

# INTERFACE

[rsif.royalsocietypublishing.org](http://rsif.royalsocietypublishing.org)

Research



CrossMark  
click for updates

**Cite this article:** te Beest DE, Birrell PJ, Wallinga J, De Angelis D, van Boven M. 2015 Joint modelling of serological and hospitalization data reveals that high levels of pre-existing immunity and school holidays shaped the influenza A pandemic of 2009 in The Netherlands. *J. R. Soc. Interface* **12**: 20141244.  
<http://dx.doi.org/10.1098/rsif.2014.1244>

Received: 10 November 2014

Accepted: 4 December 2014

## Subject Areas:

biomathematics, biometrics, biocomplexity

## Keywords:

Bayesian evidence synthesis, influenza A, serology, hospitalization incidence, mixture analysis, transmission model

## Author for correspondence:

Michiel van Boven

e-mail: [michiel.van.boven@rivm.nl](mailto:michiel.van.boven@rivm.nl)

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rsif.2014.1244> or via <http://rsif.royalsocietypublishing.org>.

**THE ROYAL SOCIETY**  
PUBLISHING

# Joint modelling of serological and hospitalization data reveals that high levels of pre-existing immunity and school holidays shaped the influenza A pandemic of 2009 in The Netherlands

Dennis E. te Beest<sup>1</sup>, Paul J Birrell<sup>2</sup>, Jacco Wallinga<sup>1</sup>, Daniela De Angelis<sup>2</sup> and Michiel van Boven<sup>1</sup>

<sup>1</sup>Centre for Infectious Disease Control, National Institute for Public Health and the Environment, PO Box 1, Bilthoven 3720AB, The Netherlands

<sup>2</sup>MRC Biostatistics Unit, Cambridge Institute of Public Health, CB2 0SR

Obtaining a quantitative understanding of the transmission dynamics of influenza A is important for predicting healthcare demand and assessing the likely impact of intervention measures. The pandemic of 2009 provides an ideal platform for developing integrative analyses as it has been studied intensively, and a wealth of data sources is available. Here, we analyse two complementary datasets in a disease transmission framework: cross-sectional serological surveys providing data on infection attack rates, and hospitalization data that convey information on the timing and duration of the pandemic. We estimate key epidemic determinants such as infection and hospitalization rates, and the impact of a school holiday. In contrast to previous approaches, our novel modelling of serological data with mixture distributions provides a probabilistic classification of individual samples (susceptible, immune and infected), propagating classification uncertainties to the transmission model and enabling serological classifications to be informed by hospitalization data. The analyses show that high levels of immunity among persons 20 years and older provide a consistent explanation of the skewed attack rates observed during the pandemic and yield precise estimates of the probability of hospitalization per infection (1–4 years: 0.00096 (95%CrI: 0.00078–0.0012); 5–19 years: 0.00036 (0.00031–0.0044); 20–64 years: 0.0015 (0.00091–0.0020); 65+ years: 0.0084 (0.0028–0.016)). The analyses suggest that in The Netherlands, the school holiday period reduced the number of infectious contacts between 5- and 9-year-old children substantially (estimated reduction: 54%; 95%CrI: 29–82%), thereby delaying the unfolding of the pandemic in The Netherlands by approximately a week.

## 1. Introduction

Worldwide, influenza A causes considerable morbidity and mortality in years with high influenza activity [1,2]. Proper assessment of the epidemiological dynamics is key for effective control. However, it is not uncommon that different data sources yield conflicting information [3]. For instance, influenza-like illness surveillance through networks of general practitioners showed increases in incidence in many countries well before increases in seropositivity and virus isolation rates, possibly because of increased public awareness. The advent of modern statistical methods combined with explosive increases in computing power has enabled systematic integration of different data sources in unifying statistical frameworks, thereby providing proper weighting of the various types of data.

© 2014 The Authors. Published by the Royal Society under the terms of the Creative Commons Attribution License <http://creativecommons.org/licenses/by/4.0/>, which permits unrestricted use, provided the original author and source are credited.

In this paper, we analyse the transmission dynamics of the influenza pandemic of 2009 by linking a dynamic transmission model to serological and hospitalization data in a Bayesian inferential framework. The serological data provide information on levels of immunity and infection attack rates, and the hospitalization data give an indication of the timing and severity of the epidemic in different age groups. The model classifies persons in one of four stages of infection (susceptible, exposed, infectious and recovered) and describes the transmission dynamics of influenza A in an age-structured population [3–5]. In the model, contacts between persons of different age are made at rates that are determined by observed human contact patterns [6].

The serological data consists of two cross-sectional surveys involving random samples from the Dutch population, one performed just before and the other just after the pandemic [7,8]. As in an earlier study, we model these data with mixture models in which each sample has a certain probability of belonging to a person who was susceptible, previously exposed and immune, or recently infected [7]. This has a distinct advantage over the more rigid classification of persons as being either susceptible, immune or infected as it takes into account biological variation in antibody concentrations. In the mixture modelling approach, there is no need to set a specific threshold for classification, even though optimal thresholds can be derived from the data [9]. Consequently, not only do the serological data inform the hospitalization data, but the hospitalization data also inform classification of serological samples. In other words, probabilistic classification of serological samples is determined by the information contained in both the serological and hospitalization data.

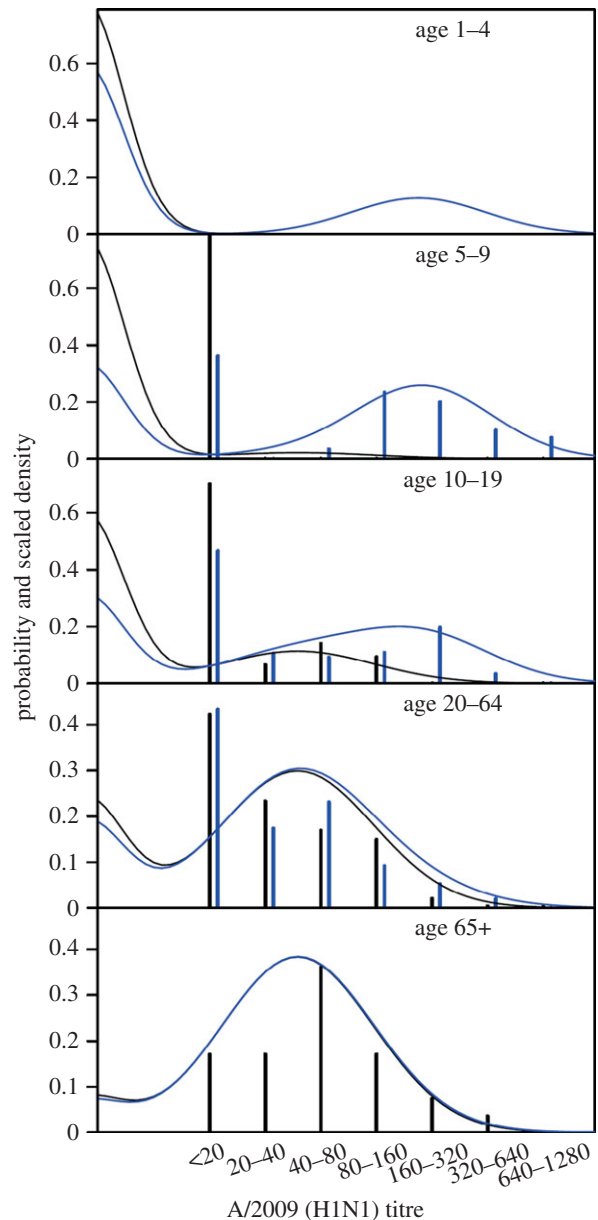
Based on the high contact rates of children, we expect an epidemic to peak first in children [5,6,10]. An unusual pattern observed in the Dutch hospitalization data for the 2009 pandemic is that the epidemic peak in young children occurred relatively late. This may have been caused by the one-week holiday in weeks 43–44, a couple of weeks before the pandemic peak. It is known that school holidays can considerably reduce the number of contacts made by children, thereby reducing transmission [11,12]. We use the inferential framework sketched above to estimate age-specific infection and hospitalization rates and to investigate to what extent the delayed peak in young children can be explained by school holidays.

## 2. Material and methods

### 2.1. Serological data

In The Netherlands, two cross-sectional serological surveys were conducted before and after the pandemic of 2009 [7]. In this study, samples were tested by a haemagglutination inhibition test (HI) to estimate age-specific infection attack rates. A structured random subset of the samples was subsequently tested against a panel of antigens using an antibody protein micro array (figure 1) [8]. The protein micro array is a novel diagnostic assay to investigate antibody responses to subunit 1 of the haemagglutinin surface glycoprotein (HA1) [13–16]. Analyses show that the micro array is more sensitive and more specific than HI in distinguishing recent infection from prior exposure [8,13]. Infection attack rates estimated with the micro array are similar to those estimated with HI, but are more precise for a given sample size [8].

A total of 357 people were tested with the micro array; 167 in the pre-pandemic study and 190 in the post-pandemic study. The age of the participants ranged from 5 to 75 years. Data were stratified



**Figure 1.** Overview of the serological data (bars) and fit of the model with school holiday (lines). Panels show results for the various age groups. Bars show the serological data aggregated in titre classes (<20, 20–40, 40–80, 80–160, 160–320 and 320–640). Black bars and black lines denote pre-pandemic data and pre-pandemic model fit, respectively. Blue bars and lines show the post-pandemic data and post-pandemic model fit. Note that no serological data are available in young children (1–4 years) and that only pre-pandemic data are included in the oldest age group (65+ years).

according to the following age categories: 5–9, 10–19, 20–64 and 65+ years, following recommendations of the Consortium for the Standardization of Influenza Seroepidemiology (consise.tghn.org), with the exception that the age groups 20–44 and 45–64 years were aggregated. The samples in the pre-pandemic subset were collected before 12 October 2009, i.e. well before the onset of widespread influenza circulation [7,8,17]. The serological data are available in the electronic supplementary material.

The post-pandemic survey was carried out several months after the pandemic period; and from these data, we excluded people who were vaccinated against the pandemic virus [7]. Exclusion of persons from the post-pandemic survey who had a pandemic vaccination also led to the exclusion of many persons with a history of seasonal vaccinations [8]. Due the underrepresentation of persons with a history of seasonal vaccination in the

post-pandemic sample in the oldest age group (65+ years), this subset of the data does not represent a random sample from the population and was not included in the analyses [8]. As a result, serological data in the oldest age group cannot directly be used to estimate the attack rate. However, the pre-pandemic sample of the oldest age group does represent a random sample of the population and was used in the statistical analyses.

## 2.2. Hospitalization data

During the pandemic, all hospitals in The Netherlands were required to notify the municipal health services of hospital admissions of patients with a laboratory-confirmed influenza A infection. Our hospitalization data represent the total, nationwide daily number of hospitalizations, i.e. the daily number summed over all municipal health services [17]. We use the same age stratification for the hospitalization data as for the serological data (5–9, 10–19, 20–64 and 65+ years), supplemented with children aged 1–4 years. Upon admittance to a hospital, persons were asked when their symptoms had started. The number of hospitalized influenza cases by day of symptom onset provides direct information on the epidemic curve. A total of 1610 cases was available for analysis, distributed over the age categories as follows: 1–4 years: 267 cases; 5–9 years: 196 cases; 10–19 years: 225 cases; 20–64 years: 800 cases; and 65+ years: 122 cases. The average time between onset of symptoms and admittance to the hospital was 2.4 days.

## 2.3. Transmission model

The hospitalization and serological data are linked through a Susceptible–Exposed–Infectious–Removed transmission model [5], thereby providing a natural weighting of the different types of data. In the model, the exposed and infectious periods are modelled using Erlang (gamma) distributions, yielding control over variation in the exposed and infectious stages [5,18]. Specifically, we include four exposed and four infectious stages, so that distributions of time in the exposed and infectious classes have means  $1/\eta$  and  $1/\gamma$ , and shape parameters  $n_E = n_I = 4$ , respectively.

In each of the five age groups, individuals make contact with individuals in other age groups at rates that are proportional to a mixing matrix  $C$ , which is specified by observed human contact patterns [6]. As in earlier analyses, a proportionality parameter  $z$  reflects the probability of transmission per contact. Hence, the basic reproduction number in a susceptible population and the reproduction number in a population with pre-existing immunity are given by the spectral radiuses of  $zC$  and  $zCS$ , respectively, where age-specific proportions that are initially susceptible are collected in the matrix  $S$  [19].

Further, if we denote by  $R_a$  the relative frequencies of removed individuals in age group  $a$ , and by  $E_{a,j}$  and  $I_{a,j}$  the relative frequencies of exposed and infectious individuals in age group  $a$  of stage  $j$  (see below), then the model dynamics is specified by the following differential equations:

$$\left. \begin{aligned} \frac{dS_a}{dt} &= -zS_a \sum_b c_{ab} I_b^* \\ \frac{dE_{a,1}}{dt} &= zS_a \sum_b c_{ab} I_b^* - n_E \eta E_{a,1}, \\ \frac{dE_{a,j}}{dt} &= n_E \eta E_{a,j-1} - n_E \eta E_{a,j}, \\ \frac{dI_{a,1}}{dt} &= n_E \eta E_{a,n_E} - n_I \gamma I_{a,1}, \\ \frac{dI_{a,j}}{dt} &= n_I \gamma I_{a,j-1} - n_I \gamma I_{a,j} \end{aligned} \right\} \quad (2.1)$$

and

$$\frac{dR_a}{dt} = n_I \gamma I_{a,n_I}$$

where the index  $a$  runs through all age groups, the index  $j$  runs through all exposed and infectious stages except for the first ( $j = 2, \dots, n_E$  for  $dE_{i,j}/dt$ , and  $j = 2, \dots, n_I$  for  $dI_{i,j}/dt$ ) and  $I_b^* = \sum_j I_{b,j}$  is the total relative frequency of infectious individuals in age group  $b$ . In line with empirical evidence, the average lengths of the exposed and infectious periods are both set to 1.5 days [20]. In this manner, the generation time is  $1/\eta + 1/\gamma (n_I + 1)/2n_I = 2.4$  days [21].

The fractions of immune persons at the start and end of the pandemic are given by  $R_a(0)$  and  $R_a(\infty)$ , and these quantities are estimated in age groups with serological data (5–9, 10–19, 20–64 and 65+ years). No serological data are available in the youngest age group (1–4 years). It is, however, plausible that there was very little pre-existing immunity in this age group, so we take  $R_{1-4}(0) = 0$ . The fractions of immune persons at the start and end of the pandemic feed into the mixture model (see below).

## 2.4. Hospitalizations likelihood

The hospitalization incidence data provide the timing of symptoms onset for confirmed influenza A infections requiring hospitalization. Making the reasonable assumption that the onset of symptoms is close to the point at which people become infectious [20], we can directly relate the incidence  $i_a(t) = n_E \eta E_{a,n_E}(t)$  in age group  $a$  at time  $t$  to the hospitalization data. Alternative assumptions, e.g. assuming that the onset of symptoms occurred halfway through or at the end of the infectious period (i.e. assuming  $i_a(t) = n_E \eta I_{a,2}(t)$  or  $i_a(t) = n_E \eta I_{a,4}(t)$ ) result in a minor shift of onset of symptoms and yield virtually identical results (not shown). Assuming further that the expected number of hospitalizations is proportional to the numbers of infections, the expected number of hospitalizations at time  $t$  in age group  $a$  is given by

$$\mu_a(t) = \psi_a N_a i_a(t). \quad (2.2)$$

where  $N_a$  denotes the number of individuals in age group  $a$ , and  $\psi_a$  is the age-dependent probability of hospitalization per infection. Thus, the log likelihood of the hospitalization data  $H_{t,a}$  is given by

$$\ell_{\text{hosp}} = \sum_t \sum_a \log [g(H_{t,a} | \mu_a(t))], \quad (2.3)$$

where the indices  $t$  and  $i$  run through all observation periods (i.e. weeks 41–51) and age groups (1–4, 5–9, 10–19, 20–64 and 65+ years), and  $g(H_{t,a} | \mu_a(t))$  is the probability density function of the hospitalizations. Throughout, we assume that hospitalizations are Poisson distributed.

## 2.5. Serology likelihood

The micro array data are fitted with a Gaussian mixture model that consists of three normally distributed components for the log-transformed serological data. The first and second components correspond to samples with antibodies present in low and intermediate concentrations, representing susceptible individuals and individuals with pre-existing immunity, respectively. The third component corresponds to samples with high antibody concentrations, consistent with recent A/2009 (H1N1) infection. Motivated by biological considerations, we fit the distributions of persons who were susceptible or had pre-existing immunity to the pre-pandemic data, and all three distributions to the post-pandemic data [7,8]. For each distribution, we estimate a mean and standard deviation  $\theta_i = (\mu_i, \sigma_i)$ . Densities of the distributions are denoted by  $f_i(x; \theta_i)$  or simply  $f(x; \theta_i)$ .

Age-specific mixing parameters are collected in vectors  $\mathbf{p}$  and  $\mathbf{q}$ . Here,  $p_a$  represents the fraction of the population in age group  $a$  that has pre-existing immunity, and  $q_a$  is the fraction in age group  $a$  that has been infected. Hence,  $1 - p_a$  and  $1 - p_a - q_a$  are the corresponding pre- and post-pandemic fractions of susceptible persons in age group  $i$ . Further, we denote by  $g_i$  the age label of sample  $i$ , by  $n_{\text{pre}}$  and  $n_{\text{post}}$  the number of samples

**Table 1.** Parameter estimates of the model that does not include the school holiday effect. Parameter estimates are represented by the medians of the posterior distribution.

parameter	age group (years)	estimate	(95% CrI)
fraction immune before the pandemic	1–4	0 <sup>a</sup>	
	5–9	0.07	(0.00–0.23)
	10–19	0.25	(0.16–0.37)
	20–64	0.70	(0.61–0.76)
	65+	0.90	(0.76–0.95)
infection attack rate	1–4	0.22	(0.20–0.25)
	5–9	0.48	(0.40–0.52)
	10–19	0.31	(0.25–0.37)
	20–64	0.05	(0.04–0.07)
	65+	0.01	(0.00–0.02)
basic reproduction number		1.9	(1.8–2.3)
reproduction number at the start of the pandemic		1.31	(1.29–1.33)
probability of hospitalization	1–4	0.0012	(0.0010–0.0014)
	5–19	0.00040	(0.00034–0.00048)
	20–64	0.0017	(0.0011–0.0022)
	65+	0.010	(0.0037–0.018)

<sup>a</sup>Pre-pandemic immunity is assumed to be absent in young children (1–4 years).

in the pre- and post-pandemic surveys, by  $d_{pre,i}$  and  $d_{post,i}$  the  $\log_2$  antibody titres in the pre- and post-pandemic studies and by  $w_i$  the contribution of sample  $i$  to the population census [7,8]. With these notational conventions, the log likelihood of the pre-pandemic data is given by

$$\ell_{pre}(\theta_{sus}, \theta_{imm}, p | d_{pre}) = \sum_{i=1}^{n_{pre}} w_i \log((1 - p_{g_i})f(d_{pre,i}; \theta_{sus}) + p_{g_i}f(d_{pre,i}; \theta_{imm})). \quad (2.4)$$

In similar fashion, the log likelihood of the post-pandemic data is given by

$$\begin{aligned} \ell_{post}(\theta_{sus}, \theta_{imm}, \theta_{inf}, p, q | d_{post}) \\ = \sum_{i=1}^{n_{post}} w_i \log((1 - p_{g_i} - q_{g_i})f(d_{post,i}; \theta_{sus}) \\ + p_{g_i}f(d_{post,i}; \theta_{imm}) + q_{g_i}f(d_{post,i}; \theta_{inf})). \end{aligned} \quad (2.5)$$

Note that the fraction of persons with pre-existing immunity in the pre-pandemic study equals the fraction of persons with pre-existing immunity in the post-pandemic study. This assumption is reasonable, given that there was only a short-time span between the two studies (less than six months) making significant loss of immunity unlikely. Combining the above, the log likelihood of the serological data is given by  $\ell_{pre} + \ell_{post}$  and the total likelihood of the joint hospitalization data and serological data is given by  $\ell_{hosp} + \ell_{pre} + \ell_{post}$ .

The age-dependent weights of the three mixture distributions have an epidemiological interpretation, representing the fractions of the population that were initially susceptible, had pre-existing immunity and had been infected (equation (2.5)). The weights are linked to the epidemic model via equation (2.1) as follows. In the epidemic model, the fractions of susceptible and immune persons before the epidemic are given by  $S_a(0)$  and  $R_a(0)$ , the number of immune and infected persons at the end of the pandemic is given by  $R_a(\infty)$ , and the attack rate is given by  $R_a(\infty) - R_a(0)$ . In the mixture model, we thus take  $p_a = R_a(0)$  and

$q_a = R_a(\infty) - R_a(0)$  (see the electronic supplementary material for details).

## 2.6. School holiday

In The Netherlands, there is a week-long autumn school holiday, which was planned in week 43 (estimated at 71% of the population) or week 44 (29% of the population). We use those estimates in the analyses with school holiday effect. The 5–9 and 10–19 year-old age groups consist largely of school-going children. The number of contacts within these age groups are discounted during the two weeks of school holiday by replacing the elements  $c_{ii}$  in the contact matrix by  $v_t r_i c_{ii} + (i - v_t) c_{ii}$ . Here,  $v_t$  represents the fraction of children that have a holiday in week  $t$ , and  $r_i$  represents the contact reduction during the holiday. We estimate separate contact reductions  $r_i$  for the two age groups 5–9 and 10–19 years.

## 2.7. Estimation

We considered two main scenarios, one without and the other with the school holiday effect. The models have 16 and 18 parameters to be estimated, respectively (see tables 1 and 2, and the electronic supplementary material, tables S2 and S3). We use a Bayesian framework to enable flexible incorporation of prior information. The parameters specifying the fractions infected and the fractions immune at the start of the epidemic ( $R_a(0)$ ), specifying the impact of the school holiday ( $r_a$ ) and determining the fraction of the population at the start of the pandemic are constrained to the domain  $[0,1]$ , and we use Jeffrey's prior distributions for these parameters. The parameters  $\theta_{sus}$ ,  $\theta_{imm}$  and  $\theta_{inf}$  are constrained to the domain  $[0, \infty]$ , and we assume (improper) uniform prior distributions for these parameters. The hospitalization probabilities ( $\psi_a$ ) are also constrained to the domain  $[0, 1]$ . It is plausible that the  $\psi_a$  should not deviate excessively between adjacent age categories. Therefore, and using information from preliminary analyses,  $\psi_a$  is estimated jointly for the age groups 5–9 and 10–19 years, while  $\psi_{1-4}$ ,  $\psi_{5-9}$  and

**Table 2.** Parameter estimates of the model that includes a potential reduction in transmission during the school holiday. Parameter estimates are represented by the medians of the posterior distribution.

parameter	age group (years)	estimate	(95% CrI)
fraction immune before the pandemic	1–4	0 <sup>a</sup>	
	5–9	0.08	(0.00–0.24)
	10–19	0.29	(0.20–0.41)
	20–64	0.72	(0.63–0.78)
	65+	0.91	(0.75–0.95)
infection attack rate	1–4	0.28	(0.24–0.33)
	5–9	0.53	(0.43–0.58)
	10–19	0.34	(0.27–0.40)
	20–64	0.05	(0.04–0.09)
	65+	0.01	(0.00–0.02)
reduction of transmission during school holiday	5–9	0.54	(0.29–0.82)
	10–19	0.10	(0.00–0.29)
basic reproduction number		2.2	(2.0–2.6)
reproduction number at the start of the pandemic		1.42	(1.37–1.48)
probability of hospitalization	1–4	0.00096	(0.00078–0.0012)
	5–19	0.00036	(0.00031–0.00044)
	20–64	0.0015	(0.00091–0.0020)
	65+	0.0084	(0.0028–0.016)

<sup>a</sup>Pre-pandemic immunity is assumed to be absent in young children (1–4 years).

$\psi_{20-64}$  are not allowed to deviate more than a factor of five from each other. For the oldest age group,  $\psi_{65+}$  is not allowed to deviate more than 10-fold from  $\psi_{20-64}$ . In a sensitivity analysis, we allow these deviations to be doubled to a 10- and 20-fold maximal difference.

Estimates of the parameters are obtained in a Markov chain Monte Carlo framework with a random-walk Metropolis algorithm using Gaussian proposal distributions and with standard deviations set to achieve acceptance ratios in the range 0.20 to 0.35 (see the electronic supplementary material). A 1/10 thinned sample of 200 000 is obtained from the posterior distribution, after a burn-in of 10 000 cycles. Inspection of convergence and mixing is assessed visually. All programs are coded in R 3.1.

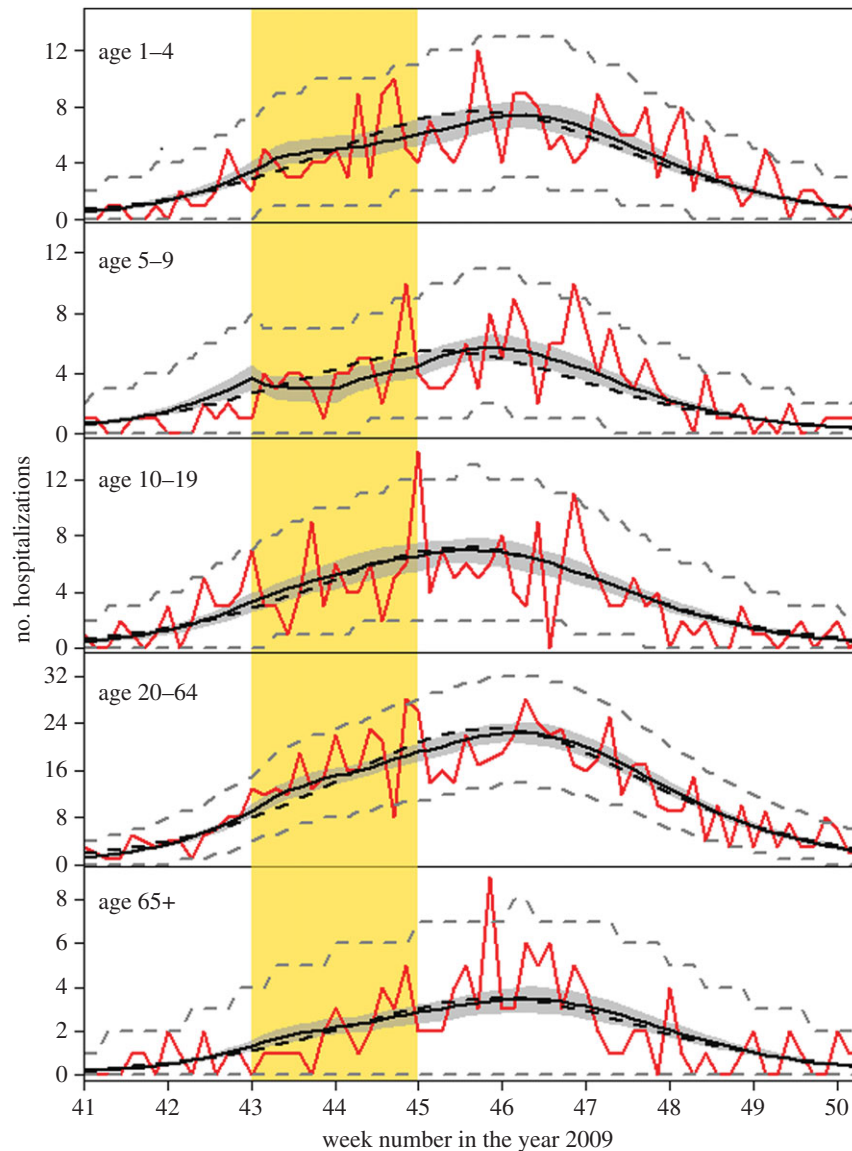
### 3. Results

Figures 1 and 2 give an overview of data and model fits, and parameter estimates of the models with and without school holiday effect are given in tables 1 and 2 and electronic supplementary material, tables S2 and S3. Visual inspection of the hospitalization data and model fits shows that both models adequately describe hospitalizations over time in the different age groups, with no systematic deviations of the predicted from the observed hospitalizations, and no evidence of overdispersion relative to the Poisson model (figure 2). A formal comparison of the two models is made using the Bayes factor (BF), which represents the relative strength of evidence for the model with hospitalization data. To this end, marginal likelihoods are estimated using the harmonic means of the posterior likelihood values of the competing models, yielding stable log-likelihood estimates of  $-1087.8$  and  $-1096.5$ . Hence, the comparison indicates that the model with holiday effect is strongly supported by

the data [22]. This is due mainly to the fact that the model with holiday effect is able to capture the lagging number of hospitalizations in the ascending phase of the epidemic in 5- to 9-year-old children in weeks 43–44. The predicted contact reduction among children in this age groups is 54% (95% CrI: 29–82%). We do not find a significant reduction among children aged 10–19 years during the school holiday (median reduction 10%, 95% CrI: 0–29%). Incorporating the school holiday not only improved the model fit in children, but also seemed to improve the model fit for the other age groups.

To investigate the impact of the autumn school holiday on the epidemic dynamics, we run the transmission model without the school holiday effect with parameters taken from the posterior distribution of the model with school holiday effect. The ensemble dynamics is presented in figure 3. The school holiday appears to have delayed the epidemic peak by approximately one week. Specifically, the estimated delay is 6.9 days (95%CrI: 3.9–9.4) in 1- to 4-year-old children, 7.9 days (95%CrI: 4.9–10.4) in 5- and 9-year-old children, 3.0 days (95%CrI: 0.2–6.2) in 10- and 19-year-old children, 5.4 days (95%CrI: 3.0–7.9) in adults (20–64 years) and 5.5 days (95%CrI: 3.0–7.9) in the elderly (65+ years). Further, our estimates indicate that the holiday period has lowered the epidemic peak by 27% (CrI: 17–35%).

Inspection of the immunity estimates reveal that a large proportion of the adult population had pre-existing immunity and that the high levels of immunity can explain the skewed attack rates (tables 1 and 2). In fact, estimates of the levels of immunity and infection attack rates are strongly correlated (figure 4). Our model also provides estimates of the attack rate in young children (1–4 years), even though no serological data are available for this age group. Specifically, for this age group, the median posterior attack rate is 0.28 (95%CrI: 0.24–0.33).



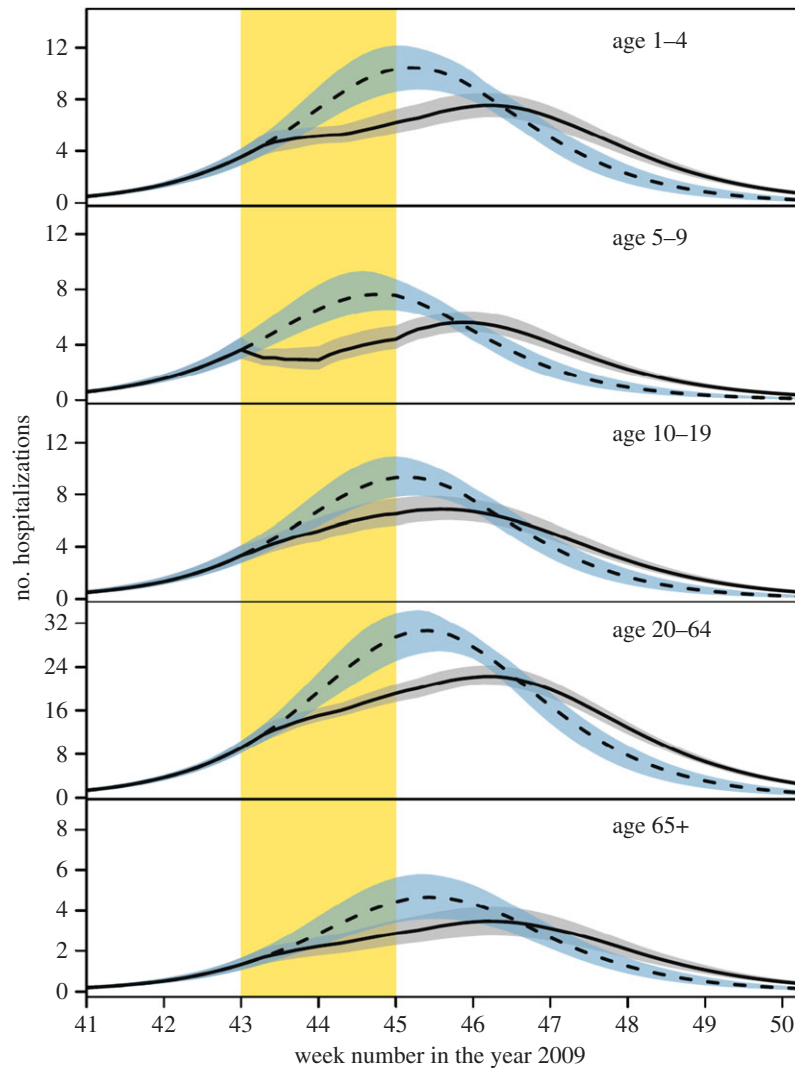
**Figure 2.** Overview of the hospitalization data (red lines) and model fits (black lines). The red line shows daily incidence of symptoms onset for influenza A requiring hospitalization. The area shaded in yellow indicates the timing of the school holiday. The solid and dashed black lines give fits of the models with and without school holiday, respectively. The shaded black area represents the 95% credible interval of the model with school holiday, and grey dashed lines indicate the Poisson 95% confidence interval of the number of hospitalizations in the model with school holiday.

In the above analyses, the susceptible component of the mixture distribution is located almost entirely below the detection limit ( $20 \text{ U ml}^{-1}$ ; figure 1), and the immune component has a high density at the detection limit. This implies that samples with an antibody concentration just above the detection limit have a high probability (99% or more) to be classified as immune. We explore how the results are affected if the component of pre-existing immunity is forced to have most (at least 95%) of its weight above the detection limit. The result is an increase in the mean and a decrease in the variance of the immune component (electronic supplementary material, table S4). Estimates of pre-existing immunity in adults remain high, but decrease from 72 to 52% (20–64 years) and from 91 to 84% (65+ years). Likewise, estimates of the infection attack rates remain high in children and very low in adults and the elderly. Estimates of the school holiday effect are negligibly affected.

In a second sensitivity analysis, we allow the probability of hospitalization to differ more between age groups (see Material and methods for details, and the electronic supplementary

material, table S5, for full results). The main effect is an increase in the estimated level of pre-existing immunity in the elderly (65+ years), accompanied by a substantial increase in the probability of hospitalization from 0.0084 to 0.017 in this age group. In the younger age groups, the probabilities of hospitalization are marginally affected, indicating that the choice of prior distributions has little impact on the posterior distribution, and that the probability of hospitalization can be identified by the data.

Finally, we analyse how our results compare to a traditional approach in which fixed thresholds determine whether an individual is classified as susceptible, immune or recently infected. In the electronic supplementary material, tables S6 and S7, we used threshold values of 20 and  $40 \text{ U ml}^{-1}$  for distinguishing susceptible from immune or infected persons. The most conspicuous difference compared with our mixture model analysis is that the estimates of pre-existing immunity are lower and attack rate estimates are higher. Further, estimates of the school holiday effect are consistently higher in the analyses with thresholds than in the mixture model.



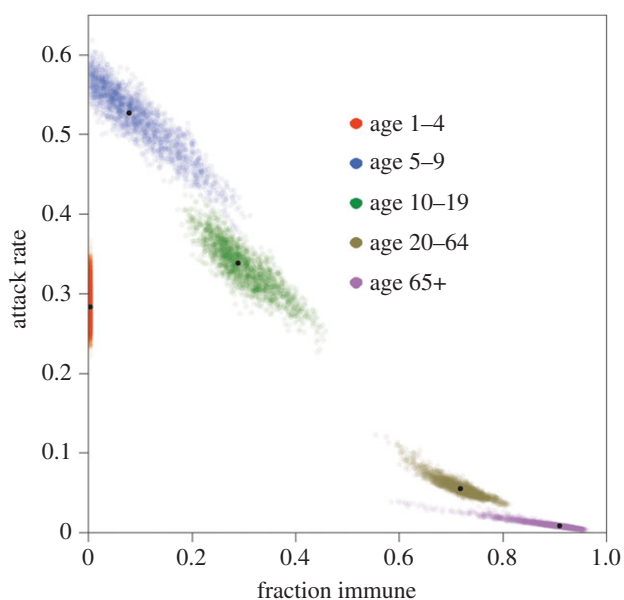
**Figure 3.** Comparison of model fit (solid black line) with predicted epidemic dynamics in the absence of the school holiday (dashed black line). The area shaded in yellow indicates the timing of the school holiday, and the grey and blue shaded areas represent 95% credible intervals of the model estimate (black) and predicted dynamics without school holiday (blue). The predicted epidemic dynamics was obtained by simulation of the dynamics without school holiday, but with parameter estimates (samples from the posterior distribution) from the model with school holiday (table 2).

#### 4. Discussion

Building on earlier work [3], our analyses demonstrate how disease incidence and infection seroprevalence data can be combined in a consistent manner to estimate infection attack rates, levels of pre-existing immunity, the impact of school holidays and probabilities of hospitalization. Our analyses improve on previous approaches by using a mixture model for probabilistic classification of serological samples. We believe that this represents a significant advance over traditional analyses that use predefined fixed thresholds. In fact, there is usually considerable biological variation in antibody response data, leading to classification uncertainties and making analyses using fixed thresholds prone to uncontrollable misclassification (electronic supplementary material, figures S6 and S7). In our mixture model approach, classification uncertainties are propagated to the epidemic model in a natural manner. Consequently, our probabilistic classification of serological samples is determined not only by serological information but also by the disease incidence data. This is the reason why the precision of estimates of the infection attack rates and levels of pre-existing immunity are comparable in the mixture model and the models with fixed thresholds; the lack of classification certainty in the mixture model is compensated by better use of the hospitalization data.

Our results suggest that high levels of immunity in persons older than 20 years, the lack of immunity in persons younger than 10 years and the week-long autumn holiday had a substantial impact on the epidemic in The Netherlands. A weighted average over the estimated contact reduction in 5–9- and 10–19-year-old children (54 and 10%) yielded an estimated contact reduction of 25% in 5- to 19-year-old children, which is similar to the reduction reported by Cauchemez *et al.* [11] for children in France. Our estimate is smaller than the reported reduction of more than 50% in children during the Canadian summer holiday [12]. Possible explanations for the relatively small estimated reduction in 10- to 19-year-old children are that they may be less dependent on school for social contacts and that a small proportion of 10- to 19-year-old children no longer attends school. The difference in the estimates in 5–9- versus 10–19-year-old children is of sufficient magnitude to suggest that the largest reduction in transmission would be achieved by closure of primary rather than secondary schools.

Parameter estimates in our analyses are informed by (i) serological data, (ii) hospitalization data, and (iii) the transmission model armed with age-specific contact patterns. Estimates of the infection attack rates are largely (but not exclusively) informed by the serological data. Particularly



**Figure 4.** Age-specific estimates of pre-existing immunity and infection attack rates. Coloured dots indicate samples from the posterior distribution, and black dots represent medians of the posterior distribution.

for the 5–9- and 10–19-year-old age groups, the post-pandemic serological data are bimodally distributed, and the distribution of infected persons is clearly identifiable (figure 1). On the other hand, attack rates have to be in line with the pandemic model and are restricted in particular by the contact patterns. The result is that confidence intervals of the attack rate estimates are smaller in our analyses than when the attack rates are estimated using the serological data alone (see below). This is especially true for the older age groups in which the serological data *per se* provide little information. Here, the information in the serological data for the younger age groups is carried over to the older age groups via the contact patterns specified by the epidemic model.

In a similar vein, estimates of the levels of pre-pandemic immunity are largely informed by the serological data. In fact, mixture model analyses without hospitalization data

are characterized by high estimated attack rates below the age of 20 years (5–9 years: 65%; 10–19 years: 27%) and low attack rates above 20 years (20–44 years: 5%; 45–64 years: approx. 0%), and levels of immunity that are low below 20 years (5–9 years: approx. 0%; 10–19 years: 37%) and high above 20 years (20–44 years: 70%; 45–64 years: 78%) [8]. The estimated attack rates and levels of immunity in our analyses are qualitatively—and to a reasonable extent also quantitatively—in agreement with attack rates and immunity rates based on serological data alone. The main quantitative difference is that estimates are more extreme when using serological data only, i.e. yielding higher attack rates in children and higher immunity estimates in adults and elderly. In our analyses, estimates are informed mainly by serological data, but also constrained by the hospitalization data and by contact patterns.

Our estimates and earlier estimates of the attack rate of the 2009 pandemic in The Netherlands are highly skewed, and range from up to 60% in young children (5–9 years) down to 0–1% in elderly [7]. The 2009 pandemic virus is structurally very similar to the 1918 H1N1 virus [23], and it has been shown that people exposed to pre-1957 H1N1 viruses have a substantial degree of pre-existing immunity to A/2009 H1N1-like viruses [24,25]. Our finding of high levels of immunity combined with relatively high probability of hospitalization after infection in older adults (approaching 1% in our main analyses) indicates that pre-existing immunity played a crucial role keeping the overall public health impact of the 2009 pandemic low [7].

**Acknowledgements.** Jan van de Kasstele is gratefully acknowledged for providing the contact matrix (electronic supplementary material, table S1). We thank Scott McDonald for valuable comments on a draft of the manuscript, and three anonymous reviewers for constructive criticism.

**Funding statement.** This project has been funded by the National Institute for Public Health and the Environment (RIVM) project S/210096. D.D.A. was supported by the Medical Research Council (U105260566) and Public Health England. P.J.B. was funded by the National Institute for Health Research HTA programme (project no. 11/46/03).

## References

- van den Wijngaard CC *et al.* 2012 Comparing pandemic to seasonal influenza mortality: moderate impact overall but high mortality in young children. *PLoS ONE* **7**, e311197. (doi:10.1371/journal.pone.0031197)
- Hardelid P, Pebody R, Andrews N. 2013 Mortality caused by influenza and respiratory syncytial virus by age group in England and Wales 1999–2010. *Influenza Other Respir. Viruses* **7**, 35–45. (doi:10.1111/j.1750-2659.2012.00345.x)
- Birrell PJ *et al.* 2011 Bayesian modeling to unmask and predict influenza A/H1N1pdm dynamics in London. *Proc. Natl Acad. Sci. USA* **108**, 18 238–18 243. (doi:10.1073/pnas.1103002108)
- Wu JT *et al.* 2010 School closure and mitigation of pandemic (H1N1) 2009, Hong Kong. *Emerg. Infect. Dis.* **16**, 538–541. (doi:10.3201/eid1603.091216)
- Lugnér AK, van Boven M, de Vries R, Postma MJ, Wallinga J. 2012 Cost effectiveness of vaccination against pandemic influenza in European countries: mathematical modelling analysis. *BMJ* **345**, e4445. (doi:10.1136/bmj.e4445)
- Mossong J *et al.* 2008 Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLOS Med.* **5**, e74. (doi:10.1371/journal.pmed.0050074)
- Steens A *et al.* 2011 Age-dependent patterns of infection and severity explaining the low impact of 2009 influenza A (H1N1): evidence from serial serologic surveys in the Netherlands. *Am. J. Epidemiol.* **174**, 1307–1315. (doi:10.1093/aje/kwr245)
- te Beest DE, De Bruin E, Imholz S, Wallinga J, Teunis P, Koopmans M, van Boven M. 2014 Discrimination of influenza infection (A/2009 H1N1) from prior exposure by antibody protein microarray analysis. *PLoS ONE* **9**, e113021. (doi:10.1371/journal.pone.0113021)
- Teunis P, Fonville M, Döpfer D, Eijck I, Molina V, Guarnera E, van Der Giessen J. 2009 Usefulness of sero-surveillance for *Trichinella* infections in animal populations. *Vet. Parasitol.* **159**, 345–349. (doi:10.1016/j.vetpar.2008.10.071)
- Wallinga J, Teunis P, Kretzschmar M. 2006 Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents. *Am. J. Epidemiol.* **164**, 936–944. (doi:10.1093/aje/kwj317)
- Cauchemez S, Valleron A-J, Boelle P-Y, Flahault A, Ferguson NM. 2008 Estimating the impact of school closure on influenza transmission from Sentinel data. *Nature* **452**, 750–754. (doi:10.1038/nature06732)
- Earn DJD, He D, Loeb MB, Fonseca K, Lee BE, Dushoff J. 2012 Effects of school closure on



- incidence of pandemic influenza in Alberta, Canada. *Ann. Intern. Med.* **156**, 173–181. (doi:10.1059/0003-4819-156-3-201202070-00005)
13. Koopmans M *et al.* 2012 Profiling of humoral immune responses to influenza viruses by using protein microarray. *Clin. Microbiol. Infect.* **18**, 797–807. (doi:10.1111/j.1469-0691.2011.03701.x)
  14. Baas DC *et al.* 2013 Detection of influenza A virus homo- and heterosubtype-specific memory B-cells using a novel protein microarray-based analysis tool. *J. Med. Virol.* **85**, 899–909. (doi:10.1002/jmv.23535)
  15. Boni MF *et al.* 2013 Population-level antibody estimates to novel influenza A/H7N9. *J. Infect. Dis.* **4**, 554–558. (doi:10.1093/infdis/jit224)
  16. Huijskens EGW *et al.* 2013 Profiling of humoral response to influenza A(H1N1)pdm09 infection and vaccination measured by a protein microarray in persons with and without history of seasonal vaccination. *PLoS ONE* **8**, e54890. (doi:10.1371/journal.pone.0054890)
  17. Donker T, van Boven M, van Ballegooijen WM, van't Klooster T, Wielders C, Wallinga J. 2011 Nowcasting pandemic influenza A/H1N1 2009 hospitalizations in The Netherlands. *Eur. J. Epidemiol.* **26**, 195–201. (doi:10.1007/s10654-011-9566-5)
  18. Mylius SD, Hagens TJ, Lugnér AK, Wallinga J. 2008 Optimal allocation of pandemic influenza vaccine depends on age, risk and timing. *Vaccine* **26**, 3742–3749. (doi:10.1016/j.vaccine.2008.04.043)
  19. Diekmann O, Heesterbeek H, Britton T. 2012 *Mathematical tools for understanding infectious disease dynamics*. Princeton, NJ: Princeton University Press.
  20. Carrat F, Vergu E, Ferguson NM, Lemaître M, Cauchemez S, Leach S, Valleron A-J. 2008 Time lines of infection and disease in human influenza: a review of volunteer challenge studies. *Am. J. Epidemiol.* **167**, 775–785. (doi:10.1093/aje/kwm375)
  21. Svensson A. 2007 A note on generation times in epidemic models. *Math. Biosci.* **208**, 300–311. (doi:10.1016/j.mbs.2006.10.010)
  22. Kass RE, Raftery AE. 1995 Bayes factors. *J. Am. Stat. Assoc.* **90**, 773–795. (doi:10.1080/01621459.1995.10476572)
  23. Xu R, Ekiert DC, Krause JC, Hai R, Crowe JE, Wilson IA. 2010 Structural basis of preexisting immunity to the 2009 H1N1 pandemic influenza virus. *Science* **328**, 357–360. (doi:10.1126/science.1186430)
  24. Hancock K *et al.* 2009 Cross-reactive antibody responses to the 2009 pandemic H1N1 influenza virus. *NE J. Med.* **361**, 1945–1952. (doi:10.1056/NEJMoa0906453)
  25. Ikonen N, Strengell M, Kinnunen L, Österlund P, Pirhonen J, Broman M, Davidkin I, Ziegler T, Julkunen I. 2010 High frequency of cross-reacting antibodies against 2009 pandemic influenza A (H1N1) virus among the elderly in Finland. *Euro Surveill.* **15**, pii:19478.