Vol 452|10 April 2008|doi:10.1038/nature06732

LETTERS

Estimating the impact of school closure on influenza transmission from Sentinel data

Simon Cauchemez[1], Alain-Jacques Valleron[2,3,4], Pierre-Yves Boëlle[2,3,4], Antoine Flahault[2,3,5] & Neil M. Ferguson[1]

The threat posed by the highly pathogenic H5N1 influenza virus requires public health authorities to prepare for a human pandemic. Although pre-pandemic vaccines and antiviral drugs might significantly reduce illness rates[1,2], their stockpiling is too expensive to be practical for many countries. Consequently, alternative control strategies, based on non-pharmaceutical interventions, are a potentially attractive policy option. School closure is the measure most often considered. The high social and economic costs of closing schools for months make it an expensive and therefore controversial policy, and the current absence of quantitative data on the role of schools during influenza epidemics means there is little consensus on the probable effectiveness of school closure in reducing the impact of a pandemic. Here, from the joint analysis of surveillance data and holiday timing in France, we quantify the role of schools in influenza epidemics and predict the effect of school closure during a pandemic. We show that holidays lead to a 20–29% reduction in the rate at which influenza is transmitted to children, but that they have no detectable effect on the contact patterns of adults. Holidays prevent 16–18% of seasonal influenza cases (18–21% in children). By extrapolation, we find that prolonged school closure during a pandemic might reduce the cumulative number of cases by 13–17% (18–23% in children) and peak attack rates by up to 39–45% (47–52% in children). The impact of school closure would be reduced if it proved difficult to maintain low contact rates among children for a prolonged period.

A thorough evaluation of the effectiveness of school closure as a pandemic mitigation measure is difficult, owing to the limited epidemiological data[3] and the current deficit in statistical methods to analyse those data. So far, it has been possible to establish that school closure is negatively correlated with influenza incidence[4,5]. Mathematical models have been used to evaluate the impact of school closure in a pandemic[1,2,6]. However, in the absence of quantitative estimates derived from epidemiological data, those models made strong assumptions about school transmission. The relatively wide range of effects they predicted[1,2,6] shows that modelling assumptions cannot replace the statistical investigation of epidemiological data.

Here we present a novel statistical approach to evaluating the impact of school closure on influenza epidemics from the joint analysis of disease surveillance data and information on the timing of school holidays in France. The hypothesis we examine is that influenza transmission changes during holidays as a result of the altered mixing patterns of children. The Sentinel network[7,8] (see http://www.sentiweb.org and Supplementary Information) is an internet-based network of French general practitioners (GPs). Since 1984, approximately 1,200 GPs have collected and sent data regularly on a dozen diseases, including influenza-like illness (case definition: sudden temperature of $>39\,^{\circ}C$, myalgia and cough/running nose). Regional daily incidences of influenza-like illness are estimated as area-weighted averages from individual GP declarations, using population data and data on the percentage of GPs participating in the surveillance network. Data on the timing of French holidays in different regions was obtained from the French Ministry of Education. In France, holidays are staggered across three geographic zones (two zones in 1986 and 1990) and the timing varies from region to region and from year to year. This provides conditions resembling those of a natural experiment.

The surveillance data consist of daily incidences for children (<18 years old) and adults (≥ 18 years old) for the two or three holiday zones in mainland France and over 21 years (1985–2006). We assume that half of all influenza patients consult his or her GP, giving a reasonable average attack rate of 11.4% (range 4.6–20.6%). We select epidemic periods (weekly incidence >160 per 100,000 inhabitants) and discard one epidemic that lasted 13 days only. This leaves 60 epidemic periods, with average duration 61 days (range 22–111 days) (Fig. 1a).

We model the spread of influenza in a population structured into households and schools (Fig. 1d; see Methods and Supplementary Information). Community transmission also occurs randomly between all members of the population. The simulated population matches the structure of the French population (Fig. 1b, c). We assume that at the start of each influenza season an average of 27% of the population is immune[9], and that immunity is distributed within the population from its expected stationary distribution (see Supplementary Information). During holidays, no transmission occurs in schools, but in other places (household, community), transmission rates may be modified. We use estimates from another study[10] to characterize household transmission and the infectiousness profile (generation time 2.4 days).

The high dimensionality of the data means that model parameters cannot be estimated using standard statistical methods, such as least-squares fitting or data augmentation[10]. We therefore designed a new statistical approach, based on the simulation of epidemics that are constrained to be consistent with the observed incidence curves (Fig. 1e; see Methods and Supplementary Information). Using simulated data, we find that, even in a context with observation errors and where transmissibility varies substantially between epidemics, the inference method gives satisfactory estimates of all parameters (see Supplementary Information). The approach also provides the relative prediction error (RPE; see Fig. 2a).

We first estimate transmission parameters under the assumption that influenza transmission is not modified during holidays (see Supplementary Information). For this model, adult and child RPEs are close to 0% during the school term (Fig. 2a). Adult RPE is also close to 0% during holidays; but child RPE drops to -24% (range -20% to -29%) during holidays. This implies that, on average,

[1]MRC Centre for Outbreak Analysis and Modelling, Department of Infectious Diseases Epidemiology, Imperial College London, Norfolk Place, London W2 1PG, UK. [2]Université Pierre et Marie Curie—Paris 6, UMR S 707, 27 rue Chaligny, Paris 75012, France. [3]INSERM, UMR S 707, 27 rue Chaligny, Paris 75012, France. [4]AP-HP, Hôpital St Antoine, 27 rue Chaligny, Paris 75012, France. [5]French School of Public Health (EHESP), 1 place du Parvis Notre-Dame, Paris F-75004, France.

holidays lead to a 24% reduction in the rate at which influenza is transmitted to children, but that they have no detectable effect on adults' contact patterns.

We then estimate transmission parameters allowing for modifications of children's contact patterns during holidays. Three quantities are needed: transmission rates within schools, in the community, and the increase in non-school transmission rates during holidays (compensatory behaviours). It is not possible to estimate those three quantities independently from the available data. We therefore fix one of them (compensatory behaviours) and estimate the other two. We then undertake a rigorous sensitivity analysis on compensatory behaviours, considering 30 parameter combinations parameterized by the increase in child-to-child community transmission ($\delta_{com} = 0\%$, 50%, 100%, 150%, 200% and up to ∞) and in child-to-child household transmission ($\delta_{hous} = 0\%$, 50%, 100%, 150% and 200%).

Irrespective of $\{\delta_{com}, \delta_{hous}\}$, we find that accounting for holidays improves the model fit: (1) log-likelihoods are larger (see Supplementary Information); (2) child RPE becomes close to 0% during holidays (Fig. 2a). We find that the proportion of transmission occurring in schools increases with $\{\delta_{com}, \delta_{hous}\}$, ranging from 7 to 20% overall (Fig. 2c) and from 20 to 54% in children (Fig. 2d). The proportion of secondary cases of children infected in schools is 16–44%. Other summary statistics are robust to a change in $\{\delta_{com}, \delta_{hous}\}$ (see Supplementary Information). Although children represent 28% of the population, they are responsible for 46–47% of all infections (Fig. 2e). Household transmission accounts for 36–39% of infections of children (and 40% of adult infections). Household members make up 48–50% of secondary cases in children. The basic reproduction number, R_0 (average number of cases generated by one typical case in a completely susceptible population) is estimated to be 1.7 (range 1.5–1.8) during school term, and 1.4 (range 1.3–1.6) in holidays. The average R_0 for child cases is 2.2 (range 2.0-2.4) during term and 1.7 (1.4–1.9) during holidays, while for adult cases it is 1.3 (range 1.2-1.4) for both terms and holidays (see Supplementary Information).

No major difference in transmission is detected between Christmas and other breaks (winter and spring breaks). For adults, RPE is 0% (range −4% to 5%) over Christmas and 1% (range −4% to 5%) during other breaks. For children, RPE is −4% (range −15% to 9%) over Christmas and −1% (range −12% to 12%) during other breaks.

Classifying each year by the dominant influenza virus type or subtype and fitting season-specific variations in transmissibility, we find that subtype B is less transmissible but more child-associated

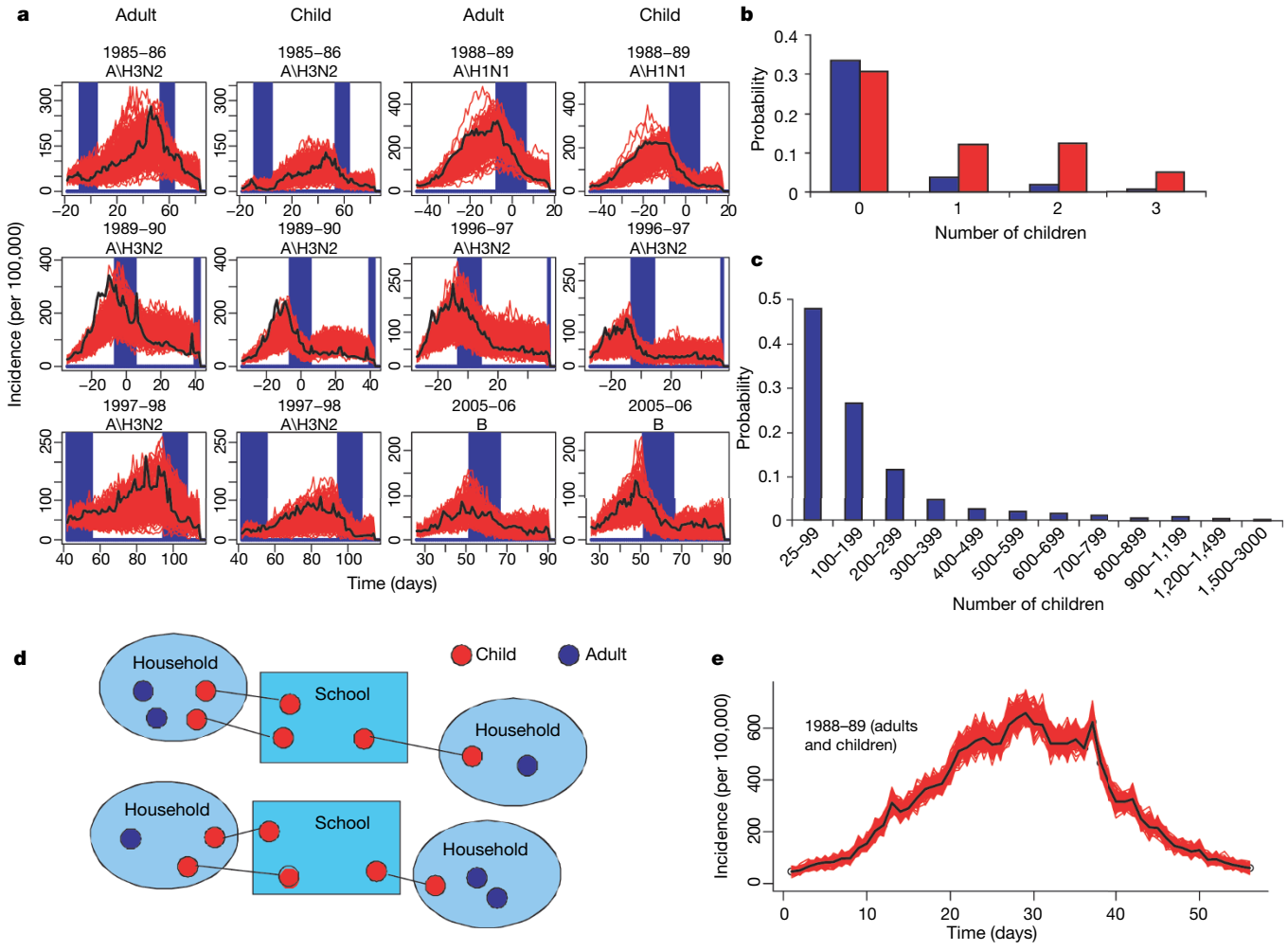

Figure 1 | Data, transmission model and inference method. **a**, Daily incidence (black line) for children (<18 years old) and adults and holiday timing (blue bars) for six of the 60 epidemic periods selected among surveillance data over 21 years (1985–2006) and three holiday zones in France. Day 0 corresponds to 1 January. Red lines show 200 simulations from the model, with parameters drawn from their posterior distribution. **b**. Size distribution of French households (1999 census). Blue, one adult; red, two adults. **c**, Size distribution of French schools (1999 census). **d**, Schematic diagram of transmission model in a population structured into households and schools (see Methods and Supplementary Information). **e**, Constrained simulations. For inference, epidemics are simulated which are constrained to be consistent with observed incidence curves. Black line, the observed incidence curve for one holiday zone in 1988–89; red lines, 200 constrained simulations. See Methods and Supplementary Information.

than subtype A\H3N2 (Wilcoxon test: probability $P = 2.6\%$ for the strength of transmission, and $P = 1.8\%$ for the relative contribution of children to transmission) and that subtype A\H1N1 has intermediate characteristics between subtype B and subtype A\H3N2 (no significant difference with subtype B, nor with A\H3N2) (Fig. 2b).

We then simulate epidemics from the model, with parameters drawn from their posterior distribution for the different model variants (Fig. 1a and Supplementary Information). Simulations start at the same time and with the same number of initial cases as observed epidemics, use season-specific transmissibility estimates, but are not otherwise constrained. For adults, 81% (83% for children) of observed daily attack rates fall between the 2.5th and 97.5th percentile of the distribution of simulated daily attack rates.

These simulations are used to assess the impact of school closure on seasonal and pandemic attack rates. Figure 3a–c shows that, irrespective of the assumptions made about $\{\delta_{com}, \delta_{hous}\}$, we obtain the same estimates of the impact of school closure on cumulative and peak attack rates. For typical holiday timings, the different model variants predict an average seasonal attack rate of 10.6–11.1%. They also predict that, if schools were always open, the attack rate would be 12.8–13.4%; that is, holidays prevent 16–18% of seasonal influenza cases (for adults 14–17%; for children 18–21%).

We then consider the pandemic context, where 100% of the population is susceptible and assume that 50% of infections are symptomatic. For typical holiday timings, 31% of the population would report being ill (37–38% of children); and the daily incidence at the peak would be 1.6–1.7% (2.1–2.2% in children). If schools were closed permanently at an early stage (for example, once daily incidence exceeds 20/100,000), with subsequent behaviour typical of holidays, the cumulative number of cases would be curbed by 13–17% overall (for children only 18–23%) and the number of cases at the peak by 39–45% (for children only 47–52%).

Contact patterns might, however, be less affected by prolonged school closure than by normal school breaks, when people go on vacation, celebrate Christmas, and so on. The reductions we predict might therefore be an upper bound of what might happen during school closure in a pandemic. If compensatory increases in contact rates $\{\delta_{com}, \delta_{hous}\}$ were 1.5-fold larger during school closure in a pandemic than for typical holidays, there would be at most a very limited reduction in cumulative/peak attack rates and in R_0 (Fig. 3d–i and Supplementary Information).

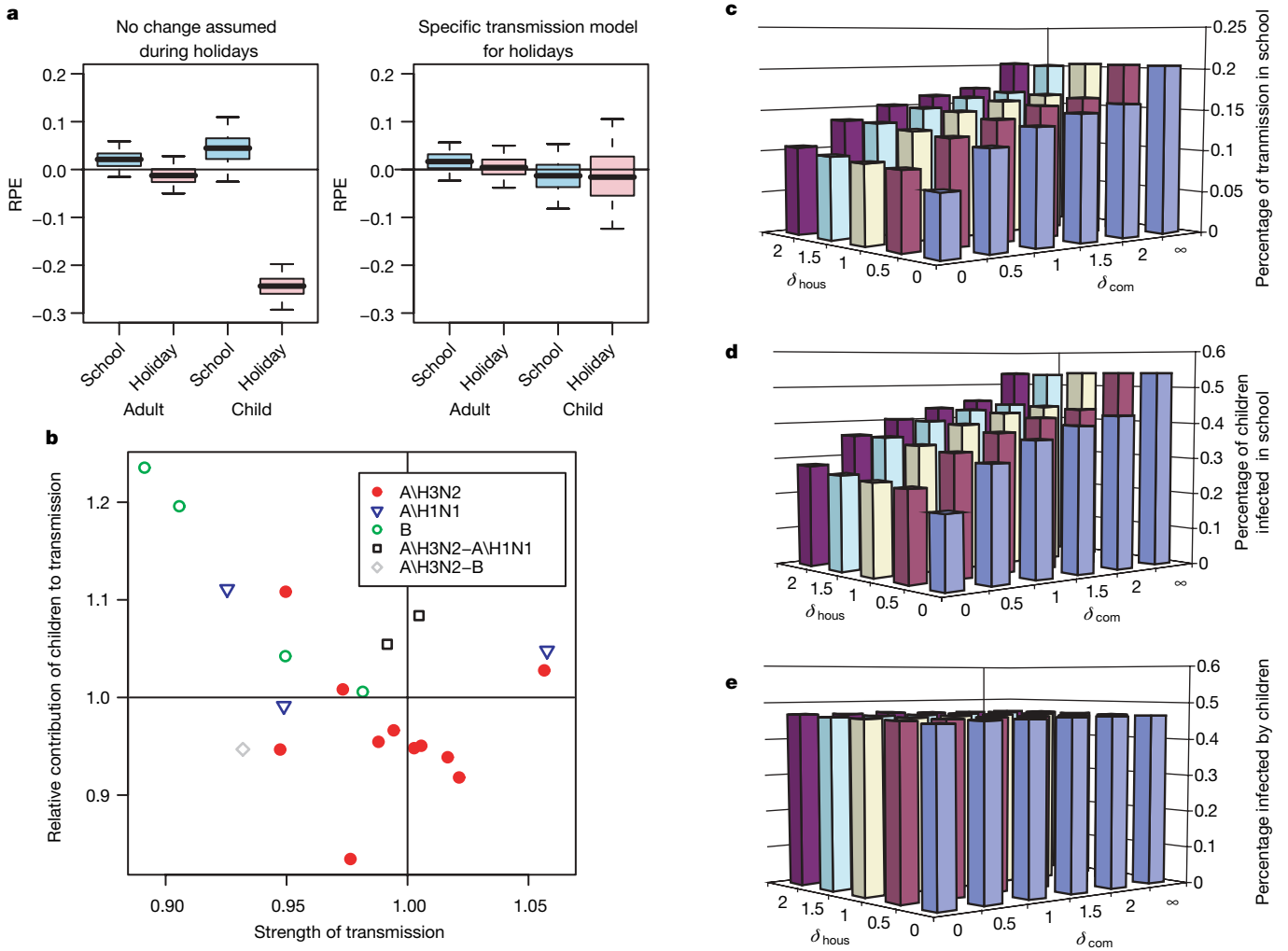

Figure 2 | Inferred influenza transmission characteristics. **a**, Posterior distribution (2.5%, 25%, 50%, 75% and 97.5% percentiles) of the RPE for adults and children, during school terms and during holidays, when no change in transmission is assumed during holidays (left) and when a specific transmission model is designed for holidays (right). RPE is the average relative error $(O_t - E_t)/E_t$ between the number O_t of cases observed at time t and the number E_t of cases predicted by the model given the observed epidemic up to time $t - 1$. An RPE close to 0% is indicative of good fit. **b**, Strength of transmission, and relative contribution of children to transmission for each epidemic season according to the circulating subtype[19] (see Methods and Supplementary Information for details of calculation). **c**, Proportion of school transmission according to the strength of compensatory behaviours in the community (δ_{com}) and in the household (δ_{hous}). **d**, Proportion of children infected in schools. **e**, Proportion of cases infected by children.

For one model variant ($\delta_{com} = \delta_{hous} = 100\%$), we then perform a sensitivity analysis and find that results are relatively robust to varying other model assumptions (Fig. 3d–i and Supplementary Information). If prolonged school closure has the same impact as holidays, the relative reduction in the cumulative number of cases in a pandemic is always below 20%, except under the unlikely (see Supplementary Information) assumptions that immunity is completely clustered in households (42% reduction) or that the generation time is as long as 4.1 days (25% reduction). Summary statistics on the place of transmission are also relatively robust to changes in modelling assumptions (see Supplementary Information).

The derivation of influenza incidence from GP reports is uncertain because some cases do not visit a GP, only a small proportion of GPs report, diagnosis is based on influenza-like illness with no viral ascertainment and there are asymptomatic infections. Three observations do however suggest that the influenza-like illness data provide a sensible description of influenza circulation. First, they are consistent with data collected independently on virus circulation (see Supplementary Information). Second, weekly mortality due to pneumonia and influenza is almost perfectly predicted by the surveillance data and the circulating strains[11]. Lastly, solely on the basis of the influenza-like illness data, we found that transmission characteristics of influenza depended on the circulating subtype (Fig. 2b), in a way that is consistent with past epidemiological studies[12–14].

Although the apparent impact of public health measures was substantial (that is, up to 50% reduction in transmission) in some US cities in 1918, it is not possible to disentangle the relative impact of different measures[15–17]. School closure was commonly adopted, and in some of the cities in which schools were closed, the total impact of all public health measures was estimated to be as low as 10% (ref. 15). Here, we used a natural experiment to estimate the specific effect of school closure on seasonal influenza transmission. Our extrapolations to the pandemic context rest on the relatively strong assumption that people will behave during a pandemic as they do during seasonal outbreaks.

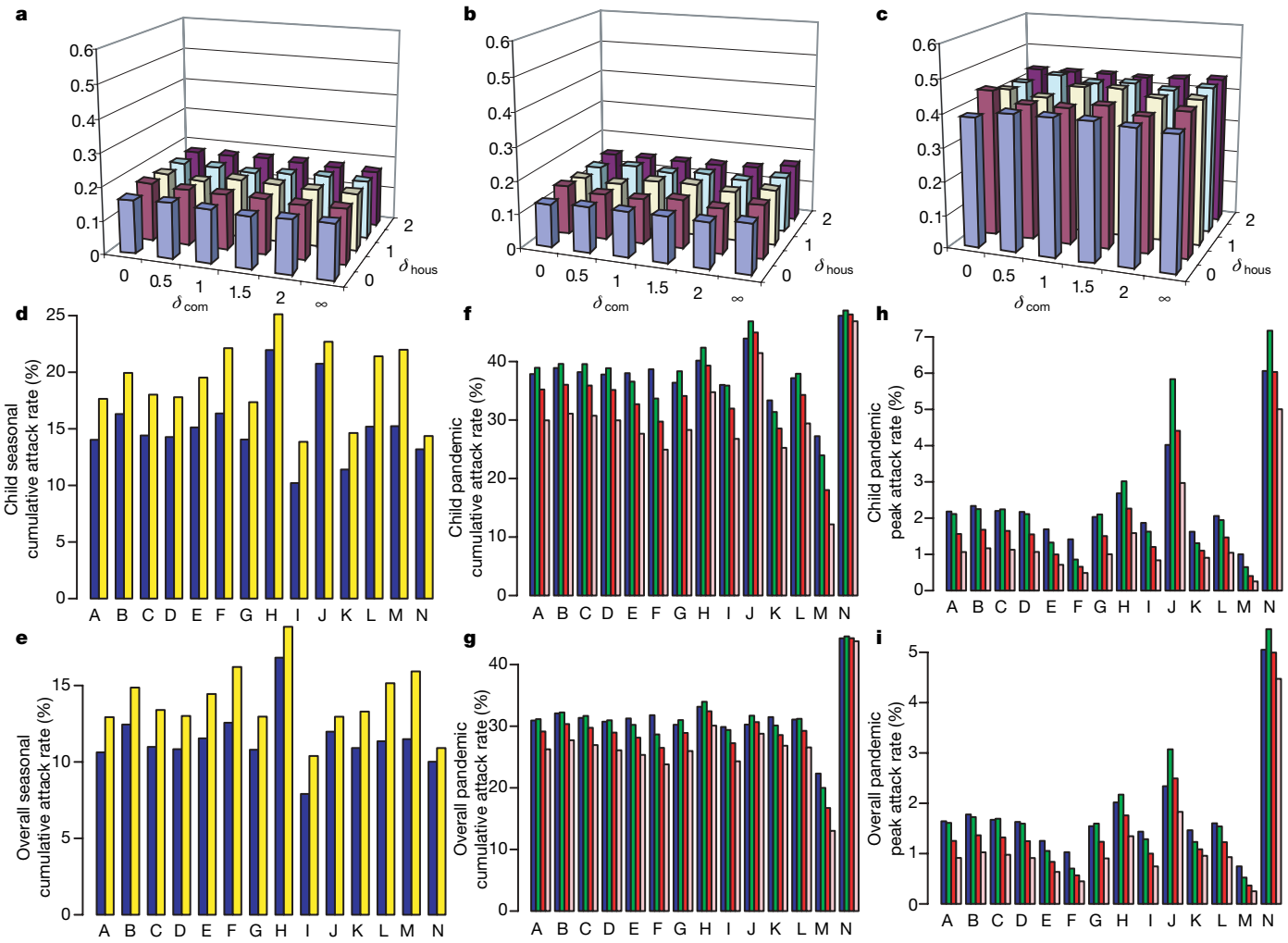

Figure 3 | Impact of school closure on seasonal and pandemic influenza. **a**, Relative reduction in seasonal influenza cumulative attack rates due to holidays, according to the assumed compensatory increase in contact rates in the community (δ_{com}) and in the household (δ_{hous}) during French holidays, and under baseline assumptions (see main text). **b**, Relative reduction in pandemic cumulative attack rates due to permanent school closure, assuming closure has the same effect on transmission as holidays. **c**, As for **b**, but for peak daily attack rate. **d–i**, Sensitivity analyses for parameters estimated assuming $\delta_{com} = \delta_{hous} = 100\%$ during French holidays. A: baseline (see main text); B: 19 smallest outbreaks discarded; C and D: epidemic period defined as weekly incidence over 120/100,000 or 200/100,000 respectively; E and F: 3.25-day and 4.11-day generation time respectively; G: household transmission rates 25% smaller than estimated in ref. 10; H–K: adult reporting rates of 30%, 70%, 50% and 50% respectively, with child reporting rates of 30%, 70%, 30% and 70% respectively; L: immunity seeded independently of household; M: immunity clustered by household; N: 50% immune. **d**, Seasonal cumulative attack rates among children during a typical holiday pattern (blue) and when schools are never closed (yellow). **e**, As for **d**, but for the whole population. **f**, Pandemic cumulative attack rates among children during a typical holiday pattern (blue); when schools are closed throughout with $\delta_{com} = \delta_{hous} = 100\%$ (pink); when schools are closed throughout but with compensatory contact rate increases 1.5-fold larger than normal holidays, that is $\delta_{com} = \delta_{hous} = 150\%$ (green); and as for green but with $\delta_{com} = \delta_{hous} = 125\%$ (red). **g**, As for **f**, but for the whole population. **h**, As for **f**, but for peak daily attack rates. **i**, As for **h**, but for the whole population.

Because demography and school holiday patterns are similar across much of Europe, we are confident that our results can be extrapolated to other European countries. Extrapolation to developing countries is more difficult because of the absence of independent data.

Compared with other studies of influenza transmission[18,19], our analysis shows that age is an important determinant of seasonal variations in influenza transmission both within (holidays fundamentally affect children's contact patterns) and between epidemics (there are large variations between seasons in the relative contribution of children to transmission).

Methods used to estimate parameters of complex transmission models (for example, data augmentation techniques[10]) have traditionally been very distinct from techniques used for prediction (that is, simulation[1]). The difficulty (and sometimes, as here, impossibility) of implementing those estimation methods for high-dimensional dynamical models largely explains why estimation often rests on naive least-squares fitting. In contrast, the new statistical method presented here, which relies on sequential Monte Carlo methods[20,21], makes it straightforward to upgrade a complex epidemic simulator to a computationally efficient likelihood-based inference tool.

Pandemic planning is a challenging task in today's highly connected world and when some key characteristics of the future pandemic virus cannot be predicted. Mathematical models provide a framework for assisting rational decision-making. However, for models to have predictive power, it is critical that they make full use of epidemiological data. Undertaking more epidemiological studies and designing statistical methods to extract maximum information from the data collected therefore remains a priority.

In public health terms, our conclusions do not rule out the use of school closure in a severe pandemic. We predict that this policy can significantly reduce the stress on healthcare systems at the peak of the pandemic. But our work should temper expectations of the scale of the reduction in overall illness and mortality achievable through this measure alone.

METHODS SUMMARY

Transmission model. The household transmission rate associated with an infective person of age a (where $a = \mathrm{A}$ for adult or C for child) is $\beta^a_{\mathrm{hous}} f(t)/n$, where n is the size of the household and $f(t)$ characterizes the relative infectiousness at time t since infection. An infectious child infects children in the same school at a rate $\beta_{\mathrm{school}} f(t)/N_{\mathrm{school}}$ where N_{school} is the size of the school. We make a distinction between adult-to-adult (A→A), child-to-child (C→C) and adult-to-child or child-to-adult (A→C, C→A) transmission in the community. During holidays, C→C community transmission increases by a factor of $1+\delta_{\mathrm{com}}$ ($1+\delta_{\mathrm{hous}}$ for household transmission). We explore a range of possible compensatory behaviours, parameterized by $\delta_{\mathrm{com}} = 0\%$, 50%, 100%, 150% 200% and ∞ ($\delta_{\mathrm{com}} = \infty$ is the extreme situation where children mix only in schools during school terms and mix only in the community during holidays) and $\delta_{\mathrm{hous}} = 0\%$, 50%, 100%, 150% and 200%. At any time t, susceptible individual i is exposed to a baseline infection risk of $\lambda_i^{\mathrm{baseline}}(t)$, which is the sum of the baseline risks of infection in their household, their school (for children) and the community. To model seasonal variations in influenza transmission, we introduce for each year y the strength of transmission σ_y and the relative contribution of children to transmission τ_y. During year y, at time t, the risk of infection $\lambda_{i,y}(t)$ for susceptible i is $\sigma_y \tau_y \lambda_i^{\mathrm{baseline}}(t)$ if i is a child and $(\sigma_y/\tau_y)\lambda_i^{\mathrm{baseline}}(t)$ if i is an adult. See the Supplementary Information for more information.

'Constrained' simulations. To approximate the likelihood of the parameters, we simulate epidemics constrained to be consistent with the observed incidence curves (Fig. 1e). At any time t, on average, the number of cases generated by the constrained simulator equals the observed number of cases. The likelihood is then approximated by sequential importance sampling[20,21], and the parameters space is explored by Markov-chain Monte-Carlo sampling[22,23]. Details on the simulation of constrained epidemics and the statistical methodology are given in the Supplementary Information and online-only Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 4 December 2007; accepted 21 January 2008.

1. Ferguson, N. M. *et al.* Strategies for mitigating an influenza pandemic. *Nature* **442,** 448–452 (2006).
2. Germann, T. C., Kadau, K., Longini, I. M. Jr & Macken, C. A. Mitigation strategies for pandemic influenza in the United States. *Proc. Natl Acad. Sci. USA* **103,** 5935–5940 (2006).
3. Bell, D. M. Non-pharmaceutical interventions for pandemic influenza, national and community measures. *Emerg. Infect. Dis.* **12,** 88–94 (2006).
4. Heymann, A., Chodick, G., Reichman, B., Kokia, E. & Laufer, J. Influence of school closure on the incidence of viral respiratory diseases among children and on health care utilization. *Pediatr. Infect. Dis. J.* **23,** 675–677 (2004).
5. Valleron, A. J., Flahault, A. & Boelle, P. Y. *Do school holidays have an impact on influenza epidemics, then on mortality?* Presentation at *International Conference on Options for the Control of Influenza V* (Okinawa, 7–11 October 2003); ⟨http://www.u707.jussieu.fr/valleron/dia/dia2003/okinawa.pdf⟩.
6. Glass, R., Glass, L., Beyeler, W. & Min, H. Targeted social distancing design for pandemic influenza. *Emerg. Infect. Dis.* **12,** 1671–1681 (2006).
7. Valleron, A. J. *et al.* A computer network for the surveillance of communicable diseases: the French experiment. *Am. J. Public Health* **76,** 1289–1292 (1986).
8. Flahault, A. *et al.* Virtual surveillance of communicable diseases: a 20-year experience in France. *Stat. Methods Med. Res.* **15,** 413–421 (2006).
9. Longini, I. M. Jr, Koopman, J. S., Haber, M. & Cotsonis, G. A. Statistical inference for infectious diseases. Risk-specific household and community transmission parameters. *Am. J. Epidemiol.* **128,** 845–859 (1988).
10. Cauchemez, S., Carrat, F., Viboud, C., Valleron, A. J. & Boelle, P. Y. A. Bayesian MCMC approach to study transmission of influenza: application to household longitudinal data. *Stat. Med.* **23,** 3469–3487 (2004).
11. Denoeud, L. *et al.* Predicting pneumonia and influenza mortality from morbidity data. *PLoS One* **2,** e464 (2007).
12. Glezen, W. P. *et al.* Age distribution of patients with medically-attended illnesses caused by sequential variants of influenza A/H1N1: comparison to age-specific infection rates, 1978–1989. *Am. J. Epidemiol.* **133,** 296–304 (1991).
13. Monto, A. S. & Sullivan, K. M. Acute respiratory illness in the community. Frequency of illness and the agents involved. *Epidemiol. Infect.* **110,** 145–160 (1993).
14. Olson, D. R. *et al.* Monitoring the impact of influenza by age: emergency department fever and respiratory complaint surveillance in New York City. *PLoS Med.* **4,** e247 (2007).
15. Bootsma, M. C. J. & Ferguson, N. M. The effect of public health measures on the 1918 influenza pandemic in U.S. cities. *Proc. Natl Acad. Sci. USA* **104,** 7588–7593 (2007).
16. Hatchett, R. J., Mecher, C. E. & Lipsitch, M. Public health interventions and epidemic intensity during the 1918 influenza pandemic. *Proc. Natl Acad. Sci. USA* **104,** 7582–7587 (2007).
17. Markel, H. *et al.* Nonpharmaceutical interventions implemented by US cities during the 1918–1919 influenza pandemic. *J. Am. Med. Assoc.* **298,** 644–654 (2007).
18. Finkenstadt, B. F., Morton, A. & Rand, D. A. Modelling antigenic drift in weekly flu incidence. *Stat. Med.* **24,** 3447–3461 (2005).
19. Xia, Y., Gog, J. R. & Grenfell, B. Semiparametric estimation of the duration of immunity from infectious disease time series: influenza as a case-study. *J. R. Stat. Soc. Ser. C* **54,** 659–672 (2005).
20. Doucet, A., de Freitas, N. & Gordon, N. *Sequential Monte Carlo Methods in Practice* (Springer, New York, 2001).
21. Liu, J. S. *Monte Carlo Strategies in Scientific Computing* (Springer, New York, 2001).
22. Gilks, W. R., Richardson, S. & Spiegelhalter, D. J. *Markov Chain Monte Carlo in Practice* (Chapman and Hall, London, 1996).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the MRC, European Union FP6 SARSTRANS and INFTRANS projects, RCUK, and the NIGMS MIDAS initiative for research funding. We thank F. Carrat for comments.

Author Contributions S.C. developed the transmission model and conceived and implemented the inference framework used, did the analysis and drafted and revised the text. All other authors edited or commented on the text. A.-J.V., P.-Y.B. and A.F. identified, collated and processed the surveillance and holiday data. P.-Y.B. also provided input on the statistical framework. N.M.F. conceived the study (building on earlier work by A.-J.V. and A.F. examining the correlation between holidays and seasonal influenza incidence), provided input on the statistical framework, model design and assumptions and gave other technical advice.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to S.C. (s.cauchemez@imperial.ac.uk).

METHODS

Simulation of 'unconstrained' epidemics. Let θ be the baseline transmission parameters and V the set of parameters characterizing annual variations in transmission. Consider an outbreak occurring in year y. Given the initial state of the system, it is straightforward to simulate epidemics from the model (we use a discrete time-step of $\Delta T = 0.25$ days). The epidemic starts at time-step $L = 0$.

Denote by Z_L the history of the epidemic (specifying those in the population who are infected and those who are immune, and the times of infection) up to time-step L. Given Z_{L-1}, the probability that susceptible person i is infected during time-step L is:

$$p^L_{i,y} = 1 - \exp\{-\lambda_{i,y}(L\Delta T)\Delta T\}$$

where $\lambda_{i,y}(t)$ is the hazard of infection. The associated density is:

$$P(Z_L|Z_{L-1},\theta,V) = \prod_{i:\text{ susceptible people at } L-1} \left(p^L_{i,y}\right)^{x^L_i} \left(1-p^L_{i,y}\right)^{1-x^L_i}$$

where x^L_i is equal to 1 if susceptible person i is infected at time-step L and zero otherwise. The unconstrained density for the complete history of the epidemic Z is:

$$g(Z|\theta,V) = P(Z_0|Y,\theta) \prod_{L=1}^{\infty} P(Z_L|Z_{L-1},\theta,V) \quad (1)$$

Seeding the initial state of the system $\{Z_0\}$ is described in the Supplementary Information.

Simulation of constrained epidemics. We modify the unconstrained simulator to a constrained simulator, which simulates epidemics consistent with the data. Assume that the constrained epidemic has been simulated up to time-step $L-1$. From the model, we compute the risk of infection $\lambda_{i,y}(L\Delta T)$ for each susceptible person i for time-step L; and derive the expected incidence for adults $E_A(L)$ and children $E_C(L)$:

$$E_A(L) = \frac{\sum_{i:a(i)=A} \lambda_{i,y}(L\Delta T)\Delta T}{N_{com}} \text{ and } E_C(L) = \frac{\sum_{i:a(i)=C} \lambda_{i,y}(L\Delta T)\Delta T}{N_{com}}$$

We denote by $Y_A(L)$ and $Y_C(L)$ the observed incidence at time step L among adults and children respectively. The ratio of observed incidence to expected incidence is $\rho_A(L) = Y_A(L)/E_A(L)$ for adults and $\rho_C(L) = Y_C(L)/E_C(L)$ for children. The constrained epidemic is obtained by simulating the infection process with corrected infection risks. For any susceptible person i, the corrected infection risk is:

$$\lambda^*_{i,y}(L\Delta T) = \rho_{a(i)}(L)\lambda_{i,y}(L\Delta T)$$

where $a(i)$ is the age (either adult or child status) of the susceptible person. It is straightforward to check that, for the constrained process, the expected incidence at time step L is $Y_A(L)$ for adults and $Y_C(L)$ for children.

The density of the constrained simulation for time-step L is:

$$P_{\text{constrained}}(Z_L|Z_{L-1},Y,\theta,V) = \prod_{i:\text{ susceptible people at } L-1} \left(p^{L*}_{i,y}\right)^{x^L_i} \left(1-p^{L*}_{i,y}\right)^{1-x^L_i}$$

where $p^{L*}_{i,y} = 1 - \exp\left\{-\lambda^*_{i,y}(L\Delta T)\Delta T\right\}$ is the constrained probability of infection. The constrained density for the complete history of the epidemic Z is:

$$h(Z|Y,\theta,V) = P(Z_0|Y,\theta) \prod_{L=1}^{\infty} P_{\text{constrained}}(Z_L|Z_{L-1},Y,\theta,V) \quad (2)$$

Approximation of the likelihood. If the complete history of the epidemic Z (who has been infected when in the structured population) was known, it would be easy to write down the probability $P(Z|\theta,V)$ (see Supplementary Information), and then likelihood-based inference, in a frequentist or bayesian setting, would be straightforward. However, the data Y consist only of daily incidences. The likelihood $P(Y|\theta,V)$ is then difficult to compute because it requires integration with respect to the (unobserved) complete history Z, which has a very high dimension:

$$P(Y|\theta,V) = \int_Z P(Y|Z)g(Z|\theta,V)\,dZ$$

The first term of the integrand is the observation model, which ensures that complete history Z is consistent with the observed curves Y: so $P(Y|Z)$ is equal to 1 if Z is consistent with Y and 0 otherwise. The second term is the sampling density for the complete history Z of the epidemic (equation (1)).

We have designed an approach based on sequential importance sampling[20,21] to approximate the likelihood. The idea is to work with simulated epidemics that are constrained to be consistent with the observation Y (see above), and have density h (equation (2)). We can rewrite the likelihood (via multiplication by 1) as:

$$P(Y|\theta,V) = \int_Z \frac{g(Z|\theta,V)}{h(Z|Y,\theta_{\text{simul}},V_{\text{simul}})} h(Z|Y,\theta_{\text{simul}},V_{\text{simul}})dZ$$

To obtain an importance sampling approximation[21] of this integral, we simulate $Z^1,\ldots,Z^N$ constrained epidemics (sampling from density h, with parameters $\{\theta_{\text{simul}}, V_{\text{simul}}\}$), and approximate the likelihood by:

$$P(Y|\theta,V) \approx \frac{1}{N}\sum_{n=1}^{N} \frac{g(Z|\theta,V)}{h(Z|Y,\theta_{\text{simul}},V_{\text{simul}})}$$

There is much less stochastic fluctuation in the sampling process than for unconstrained epidemics, so we do not have to simulate large numbers of epidemics per observed curve. We found that the integral was well evaluated using only one simulation per observed curve ($N = 1$; see Supplementary Information). So, in practice, we used $N = 1$. This point makes the whole estimation process extremely efficient.

Another typical feature of sequential importance sampling is that the epidemic trajectory supporting estimation of the log-likelihood does not have to be obtained with the parameters of interest, that is $\{\theta_{\text{simul}}, V_{\text{simul}}\}$ may be different from $\{\theta, V\}$. Therefore, using a trajectory simulated with well-chosen values of $\{\theta_{\text{simul}}, V_{\text{simul}}\}$, the approximation of the log-likelihood is possible in a larger region of the parameter space. This property is at the heart of our estimates of the annual variations in transmission V (see Supplementary Information).

Statistical framework. In a bayesian context, we explore the posterior distribution of the parameters by Markov-chain Monte Carlo sampling[22]. When we account for annual variations in influenza transmission V, we rely on the profile likelihood for θ:

$$\text{LP}(\theta,Y) = P\left(Y|\theta,\hat{V}(\theta)\right)$$

where $\hat{V}(\theta)$ maximizes $P(Y|\theta,V)$ with respect to V and satisfies identifiability constraints (see Supplementary Information).