

Multiannual forecasting of seasonal influenza dynamics reveals climatic and evolutionary drivers

Jacob Bock Axelsen^a, Rami Yaari^{a,b}, Bryan T. Grenfell^{c,d}, and Lewi Stone^{a,e,1}

^aBiomathematics Unit, Department of Zoology, Faculty of Life Science, and ^bPorter School of Environmental Studies, Tel Aviv University, Ramat Aviv 69978, Israel; ^cWoodrow Wilson School of Public and International Affairs, Princeton University, Princeton, NJ 08544; ^dFogarty International Center, National Institutes of Health, Bethesda, MD 20892-2220; and ^eSchool of Mathematical and Geospatial Sciences, RMIT University, Melbourne, VIC 3001, Australia

Edited by Bruce R. Levin, Emory University, Atlanta, GA, and approved May 20, 2014 (received for review November 22, 2013)

Human influenza occurs annually in most temperate climatic zones of the world, with epidemics peaking in the cold winter months. Considerable debate surrounds the relative role of epidemic dynamics, viral evolution, and climatic drivers in driving year-to-year variability of outbreaks. The ultimate test of understanding is prediction; however, existing influenza models rarely forecast beyond a single year at best. Here, we use a simple epidemiological model to reveal multiannual predictability based on high-quality influenza surveillance data for Israel; the model fit is corroborated by simple metapopulation comparisons within Israel. Successful forecasts are driven by temperature, humidity, antigenic drift, and immunity loss. Essentially, influenza dynamics are a balance between large perturbations following significant antigenic jumps, interspersed with nonlinear epidemic dynamics tuned by climatic forcing.

model forecasting | infectious disease | climate | Bayesian epidemic model | predictive model

Influenza outbreaks have been documented in the scientific literature in records that extend back to at least 1650 (1), making it an exceptional example of a persisting, recurrent disease. Being a respiratory infection, influenza spreads rapidly from person to person through a population in the form of virus particles airborne as respiratory droplets or aerosols. Depending on the circumstances, influenza typically infects between 10% and 50% of a given population and has become a source of considerable human morbidity and mortality (2). There is much controversy in identifying the seasonal drivers that generate annual influenza epidemics and the processes that give rise to their large variability (3–12). This is an outstanding problem of influenza research today. Using long-term modeling, a recent study (9) gave support to the possibility that absolute humidity is the predominant determinant of influenza seasonality in temperate zones, driving disease transmission and controlling the timing of individual wintertime outbreaks. Another study investigated the physical properties of absolute humidity on influenza virus transmission and influenza virus survival (3). However, a general understanding of the mechanisms underlying influenza seasonal variation remains quite limited (8). Here, we use a simple mathematical model to unravel the interplay between climate and evolution to predict long-term influenza dynamics correctly for the years since June 2010.

A requirement for the generation of recurrent epidemics is a sufficient and continuous source of new susceptible individuals arising in the population, enough to fuel each new outbreak (13). In the case of influenza, infected individuals recover with immunity but eventually become susceptible again because of the rapidly evolving nature of the influenza virus (7, 14). Positive selection exerted by the host immune system leads to a continual antigenic drift of the influenza virus's glycoproteins, particularly the main antigen, hemagglutinin, thus allowing the virus to eventually evade the immune system (15). The process of antigenic drift thereby creates an important renewed source of susceptible individuals. Hence, evolutionary forces are considered tremendously important in shaping complex recurrent patterns of infectious

diseases and explain why influenza is regarded as “an invariable disease caused by a variable virus” (1).

The changing rate of antigenic drift also has a significant impact on the timing and amplitude of influenza outbreaks (16). Recent studies reveal that the evolution of influenza A H3N2's main antigen is punctuated in character such that the drift occurs within discrete antigenic clusters (neutral periods), but with jumps to newly arising clusters after irregular periods (17, 18). A significant jump for the A H3N2 lineage last occurred during the 2003–4 season with the appearance of the A/Fujian virus strain coinciding with a sharp influenza outbreak approximately 2 mo earlier than usual, with a normal attack rate. Nevertheless, it is difficult to demonstrate a consistent and conclusive direct link between the size of antigenic jumps and changes in influenza dynamics at the population level (6, 19–21).

Results

Our modeling is based on a high-quality 11.5-y dataset from June 2001 to January 2013 of daily influenza-like illness (ILI) cases reported in Israel's largest city, Tel Aviv, as plotted in Fig. 1. The data were obtained from the Maccabi Health Maintenance Organization (HMO), whose medical surveillance covers ~45% of the local Tel Aviv population. An analysis of laboratory tests of ILI cases from sentinel clinics showed the ILI dataset to be highly correlated with the incidence of confirmed influenza cases (*SI Appendix*). Its high level of coverage and the data quality make this one of the finest available influenza surveillance datasets by world standards. We make use of the classical

Significance

Seasonality in disease incidence is ubiquitous among infectious diseases. Seasonal drivers include weather variables, such as temperature and humidity, and social factors (e.g., contact patterns). Attempts to make long-term predictions of infectious diseases are hampered by the inability to understand the complex interplay of these factors. Here, we model the dynamics of seasonal influenza based on a high-quality 12-year Israeli dataset. The dynamics are completely explainable by the time evolution of the model equations, the antigenic jumps of the virus, and the climatic forcing, yielding high-correlation fits to the data ($r = 0.94$). The forecasting ability is greatly increased through the predictability of the system's transient dynamics, resulting in accurate predictions ($r = 0.93$) that have not yet been found elsewhere.

Author contributions: J.B.A., R.Y., B.T.G., and L.S. designed research; J.B.A. and L.S. performed research; J.B.A. and L.S. contributed new reagents/analytic tools; J.B.A., R.Y., and L.S. analyzed data; and J.B.A., B.T.G., and L.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

¹To whom correspondence should be addressed. E-mail: lewistone2@gmail.com.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1321656111/-DCSupplemental.

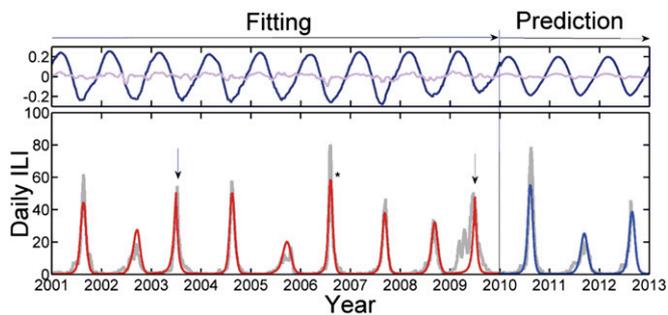


Fig. 1. (Lower) The smoothed daily ILI cases reported in Tel Aviv (light gray), with June 1 indicated on the x-axis for each year. Time series from the climate-driven deterministic SIRS model fitted to the ILI data (2001–2010) is shown (red), with fit correlation $r = 0.94$ to the period leading up to the pandemic in 2009. (Upper) The climate data consist of centered, normalized daily temperature (dark blue) and relative humidity readings (light violet), both scaled according to estimated weight. The pure prediction from June 1, 2010 is driven by the average climate of all years with fit $r = 0.93$. The outbreaks dominated by influenza B are indicated, and the asterisk highlights the abnormally high amplitude of 2006–2007. The epochal jumps in antigenic drift are indicated with arrows.

susceptible–infected–recovered–susceptible (SIRS) epidemic model (*Materials and Methods*) in which all individuals in the population are assumed to be in only one of three classes: susceptible (S), infected (I), and recovered (R). S individuals move to the I class after contact with an infective and transmission of the disease. Infective individuals eventually recover and move to the R class, whereas R individuals eventually become susceptible once more, closing the SIRS loop ($S \rightarrow I \rightarrow R \rightarrow S$) in a manner that mimics the effects of antigenic drift (14).

The model requires six basic parameters, which we provide with the best-fitting estimates (and $\sim 95\%$ credible intervals) found from the modeling procedure: (i) The parameter $\Lambda = 0.17 \pm 0.01$ describes the rate of immunity loss due to antigenic drift and the rate of new susceptible individuals entering the population per year. (ii) The reproduction number R_0 represents the number of infections transmitted by a typical individual over the lifetime of the disease in a wholly susceptible population. The model assumes that the basic reproduction rate, estimated as $R_0 = 2.9 \pm 0.1$, is constant for all years, as discussed further below. (iii) and (iv) Seasonal forcing is included by modulating R_0 with the locally observed temperature and relative humidity time series seen in Fig. 1 (Upper). The forcing $\delta(t)$ is specified by weights that control the influence of these climatic variables: temperature weight $w_T = 0.13 \pm 0.02$ and relative humidity weight $w_{RH} = 0.028 \pm 0.003$. Although for generality we include relative humidity and temperature as climate variables (22), similar but slightly inferior results also are obtained by using only absolute humidity as a single driver (3). (v) and (vi) Two separate parameters are used to take into account the special characteristics of years with epochal antigenic jumps. One parameter enhances the loss-of-immunity rate $\Lambda' = \Lambda + \Lambda^*$ over the entire year (*Materials and Methods* and *SI Appendix*). The other parameter makes it possible to accommodate the early arrival of outbreaks observed in epochal jump years, e.g., the A/Fujian outbreak in 2003–4. This may be achieved by enhancing R_{eff} over the first 6 mo of the season when the epidemic grows, by $\delta' = \delta + \Delta$. We fit the two parameters Λ^* and Δ over the epochal jump years 2003–2004 and 2009–2010 only. Finally, the model requires a single boundary condition, $S(t_0) = 0.31 \pm 0.02$ (fitted value), which represents the fractional size of the susceptible population on the first of June 2001, at the beginning of the modeling period. All model runs start out of season on the first of June; thus, we always have $I(t_0) = 0$ although there is a small continuous migration influx (*Materials and Methods*).

The simplicity of this elementary SIRS model would deem it unlikely to mimic Israel's complex influenza dynamics, as seen in Fig. 1. However, the model fit of the first 8 y of the data (June 2001 to June 2009) correlates with the observed data to an unusual degree, with correlation $r = 0.94$ (and $r = 0.90$ for the full fitting period, including the unusual pandemic year 2009/10). Note that all model time series correlations with observed data reported here are highly significant ($P < 0.001$).

To assess the quality of the fit further, we calculated the attack rates and the week of peak incidence in the data and the fit for each of the seasons (not including the last season, for which we have only partial data). We then computed the Pearson correlation between the data and the fit of these attributes. Fig. 2 shows the result of this analysis. For the attack rates, we obtained a correlation of $r = 0.74$ ($P = 0.01$). The correlation was driven down because of the pandemic season of 2009–2010, in which there were several unusual waves in a single year. Without this season, the correlation in the attack rates would exceed $r = 0.9$. For the peak epidemic week, we obtained a correlation of $r = 0.94$ ($P < 0.001$).

The excellent fit is a feature that also holds with the fully stochastic version of the model (*SI Appendix*). Because there are only six degrees of freedom and one boundary condition, this is not even one parameter per year. Also note that the large year-to-year variability in epidemic size and timing appears to be random in character, making it surprising that the mechanistic model can capture the intricacies of the data despite the major qualitative differences in the shape of the epidemic from year to year. For example, compare the large-scale symmetrically shaped epidemic in 2004–5 with the small, curtailed asymmetric outbreak in 2005–6. The model nevertheless captures both epidemics accurately. This accuracy is a result of the complex interaction among the changing supply of new susceptible individuals arising due to loss of immunity in the population through antigenic drift, the strong transient dynamics following the appearance of a new strain, and the timing of the climate signal each year (Fig. 1). Note that the years dominated by influenza B fall effortlessly within the model dynamics, providing possible population-level observational support for the notion of broader cross-reactivity of the immune response to influenza, as previously reported (23–25).

The recent global H1N1 pandemic in 2009–10 deviates the most from the normal influenza outbreak dynamics. Given that influenza epidemics almost always are triggered at the beginning of winter, one cannot expect our model to fit an outbreak beginning as early as April, as was the case for the H1N1 pandemic. The model can accommodate only the main peak of the outbreak, which occurred at the same time as the 2003–4 outbreak.

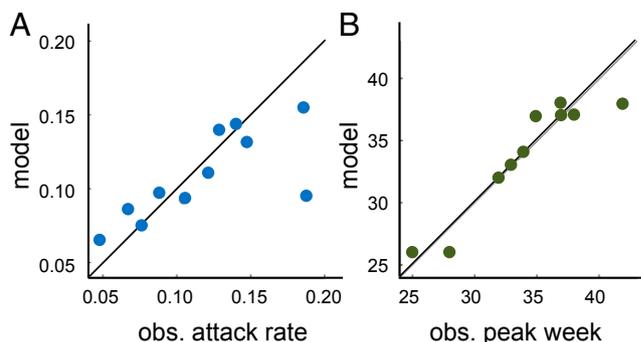


Fig. 2. Quality of model fit as assessed by correspondence of (A) attack rate at $r = 0.55$, $P = 0.01$. The main outlier in the match of attack rates is the H1N1 outbreak in 2009, without which the fit exceeds $r = 0.9$. (B) Peak epidemic week vs. model fit correspondence is $r = 0.94$, $P < 0.001$.

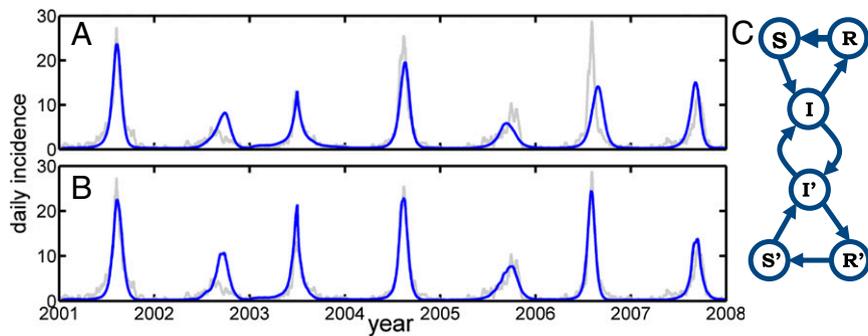


Fig. 4. Jerusalem ILI. (A) Data (gray) and model (blue) of a fit to Jerusalem influenza epidemics, $r = 0.84$. We fit Jerusalem and Tel Aviv simultaneously, although uncoupled, with the same universal R_0 . (B) The metapopulation model with mutual exchange of infectives between Tel Aviv and Jerusalem, $r = 0.94$. (C) Network diagram of the coupled populations.

access to high-accuracy prediction of local climate, which currently is not available.

In Fig. 3C, we compare the influenza dynamics following the appearance of new strains by plotting the 4 y beginning with the A/Fujian year in 2003–4 together with the 3.5 y beginning with the A/H1N1 pandemic in 2009–10. Also included are model runs (red and blue) based on a pure sinusoidal climatic driver. The data seemingly collapse in the sense that the observed data in both windows sit closely on top of each other, except for the fourth year. In other words, they form the same HHL(H) motif. It appears more than just coincidence that on the two occasions an epochal jump year appears, the double-period damping transient or the motif in the time series ensues, as this data collapse identifies. To explore the randomness of the underlying motif idea, we can assume that years have random independent peak heights (H, M, or L). The probability of the motif occurring by chance alone is significantly small, $P < 1\%$ (*SI Appendix*). Unsurprisingly, the major deviance of the data collapse—and the worst correspondence to the model—is the early part of the unusual 2009 H1N1 pandemic, during summer/fall of 2009 (visible as two “humps” before the main winter outbreak). In Fig. 3C, the fourth-year predictions also are degraded, as the transient oscillations will have died down to a level at which the climate starts to dominate again. This explains the prediction errors in Fig. 3B, in which sinusoidal rather than true climatic forcing is used. The true climatic signal provides a superior prediction of the 2006–7 outbreak (compare Fig. 1 with Fig. 3B and *SI Appendix*). Similarly, by using average climate data, we successfully trace the 2012–13 outbreak, predicting a normal outbreak initiation time and matching the amplitude peak closely (Fig. 1, time series end).

That the simple model fits the data well raises many questions. Most plausibly, the population mixing inside Israel conforms closely to the basic SIRS model assumptions so that the mean field dynamics provide a faithful caricature of influenza transmission and immunity loss in a large population. Historically, the success of the SIR model for single epidemics derives from this feature, but what is new here is the unusual long-term year-to-year predictability reproduced in addition. It would appear that the dynamics are completely explainable by the combined time evolution of the model equations, the antigenic jumps, and the climatic forcing. The hypothesis outlined here prescribes not only that climate is a key seasonal driver of influenza in semi-isolated coastal Mediterranean populations (Tel Aviv) with hot and humid climates, but also that one city may act as a hub for disease and drive satellites. The latter of these statements we derive from the fact that the model fit to Jerusalem influenza dynamics improves considerably when coupled with Tel Aviv, but not vice versa. The forecasting ability is increased strongly through the predictability of the system’s transient dynamics.

These characteristics serve to enhance the effect of climate on influenza spread in a truly unique and predictable fashion, and they help explain why the quality of our model fits has not yet been found elsewhere.

Materials and Methods

The ILI data were obtained from the Maccabi HMO for the period June 1, 2001 to January 2013. The period from June 2001 marked a fundamental switch in surveillance and coding used by the HMO. We smoothed the data by interpolating the missing weekend reports and subsequently using a wavelet approximation to minimize noise (*SI Appendix, section 3*). All ILI and climate datasets may be obtained from the following link: www.tau.ac.il/~lewi/Data_flu_climate.txt.

Climate data of relative humidity and temperature in Tel Aviv were obtained from monitoring stations of the Israeli Meteorological Society and the Ministry of Environmental Protection of Israel. The time series were smoothed analogously to the ILI time series above. The resulting time series then were centered to zero mean and normalized to unit variance, which is why there are no units in Fig. 1, *Upper*.

The dynamics of the epidemic outbreaks were modeled based on a deterministic age-of-infection SIR model (28), which we extended to an SIRS model by adding a simple feedback from the recovered population to the susceptible population via loss of immunity caused by steady antigenic drift. In this model, the number of newly infected at time t is

$$i(t) = \frac{S(t)}{N} R_0 (1 + \delta(t)) \sum_{\tau=1}^d P_{\tau} i(t - \tau), \quad [2]$$

where S is the number of susceptible, R_0 is the basic reproduction rate of a fully susceptible population, P is the 7-d infectivity profile ($d = 7$) chosen as a gamma distribution with a mean of 2.7 d and variance of 1.8 d (11), and N is the reporting population at a reporting rate of 10%. The average period of immunity of a typical individual immunity, here $1/\Lambda$, usually is 2–10 y ($\Lambda \in [0.1: 0.5]$). For a model with a daily time step, we set $\lambda = \Lambda/365$. The infectives pool follows: $I(t) = I(t - 1) + i(t) - I(t - d)$; the dynamics of the susceptible pool then is

$$S(t) = S(t - 1) - I(t) + \lambda R(t - 1). \quad [3]$$

The infected individuals recover after d days; thus, the recovered population growth is

$$R(t) = R(t - 1) + I(t - d) - \lambda R(t - 1). \quad [4]$$

Finally, there also is a constant influx of influenza arriving in Tel Aviv from any destination in the world, which was set at one person per day all year round. The influenza influx cannot be higher than 10 per day, as this gives a model baseline ILI above one influenza case per day in the low season ($10 \times 10\% = 1$, taking into account a 10% reporting rate). Conversely, an influx rate much below 0.1 sick persons per day eventually will render the model unable to reproduce the “small” outbreaks because the infected and susceptible populations may never reach the epidemic threshold $R_{\text{eff}} = 1$ in those years.

In this model, the disease dynamics are modeled neutrally for all strains and compositions thereof between punctuations, i.e., the ILI cases are treated

as one disease only. Although this is a nontrivial approximation, it is not uncommon in the seasonal flu-modeling literature (4, 10, 11, 29); furthermore, recent findings indicate the possibility of broad cross-reactivity, further justifying the approximation. These recent studies seem to find in vivo and in vitro evidence of functional cross-reactivity between human monoclonal antibodies and epitopes from hemagglutinin of both influenza A and influenza B (30).

The simple climate driver is modeled as $R_0 = \bar{R}_0[1 + \delta \sin(\omega t)]$. The real climate driver is expressed as a linear combination; thus, $\delta(t) = w_T T(t) + w_{RH} RH(t)$, where T is daily average temperatures and RH is relative humidity, both normalized to the interval $[-0.5, 0.5]$. The antigenic jump is a modification first to $\delta'(t) = \delta(t) + \Delta$, $\Delta = 0.26(0.02)$, for the period June 1 to December 1. A modification in immunity-loss rate $\lambda' = \lambda + \lambda^*$, $\lambda^* = 0.04(0.01)$ is modeled throughout the year.

Inference is performed by Bayesian Monte Carlo sampling of the a priori likelihood function; see *SI Appendix* for details and analysis.

ACKNOWLEDGMENTS. We thank Drs. Jodie McVernon, James McCaw, Haggai Katriel, and Amit Huppert. We are grateful for the support of the European Union FP7 EPIWORK grant, the Israel Science Foundation, and the Israel Ministry of Health. J.B.A. is supported by the Carlsberg Foundation. R.Y. is supported by the Israel National Institute for Health Policy and Health Services Research. We appreciate the support of Maccabi Health Care Services for providing the ILI data. B.T.G. was supported by the Bill and Melinda Gates Foundation, the Science and Technology Directorate, Department of Homeland Security Contract HSHQDC-12-C-00058, and the Research and Policy in Infectious Disease Dynamics program of the Science and Technology Directorate, Department of Homeland Security, and the Fogarty International Center, National Institutes of Health.

- Potter CW (2001) A history of influenza. *J Appl Microbiol* 91(4):572–579.
- Cox NJ, Subbarao K (2000) Global epidemiology of influenza: past and present. *Annu Rev Med* 51:407–421.
- Shaman J, Kohn M (2009) Absolute humidity modulates influenza survival, transmission, and seasonality. *Proc Natl Acad Sci USA* 106(9):3243–3248.
- Finkenstädt BF, Morton A, Rand DA (2005) Modelling antigenic drift in weekly flu incidence. *Stat Med* 24(22):3447–3461.
- Finkelman BS, et al. (2007) Global patterns in seasonal activity of influenza A/H3N2, A/H1N1, and B from 1997 to 2005: Viral coexistence and latitudinal gradients. *PLoS One* 2(12):e1296.
- Chowell G, Miller MA, Viboud C (2008) Seasonal influenza in the United States, France, and Australia: Transmission and prospects for control. *Epidemiol Infect* 136(6): 852–864.
- Earn DJD, Dushoff J, Levin SA (2002) Ecology and evolution of the flu. *Trends in Ecology and Evolution* 17(7):334–340.
- Lipsitch M, Viboud C (2009) Influenza seasonality: Lifting the fog. *Proc Natl Acad Sci USA* 106(10):3645–3646.
- Shaman J, Pitzer VE, Viboud C, Grenfell BT, Lipsitch M (2010) Absolute humidity and the seasonal onset of influenza in the continental United States. *PLoS Biol* 8(2): e1000316.
- van Noort S, Águas R, Ballesteros S, Gomes M (2011) The role of weather on the relation between influenza and influenza-like illness. *J Theor Biol* 298:131–137.
- Truscott J, et al. (2012) Essential epidemiological mechanisms underpinning the transmission dynamics of seasonal influenza. *J R Soc Interface* 9(67):304–312.
- Yaari R, Katriel G, Huppert A, Axelsen JB, Stone L (2013) Modelling seasonal influenza: The role of weather and punctuated antigenic drift. *J R Soc Interface* 10(84): 20130298.
- Stone L, Olinsky R, Huppert A (2007) Seasonal dynamics of recurrent epidemics. *Nature* 446(7135):533–536.
- Pease CM (1987) An evolutionary epidemiological mechanism, with applications to type A influenza. *Theor Popul Biol* 31(3):422–452.
- Nelson MI, Holmes EC (2007) The evolution of epidemic influenza. *Nat Rev Genet* 8(3): 196–205.
- Koelle K, Cobey S, Grenfell B, Pascual M (2006) Epochal evolution shapes the phylodynamics of inter-pandemic influenza A (H3N2) in humans. *Science* 314(5807): 1898–1903.
- Smith DJ, et al. (2004) Mapping the antigenic and genetic evolution of influenza virus. *Science* 305(5682):371–376.
- Rambaut A, et al. (2008) The genomic and epidemiological dynamics of human influenza A virus. *Nature* 453(7195):615–619.
- Wu A, Peng Y, Du X, Shu Y, Jiang T (2010) Correlation of influenza virus excess mortality with antigenic variation: application to rapid estimation of influenza mortality burden. *PLoS Comput Biol* 6(8):e1000882.
- Wolf YI, et al. (2010) Projection of seasonal influenza severity from sequence and serological data. *PLoS Curr* 2:RRN1200.
- Viboud C, et al. (2006) Transmissibility and mortality impact of epidemic and pandemic influenza, with emphasis on the unusually deadly 1951 epidemic. *Vaccine* 24(44-46):6701–6707.
- Earn DJ, et al. (2012) Effects of school closure on incidence of pandemic influenza in Alberta, Canada. *Ann Intern Med* 156(3):173–181.
- Lavenu A, Valleron AJ, Carrat F (2004) Exploring cross-protection between influenza strains by an epidemiological model. *Virus Res* 103(1-2):101–105.
- Cobey SE (2009) Ecological and evolutionary dynamics of influenza viruses. PhD thesis (Univ of Michigan, Ann Arbor).
- Towers S, Feng Z, Hupert N (2010) Short-term heterologous immunity after severe influenza A outbreaks. arXiv:1007.3017.
- Earn DJ, Rohani P, Bolker BM, Grenfell BT (2000) A simple model for complex dynamical transitions in epidemics. *Science* 287(5453):667–670.
- Aron JL, Schwartz IB (1984) Seasonality and period-doubling bifurcations in an epidemic model. *J Theor Biol* 110(4):665–679.
- Katriel G, Yaari R, Huppert A, Roll U, Stone L (2011) Modelling the initial phase of an epidemic using incidence and infection network data: 2009 H1N1 pandemic in Israel as a case study. *J R Soc Interface* 8(59):856–867.
- Xia Y, Gog JR, Grenfell BT (2005) Semiparametric estimation of the duration of immunity from infectious disease time series: Influenza as a case-study. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 54:659–672.
- Dreyfus C, et al. (2012) Highly conserved protective epitopes on influenza B viruses. *Science* 337(6100):1343–1348.