  $y_k = y_k^* \exp(\xi t), z_k = z_k^* \exp(\xi t)$ , with  $\xi \in \mathbb{C}$ , and substitute into (A.1), getting

$$\begin{cases} \xi y_k^* = \sum_{j=1}^n (\beta_{kj} y_j^* + \hat{\beta}_{kj} z_j^*) - (\gamma_k + \mu_k) y_k^* \\ \xi z_k^* = \delta_k (y_k^* - z_k^*). \end{cases} \quad (\text{A.2})$$

We assume by contradiction that the steady state is not stable, so  $\Re \xi \geq 0$ , and derive from (A.2)

$$\begin{aligned} z_k^* &= \frac{1}{1 + \xi / \delta_k} y_k^* \\ (\xi + \gamma_k + \mu_k) y_k^* &= \sum_{j=1}^n (\beta_{kj} y_j^* + \hat{\beta}_{kj} z_j^*). \end{aligned}$$

Since  $|1 + \xi|^2 = (1 + \Re \xi)^2 + (\Im \xi)^2 \geq 1$ , we have that

$$\begin{aligned} |z_k^*| &\leq |y_k^*| \\ \left| 1 + \frac{\xi}{\gamma_k + \mu_k} \right| |y_k^*| &\leq \sum_{j=1}^n \frac{\beta_{kj} + \hat{\beta}_{kj}}{\gamma_k + \mu_k} |y_j^*|. \end{aligned}$$

Let  $\varepsilon = \min_{k=1,\dots,n} \Re \left( \frac{\xi}{\gamma_k + \mu_k} \right) > 0$ ; therefore, since

$$\left| 1 + \frac{\xi}{\gamma_k + \mu_k} \right| \geq \Re \left( 1 + \frac{\xi}{\gamma_k + \mu_k} \right) \geq 1 + \varepsilon,$$

we get

$$\left| 1 + \frac{\xi}{\gamma_k + \mu_k} \right| |y_k^*| \leq \sum_{j=1}^n \frac{\beta_{kj} + \hat{\beta}_{kj}}{\gamma_k + \mu_k} |y_j^*|. \quad (\text{A.3})$$

We can thus apply the following lemma:

**Lemma A.1** (Thieme, 2003). *Let  $M$  be a positive matrix and  $\eta \geq 0$  such that  $M^h v \geq \eta v$  for some natural number  $h > 0$  and some vector  $v \in \mathbb{R}^m$  such that  $-v \notin \mathbb{R}_+^m$ . Then the spectral radius of  $M$ ,  $\rho(M)$ , is such that  $\rho(M) \geq \eta^{1/h}$ .*

We take  $M = \begin{pmatrix} \beta_{kj} + \hat{\beta}_{kj} \\ \gamma_k + \mu_k \end{pmatrix}$ ,  $\eta = 1 + \varepsilon$ ,  $q = 1$ ,  $v = (|y_1^*|, \dots, |y_n^*|) \in \mathbb{R}^n$ : the hypotheses of Lemma A.1 are satisfied thanks to (A.3), consequently

$$\rho \left\{ \begin{pmatrix} \beta_{kj} + \hat{\beta}_{kj} \\ \gamma_k + \mu_k \end{pmatrix} \right\} = R_0 \geq 1 + \varepsilon > 1;$$

but this contradicts the hypothesis that  $R_0 < 1$ . Therefore all eigenvalues of  $L$  have a negative real part, and so the disease-free equilibrium is locally asymptotically stable for  $R_0 < 1$ .

Now we prove that this steady state is unstable for  $R_0 > 1$ . Let us write  $L$  more extensively as given in Box I.

The following lemma turns out to be useful:

**Lemma A.2** (Thieme, 2003). *Let  $M$  be a quasi-positive matrix. If there exist a positive vector  $v$  and  $\eta \in \mathbb{R}$  such that  $Mv \geq \eta v$ , then the spectral bound of  $M$ ,  $s(M)$ , satisfies  $s(M) \geq \eta$ .*

We take  $M = L, \eta = 0, v = (i_1^*, \dots, i_n^*, a_1^*, \dots, a_n^*) = (i_1^*, \dots, i_n^*, i_1^*, \dots, i_n^*)$ , where  $i_k^*, a_k^*$  are the components of the endemic equilibrium; notice that it exists since we are considering the case  $R_0 > 1$ . Then  $Lv$  gives

$$\sum_{j=1}^n (\beta_{kj} + \hat{\beta}_{kj}) i_j^* - (\gamma_k + \mu_k) i_k^*$$

for  $k = 1, \dots, n$  (i.e., the first  $n$  components), and 0 for the components  $n+1, \dots, 2n$ . This vector is non-negative thanks to Eq. (6); therefore Lemma A.2 guarantees that  $s(L) \geq 0$ , that is, there is at least an eigenvalue of  $L$  with positive real part, so the disease free equilibrium is unstable (and hence the infection becomes endemic) for  $R_0 > 1$ .

$$L = \begin{pmatrix} \beta_{11} - (\gamma_1 + \mu_1) & \beta_{12} & \cdots & \beta_{1n} & \hat{\beta}_{11} & \hat{\beta}_{12} & \cdots & \hat{\beta}_{1n} \\ \beta_{21} & \beta_{22} - (\gamma_2 + \mu_2) & \cdots & \beta_{2n} & \hat{\beta}_{21} & \hat{\beta}_{22} & \cdots & \hat{\beta}_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \beta_{n1} & \beta_{n2} & \cdots & \beta_{nn} - (\gamma_n + \mu_n) & \hat{\beta}_{n1} & \hat{\beta}_{n2} & \cdots & \hat{\beta}_{nn} \\ \delta_1 & 0 & \cdots & 0 & -\delta_1 & 0 & \cdots & 0 \\ 0 & \delta_2 & \cdots & 0 & 0 & -\delta_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \delta_n & 0 & 0 & \cdots & -\delta_n \end{pmatrix}$$

Box I.

**Proposition 3.4.** *Whenever it exists, the strongly endemic equilibrium (5)–(6) is locally asymptotically stable.*

We linearize system (3) in a neighborhood of the strongly endemic equilibrium  $(i^*, r^*, a^*)$ , thus obtaining, for  $k = 1, \dots, n$ ,

$$\begin{cases} x'_k = (1 - i_k^* - r_k^*) \sum_{j=1}^n (\beta_{kj} x_j + \hat{\beta}_{kj} z_j) \\ \quad - (\gamma_k + \mu_k) x_k - (x_k + y_k) \lambda_k^* \\ y'_k = \gamma_k x_k - \mu_k y_k \\ z'_k = \delta_k (x_k - z_k). \end{cases}$$

Setting  $x_k = x_k^* \exp(\xi t)$ ,  $y_k = y_k^* \exp(\xi t)$ ,  $z_k = z_k^* \exp(\xi t)$  (where  $\xi \in \mathbb{C}$ ) we get

$$\begin{cases} \xi x_k^* = (1 - i_k^* - r_k^*) \sum_{j=1}^n (\beta_{kj} x_j^* + \hat{\beta}_{kj} z_j^*) \\ \quad - (\gamma_k + \mu_k) x_k^* - (x_k^* + y_k^*) \lambda_k^* \\ \xi y_k^* = \gamma_k x_k^* - \mu_k y_k^* \\ \xi z_k^* = \delta_k (x_k^* - z_k^*). \end{cases} \tag{A.4}$$

In order to prove local asymptotic stability of the strongly endemic equilibrium, we assume by contradiction that  $\Re \xi \geq 0$  and solve for  $y_k^*$  and  $z_k^*$  in (A.4):

$$\begin{aligned} y_k^* &= \frac{\gamma_k / \mu_k}{1 + \xi / \mu_k} x_k^* \\ z_k^* &= \frac{1}{1 + \xi / \delta_k} x_k^*, \end{aligned} \tag{A.5}$$

therefore

$$\begin{aligned} |y_k^*| &\leq \frac{\gamma_k}{\mu_k} |x_k^*| \\ |z_k^*| &\leq |x_k^*|. \end{aligned} \tag{A.6}$$

Now, since from (A.4) we have

$$\begin{aligned} &\left| \left( 1 + \frac{\xi}{\gamma_k + \mu_k} \right) x_k^* + \frac{\lambda_k^*}{\gamma_k + \mu_k} (x_k^* + y_k^*) \right| \\ &\leq (1 - i_k^* - r_k^*) \sum_{j=1}^n \frac{\beta_{kj} |x_j^*| + \hat{\beta}_{kj} |z_j^*|}{\gamma_k + \mu_k}, \end{aligned}$$

substitution of (A.5) and (A.6) into the last inequality yields

$$\begin{aligned} &\left| 1 + \frac{\xi}{\gamma_k + \mu_k} + \frac{\lambda_k^*}{\gamma_k + \mu_k} \psi_k(\xi) \right| |x_k^*| \\ &\leq (1 - i_k^* - r_k^*) \sum_{j=1}^n \frac{\beta_{kj} + \hat{\beta}_{kj}}{\gamma_k + \mu_k} |x_j^*|, \end{aligned} \tag{A.7}$$

where we have defined  $\psi_k(\xi) = 1 + \frac{\gamma_k}{\xi + \mu_k}$ .

Let us suppose that the strongly endemic equilibrium is not locally asymptotically stable: this means that  $\Re \xi \geq 0$ .

Notice that, if  $\Re \xi \geq 0$ , then also  $\Re \psi_k(\xi) > 0$ ; in fact,

$$\begin{aligned} \Re \psi_k(\xi) &= 1 + \Re \left( \frac{\gamma_k}{\xi + \mu_k} \right) \\ &= 1 + \frac{\gamma_k (\Re \xi + \mu_k)}{(\Re \xi + \mu_k)^2 + \Im^2 \xi} \geq 1 > 0. \end{aligned}$$

Similarly to what has been done in the proof of Proposition 3.3, let

$$\varepsilon = \min_{k=1, \dots, n} \Re \left( \frac{\xi}{\gamma_k + \mu_k} + \frac{\lambda_k^*}{\gamma_k + \mu_k} \psi_k(\xi) \right) > 0;$$

since

$$\begin{aligned} &\left| 1 + \frac{\xi}{\gamma_k + \mu_k} + \frac{\lambda_k^*}{\gamma_k + \mu_k} \psi_k(\xi) \right| \\ &\geq \left| \Re \left( 1 + \frac{\xi}{\gamma_k + \mu_k} + \frac{\lambda_k^*}{\gamma_k + \mu_k} \psi_k(\xi) \right) \right| \geq 1 + \varepsilon, \end{aligned}$$

from (A.7) we get

$$(1 + \varepsilon) |x_k^*| \leq (1 - i_k^* - r_k^*) \sum_{j=1}^n \frac{\beta_{kj} + \hat{\beta}_{kj}}{\gamma_k + \mu_k} |x_j^*|.$$

Let  $\eta = \max_{k=1, \dots, n} \frac{|x_k^*|}{w_k}$ , where  $w = (w_1, \dots, w_n)$  is a solution to  $w_k = (1 - i_k^* - r_k^*) \sum_{j=1}^n \frac{\beta_{kj} + \hat{\beta}_{kj}}{\gamma_k + \mu_k} |w_j^*|$  and  $w_k > 0$  for  $k = 1, \dots, n$ .

The equilibrium is strongly endemic, so at least one of the  $x_k^*$  is strictly positive, and this means that  $\eta > 0$ ; moreover, for  $k = 1, \dots, n$ ,  $|x_k^*| \leq \eta w_k$  and there exists  $\bar{k}$  such that  $|x_{\bar{k}}^*| = \eta w_{\bar{k}}$ . Consequently,

$$\begin{aligned} (1 + \varepsilon) |x_{\bar{k}}^*| &\leq (1 - i_{\bar{k}}^* - r_{\bar{k}}^*) \sum_{j=1}^n \frac{\beta_{\bar{k}j} + \hat{\beta}_{\bar{k}j}}{\gamma_{\bar{k}} + \mu_{\bar{k}}} |x_j^*| \\ &\leq (1 - i_{\bar{k}}^* - r_{\bar{k}}^*) \sum_{j=1}^n \frac{\beta_{\bar{k}j} + \hat{\beta}_{\bar{k}j}}{\gamma_{\bar{k}} + \mu_{\bar{k}}} \eta w_j \\ &= \eta w_{\bar{k}} \\ &= |x_{\bar{k}}^*|; \end{aligned}$$

but this contradicts the fact that  $\varepsilon > 0$ . Thus, the strongly endemic equilibrium is locally asymptotically stable.

**References**

Abad, F.X., Pinto, R.M., Bosch, A., 1994. Survival of enteric viruses on environmental fomites. *Appl. Environ. Microbiol.* 60 (10), 3704–3710.  
 Ajelli, M., Iannelli, M., Manfredi, P., Ciofi degli Atti, M.L., 2008. Basic mathematical models for the temporal dynamics of HAV in medium-endemicity Italian areas. *Vaccine* 26 (13), 1697–1707.  
 Ajelli, M., Merler, S., 2008. The impact of the unstructured contacts component in influenza pandemic modeling. *PLoS One* 3 (1), e1519.  
 Ajelli, M., Merler, S., 2009. An individual-based model of hepatitis A transmission. *J. Theor. Biol.* 259 (3), 478–488.

- Amariei, R., Willms, A.R., Bauch, C.T., 2008. The United States and Canada as a coupled epidemiological system: an example from hepatitis A. *BMC Infect. Dis.* 8, 23.
- Ansaldi, F., Bruzzone, B., Rota, M.C., Bella, A., Ciofi degli Atti, M.L., Durando, P., Gasparini, R., Icardi, G., 2008. Hepatitis A incidence and hospital-based seroprevalence in Italy: a nation-wide study. *Eur. J. Epidemiol.* 23 (1), 45–53.
- Armstrong, G.L., Bell, B.P., 2002. Hepatitis A virus infection in the United States: model-based estimates and implications for childhood immunization. *Pediatrics* 109, 839–845.
- Averhoff, F., Shapiro, C.N., Bell, B.P., Hyams, I., Burd, L., Deladisma, A., Simard, E.P., Nalin, D., Kuter, B., Ward, C., Lundberg, M., Smith, N., Margolis, H.S., 2001. Control of hepatitis A through routine vaccination of children. *J. Am. Med. Assoc.* 286 (23), 2968–2973.
- Balcan, D., Hu, H., Gonçalves, B., Bajardi, P., Poletto, C., Ramasco, J., Paolotti, D., Perra, N., Tizzoni, M., Broeck, W., Colizza, V., Vespignani, A., 2009. Seasonal transmission potential and activity peaks of the new influenza A(H1N1): a Monte Carlo likelihood analysis based on human mobility. *BMC Med.* 7, 45.
- Bauch, C.T., Srinivasa Rao, A.S., Pham, B.Z., Krahn, M., Gilca, V., Duval, B., Chen, M.H., Tricco, A.C., 2007. A dynamic model for assessing universal hepatitis A vaccination in Canada. *Vaccine* 25 (10), 1719–1726.
- Biziagos, E., Passagot, J., Crance, J.M., Deloince, R., 1988. Long-term survival of hepatitis A virus and poliovirus type 1 in mineral water. *Appl. Environ. Microbiol.* 54 (11), 2705–2710.
- CDC, 2007. Health information for international travel 2008. Atlanta, US. Department of Health and Human Services, Public Health Service. Centers for Disease Control and Prevention.
- Chitambar, S.D., Chadha, M.S., Yeolekar, L.R., Arankalle, V.A., 1996. Hepatitis A in day care centre. *Indian J. Pediatr.* 63 (6), 781–783.
- 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 (3), e1790.
- Codeço, C.T., 2001. Endemic and epidemic dynamics of cholera: the role of the aquatic reservoir. *BMC Infect. Dis.* 1, 1.
- Colizza, V., Barrat, A., Barthelemy, M., Vespignani, A., 2007. Predictability and epidemic pathways in global outbreaks of infectious diseases: the SARS case study. *BMC Med.* 5, 34.
- Das, A., 2003. An economic analysis of different strategies of immunization against hepatitis A virus in developed countries. *Hepatology* 29 (2), 548–552.
- Fiore, A.E., 2004. Hepatitis A transmitted by food. *Clin. Infect. Dis.* 38, 705–715.
- Galil, K., Lee, B., Strine, T., Carraher, C., Baughman, A.A., Eaton, M., Montero, J., Seward, J., 2002. Outbreak of varicella at a day-care center despite vaccination. *New Engl. J. Med.* 347 (24), 1909–1915.
- Grenfell, B.T., Bjørnstad, O.N., Kappey, J., 2001. Travelling waves and spatial hierarchies in measles epidemics. *Nature* 414 (6865), 716–723.
- Hastie, T., Tibshirani, R., Friedman, J., 2001. *The Elements of Statistical Learning*. Springer.
- Hethcote, H.W., 1978. An immunization model for a heterogeneous population. *Theor. Popul. Biol.* 14 (3), 338–349.
- Hethcote, H.W., Thieme, H.R., 1985. Stability of the endemic equilibrium in epidemic models with subpopulations. *Math. Biosci.* 75, 205–227.
- Hoos, H.H., Stutzle, T., 2005. *Stochastic Local Search: Foundations and Applications*. Morgan Kaufmann Publishers.
- Iannelli, M., Manfredi, P., 2007. Demographic change and immigration in age-structured epidemic models. *Math. Popul. Stud.* 14 (3), 169–191.
- ISTAT, 2001. XIV censimento generale della popolazione e delle abitazioni. Provided by the Italian Institute of Statistics at: <http://dawinci.istat.it/MD/> (in Italian).
- ISTAT, 2002. Statistiche del turismo. Provided by the Italian Institute of Statistics at: [http://www.istat.it/dati/catalogo/20021009\\_00/](http://www.istat.it/dati/catalogo/20021009_00/) (in Italian).
- ISTAT, 2006. *Annuario statistico Italiano*. Provided by the Italian Institute of Statistics. Years 1980–2006 (in Italian).
- ISTAT, 2007. Life tables. Provided by the Italian Institute of Statistics at: <http://demo.istat.it/unitav/index.html?lingua=eng> (Years 1974–2007).
- ISTAT, 2009a. Indicatori demografici. Provided by the Italian Institute of Statistics at: <http://demo.istat.it/altridati/indicatori/index.html> (in Italian).
- ISTAT, 2009b. Popolazione residente. Provided by the Italian Institute of Statistics at: <http://demo.istat.it/pop2009/index.html> (in Italian).
- Keeling, M.J., Gilligan, C.A., 2000. Metapopulation dynamics of bubonic plague. *Nature* 407 (6806), 903–906.
- Koopman, J.S., Chick, S.E., Simon, C.P., Riolo, C.S., Jacquez, G., 2002. Stochastic effects on endemic infection levels of disseminating versus local contacts. *Math. Biosci.* 180 (1–2), 49–71.
- Lopalco, P.L., Malfait, P., Menniti-Ippolito, F., Prato, R., Germinario, C., Chironna, M., Quarto, M., Salmaso, S., 2005. Determinants of acquiring hepatitis A virus disease in a large Italian region in endemic and epidemic periods. *J. Viral Hepatitis* 12 (3), 315–321.
- Lucioni, C., Cipriani, V., Mazzi, S., Panunzio, M., 1998. Cost of an outbreak of hepatitis A in Puglia, Italy. *Pharmacoeconomics* 13 (2), 257–266.
- Martinelli, D., Bitetto, I., Tafuri, S., Lopalco, P.L., Mininni, R.M., Prato, R., 2010. Control of hepatitis A by universal vaccination of children and adolescents: an achieved goal or a deferred appointment? *Vaccine* 28 (41), 6783–6788.
- Mbithi, J.N., Springthorpe, V.S., Sattar, S.A., 1991. Effect of relative humidity and air temperature on survival of hepatitis A virus on environmental surfaces. *Appl. Environ. Microbiol.* 57 (5), 1394–1399.
- Mele, A., Tosti, M.E., Spada, E., Mariano, A., Bianco, E., Group, S.C., 2006. Epidemiology of acute viral hepatitis: twenty years of surveillance through SEIEVA in Italy and a review of the literature. *ISTISAN Rep.* 12 (6), 1–39.
- Merler, S., Ajelli, M., 2010. The role of population heterogeneity and human mobility in the spread of pandemic influenza. *Proc. R. Soc. B* 277 (1681), 557–565.
- Murray, C.J.L., Lopez, A.D., 1997. Mortality by cause for eight regions of the world: global burden of disease study. *Lancet* 349, 1269–1276.
- Roche, B., Lebarbenchon, C., Gauthier-Clerc, M., Chang, C.M., Thomas, F., Renaud, F., van der Werf, S., Guégan, J.F., 2009. Water-borne transmission drives avian influenza dynamics in wild birds: the case of the 2005–2006 epidemics in the Camargue area. *Infect. Genet. Evol.* 9 (5), 800–805.
- Rohani, P., Breban, R., Stallknecht, D.E., Drake, J.M., 2009. Environmental transmission of low pathogenicity avian influenza viruses and its implications for pathogen invasion. *Proc. Natl. Acad. Sci. USA* 106 (25), 10365–10369.
- Rvachev, L.A., Longini, I.M.J., 1985. A mathematical model for the global spread of influenza. *Math. Biosci.* 75 (1), 3–22.
- Salamina, G., D'Argenio, P., 1998. Shellfish consumption and awareness of risk of acquiring hepatitis A among Neapolitan families—Italy 1997. *Euro Surveill.* 3 (10).
- Samandari, T., Bell, B.P., Armstrong, G.L., 2004. Quantifying the impact of hepatitis A immunization in the United States, 1995–2001. *Vaccine* 22, 4342–4350.
- Sattenspiel, L., Dietz, K., 1995. A structured epidemic model incorporating geographic mobility among regions. *Math. Biosci.* 128, 71–91.
- Sattenspiel, L., Simon, C.P., 1988. The spread and persistence of infectious diseases in structured populations. *Math. Biosci.* 90, 341–366.
- Srinivasa Rao, A.S., Chen, M.H., Pham, B.Z., Tricco, A.C., Gilca, V., Duval, B., Krahn, M.D., Bauch, C.T., 2006. Cohort effects in dynamic models and their impact on vaccination programmes: an example from hepatitis A. *BMC Infect. Dis.* 6, 174.
- Stapleton, J.T., Lemon, S.M., 1994. *Infectious Diseases*. Lippincott Co., Philadelphia, US, pp. 790–797 (Chapter Hepatitis A and hepatitis E).
- Thieme, H.R., 2003. *Mathematics in Population Biology*. Princeton University Press.
- Viboud, C., Bjørnstad, 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 (5772), 447–451.
- Woolhouse, M.E.J., Stringer, S.M., Matthews, L., Hunter, N., Anderson, R.M., 1998. Epidemiology and control of scrapie within a sheep flock. *Proc. R. Soc. B* 265, 1205–1210.