Reconstructing influenza incidence by deconvolution of daily mortality time series

Edward Goldstein[a,1], Jonathan Dushoff[b,c], Junling Ma[d], Joshua B. Plotkin[e], David J. D. Earn[c,d], and Marc Lipsitch[a]

[a]Center for Communicable Disease Dynamics, Department of Epidemiology, Harvard School of Public Health, Boston, MA 02115; [b]Department of Biology, McMaster University, Hamilton, ON, Canada L8S 4K1; [c]M. G. Degroote Institute for Infectious Disease Research, McMaster University, Hamilton, ON, Canada L8N 3Z5; [d]Department of Mathematics and Statistics, McMaster University, Hamilton, ON, Canada L8S 4K1; and [e]Department of Biology, University of Pennsylvania, Philadelphia, PA 19104

We propose a mathematically straightforward method to infer the incidence curve of an epidemic from a recorded daily death curve and time-to-death distribution; the method is based on the Richardson–Lucy deconvolution scheme from optics. We apply the method to reconstruct the incidence curves for the 1918 influenza epidemic in Philadelphia and New York State. The incidence curves are then used to estimate epidemiological quantities, such as daily reproductive numbers and infectivity ratios. We found that during a brief period before the official control measures were implemented in Philadelphia, the drop in the daily number of new infections due to an average infector was much larger than expected from the depletion of susceptibles during that period; this finding was subjected to extensive sensitivity analysis. Combining this with recorded evidence about public behavior, we conclude that public awareness and change in behavior is likely to have had a major role in the slowdown of the epidemic even in a city whose response to the 1918 influenza epidemic is considered to have been among the worst in the U.S.

1918 pandemic | incidence curve | death curve | Richardson–Lucy deconvolution | infectivity ratios

In characterizing a newly emerging or historical epidemic, one often has access to an epidemic curve that provides the number of persons who became ill on a certain day (the symptom curve), or the number of cases that were reported each day (the report curve), or the number of dying that were reported each day (the death curve). Of more direct interest, for the purposes of visualizing the spread of the epidemic and calculating relevant quantities such as the daily reproductive number, is the (usually unobserved) epidemic curve of the number of persons becoming infected on each day (the incidence curve). The other three curves—the symptom, report, and death curves—provide information about the incidence curve but are imperfect representations of it, first because not all cases may appear in these other curves—asymptomatics, unreported cases, or nonfatal cases respectively will be missed—and perhaps more importantly because the delay between infection and subsequent events—symptom onset, report, or death—is a random variable that adds horizontal variation or "smear" to the epidemic curve. For some infections (e.g., HIV), diagnostic symptoms (i.e., AIDS-defining illness) may occur years after infection, so the symptom curve is a poor reflection of the evolution of the epidemic; there is much literature on the problem of deconvolution to estimate the HIV incidence curve from the symptom (AIDS-incidence) curve (1). For acute infections, such as SARS and influenza, the incubation period is relatively short compared with the growth rate of the epidemic; hence the symptom curve and the incidence curve are similar. The time to death for such infections, however, can be several weeks, equivalent to three or more disease generations, and is highly variable (2–4), making the death curve a poor surrogate for the incidence curve.

In this paper, we propose a mathematically straightforward method to infer the incidence curve from a death curve. The incidence curve is then used to estimate the basic epidemiological

quantities of interest, such as the daily reproductive numbers and infectivity ratios; the latter can be utilized to assess major changes in the epidemic's dynamics.

First, we describe a technique for deconvolution to estimate the incidence curve from the death curve and the time-to-death distribution. The technique, called the Richardson-Lucy (RL) deconvolution, was originally developed for use in optics (5, 6), and is adapted in a simple way to the slightly different setting of the death-to-incidence deconvolution problem. The technique is illustrated for the 1918 influenza epidemic in Philadelphia and New York State (2, 7, 8).

Second, we use the reconstructed incidence curves to perform inference on daily reproductive numbers in that epidemic. One level of difficulty arises from the fact that because of the rapid progression of the epidemic, there is saturation of susceptibles and a possible change in behavior during the course of infection for persons infected on a given day (9). To deal with this issue, we have rederived the Wallinga–Teunis estimator by using an approach similar to the one in ref. 10. The derivation uses daily infectivity ratios, a concept which essentially appeared before in refs. 10 and 11.

Third, we examine the spread of pandemic influenza in the city of Philadelphia around the end of September and the beginning of October, 1918. On September 28, a 200,000-person Liberty Loan Drive took place on the streets of Philadelphia against the advice of medical professionals (12). Within 72 hours, "every single bed in the city's 31 hospitals was filled" (ref. 13, p. 220). By October 1, residents were often encouraged to stay home to stem spread of the disease (14). On the evening of October 3, the closure of schools, churches, and places of public amusement was adopted by the Philadelphia city council (ref. 15, p. 74). The deconvolved incidence curve, which peaks around October 1–2, shows a drastic change in the growth patterns/infectivity ratios between September 26 and October 3. In particular, the infectivity ratios, representing the average number of infections by an infector on a given day, dropped by more than half. We estimate, assuming that case-fatality ratio is at least 2% (16, 17), that during this period depletion of susceptibles was at most 16% (here and elsewhere, depletion is characterized in terms of the number of susceptibles on September 26, which equals Philadelphia's population minus the number of infected by September 25). This difference suggests that depletion of susceptibles played only a modest role in slowing the growth of the epidemic during that period, which took place before the closures went into effect. Our main conclusion, which is not merely of historical interest, is that public awareness

MEDICAL SCIENCES

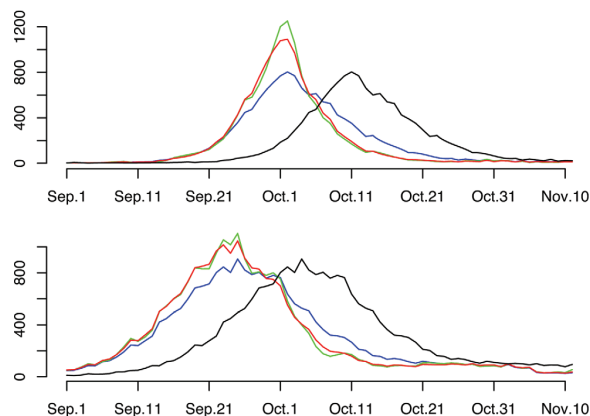**1918 influenza epidemic in Philadelphia (above) and NY State (below)**

Fig. 1. The influenza epidemic in Philadelphia (*Upper*) and New York State (*Lower*) in 1918. Black is death curve. Blue is the initial condition for RL iterations (death curve, shifted back by nine days). Red is the estimated incidence (scaled by a factor of *p*, where *p* is the case fatality ratio), obtained from the first iteration for which the χ^2 value from Eq. **3** is below 1 (sixth iteration for Philadelphia and fourth iteration for New York State). Green are results of RL after 30 iterations.

and changes in behavior are likely to have had a major role in the slowdown of the epidemic before the official control measures were in place.

In the *SI Appendix* we use simulations and sensitivity analyses to test the robustness of our methodology/conclusions. We examine how well the deconvolved incidence curve represents the original one, and what is the best way to estimate the growth rate and the initial reproductive number of the original incidence curve. We also study the epidemic progression in Philadelphia between September 26 and October 3, considering various "incidence" curves deconvolved by using a collection of randomly generated time-to-death distributions and also allowing for time dependency of the case-fatality ratios due to potential changes in demographics of the infected. Our general conclusion remains valid in all cases: Within that period, a drastic change in the epidemic's dynamics took place, and depletion of susceptibles was probably insufficient to explain that development.

Results

Reconstructing the Daily Incidence Curves for the 1918 Influenza Epidemic. We apply the procedure from the *Methods* section to the 1918 influenza and pneumonia death curves in Philadelphia and New York State (excluding the city) to obtain the daily incidence in those places. Fig. 1 depicts the reconstructed incidence curves, which are scaled by a factor of *p*, where *p* is the case fatality ratio. Notably, the peaks of the incidence curves are shifted backward by approximately nine days, but the incidence curves have narrower peaks with steeper upward and downward trajectories as one would expect when removing the smear induced by the time-to-death distribution, which is analogous to removing the blur due to camera motion from a photographic image.

Estimates of the Daily Reproductive Numbers. The effective reproductive number R_t on day *t* measures the mean number of infections caused by persons who become infected on day *t*. Fig. 2 plots the estimated reproductive numbers for certain periods of the epidemic's progression, in both Philadelphia and New York State. The reproductive numbers fluctuate initially because of the small values for the deconvolved incidence. The reproductive numbers descend in the later part of the exponential growth period, representing future infections which happen under saturation of susceptibles or behavior change.

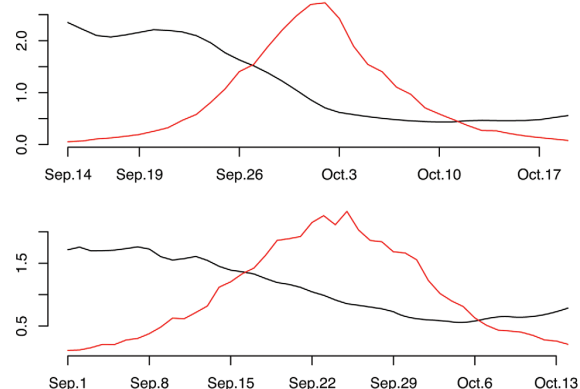**Daily reproductive numbers, Philadelphia (above) and NY State (below)**

Fig. 2. Epidemic progression in Philadelphia (*Upper*) and New York State (*Lower*), for certain time periods. Red is deconvolved incidence (scaled), and black represents the estimated reproductive numbers.

With the fluctuation of the reproductive numbers, one may wonder what is the best way to estimate R_0, the mean number of new infections caused by an infected individual during the early, exponential growth stage of the actual epidemic curve in Philadelphia. We address this via simulations in the *SI Appendix*. Our estimate is $R_0 = 2.14$, with a standard error of 0.13.

Epidemic Peak in Philadelphia. In this section, we examine the peak of the Philadelphia epidemic in more detail. We first compute the daily infectivity ratios, representing the number of infections on a given day caused by a weighted average of previously infected individuals (Eq. **4**). Fig. 3 plots the estimated infectivity ratios for the period of September 14–October 19.

The infectivity ratio IR_t on a given day *t* presents a basic assessment of the epidemic's state on day *t*; in a mass action model, the infectivity ratio is proportional to the number of susceptibles left on that day. The reproductive number R_t on day *t* measures the number of infections caused during an infectious period of an average person who was infected on day *t*; it may be harder to relate R_t to the epidemic's state on day *t* because during that infectious period further depletion of susceptibles may occur, and conditions related to the epidemic's progression, such as an introduction of control measures and an increase in public awareness, may take place.

The estimated infectivity ratios dropped by more than half between September 26 and October 3. At the same time, depletion of susceptibles was much smaller. Philadelphia had a population of 1.7 million, with another 300,000 added by the war industry (15). As explained in Methods, the estimated incidence curve plotted in Figure 1 is scaled by a factor of *p*, where *p* is the case fatality ratio. Various estimates of *p*, according to location, age, gender and

Daily infectivity ratios for the 1918 influenza epidemic in Philadelphia
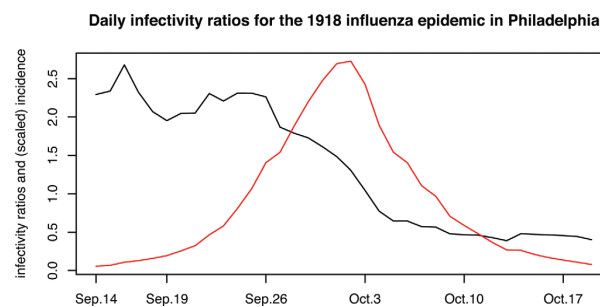
Fig. 3. Estimated infectivity ratios (black) vs. (scaled) incidence (red) in Philadephia, September 14–October 19, 1918.

race, are given in refs. 16 and 17. There is considerable geographic variation in case-fatality ratios, and for the surveyed communities in the Northeastern United States (from Baltimore, MD to New London, CT) case-fatality ratios were above 2%, topping 3% in some places. Assuming $P \geq 0.02$ for Philadelphia and by using the deconvolved incidence curve, we estimate that at most 16% of the susceptible population was depleted between September 26 and October 3, which cannot account for the over 50% drop in the infectivity ratios during that period; the latter finding was subjected to extensive sensitivity analysis with regard to the time-to-death distribution and other parameters (see the *SI Appendix* for more details). At the same time, we have evidence from refs. 12 and 14 about public awareness of the epidemic prior to October 3. We conclude that public awareness and change in behavior likely played a major role in the slowdown of the epidemic in Philadelphia before the official control measures were implemented on October 4.

Discussion

We have described and demonstrated the usefulness of a RL-type deconvolution approach to reconstruct incidence curves from epidemic curves showing the times of death in an epidemic. Applied to the 1918 influenza data, this approach reconstructs an epidemic curve that is more sharply peaked than the death curve—as one would expect given the smear introduced by the large variance in the time-to-death distribution. Moreover, this reconstructed curve allows the direct estimation of a daily value for the reproductive number R_t, bypassing the more-cumbersome methods used previously, which involved extensive simulations and allowed the estimation of only a single value for R (4), or a parametric estimation of R_t (18). The estimate of R_t is given by the Wallinga–Teunis formula (9). Our derivation of this estimate relies on the fact that the number of infected individuals is large hence the notion of an average-infectiousness profile makes sense; this is different from the original setting of an emerging epidemic in ref. 9, where it is assumed that all infected individuals had the same infectiousness profile. Finally, the deconvolved incidence curve was used to show that public awareness and change in behavior is likely to have played a major role in the slowdown of the epidemic even in Philadelphia, a city whose response to the 1918 influenza epidemic is considered to have been among the worst in the U.S. (13, 19). A related point regarding the belatedness of official control measures in Philadelphia was made in ref. 18.

The latter qualitative conclusion was subjected to extensive sensitivity analysis. The population of Philadelphia was not homogeneous, with case-fatality ratios, time-to-death distributions, infectivity and susceptibility to infection varying by age and even gender. Although we cannot recover the complexity of the population network in Philadelphia, we have tested our observation that the drop in infectivity ratios during the week between September 26 and October 3 greatly exceeded the depletion of susceptibles under various scenarios; we have examined the sensitivity of that conclusion with respect to the time-to-death distribution, and allowed for certain forms of time dependency for the case-fatality ratios, reflecting upon potential changes in the demographics of the infected as the epidemic progressed. The main conclusion persisted with remarkable stability in the simulations we have performed. Furthermore, visual inspection of the growth rate of the death curve (Fig. 4) shows a major change in its slope by October 5–6, which was likely predated by major changes in incidence patterns. Combining this with contemporary journalistic evidence (14) suggests that our hypothesis about the change in behavior before the official control measures were implemented is likely to be real.

We have not considered asymptomatic infections in our analysis. Data from the 1957 and the 1968 influenza epidemics yielded estimates of 20–42% for the rate of asymptomatic infections (20, 21). A review of several volunteer challenge studies (22) gives an estimate of 33% for the rate of asymptomatic infections. There is no
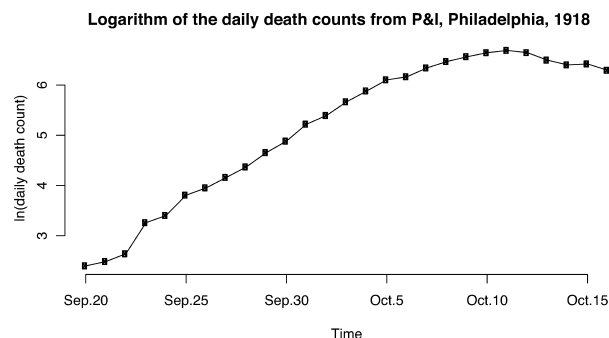
Logarithm of the daily death counts from P&I, Philadelphia, 1918

Fig. 4. Natural logarithm of the recorded daily number of deaths from pneumonia and influenza in Philadelphia, 1918.

data for the rate of asymptomatic infections for the 1918 influenza pandemic. Given its virulence, that rate could have been lower than the rates above. In our simulations, the drop in infectivity ratios between September 26 and October 3 surpasses depletion of susceptibles by a factor of at least 2.5. Thus, even if a third of all infections were asymptomatic, depletion of susceptibles during that period cannot explain the drop in infectivity ratios.

Our approach in principle gives a general method for reconstructing an incidence curve from the death curve. In practice, to have reliable estimates one needs the standard errors of the counts to be small as compared with the means—say, if the (Poisson) counts reach into the hundreds. We recommend performing simulations following the protocol in the *SI Appendix* to see how well the deconvolution process works.

We note that estimated infectivity ratios in Philadelphia (Fig. 3) first drop on September 27, one day before the notorious Liberty Loan parade, and seven days before the implementation of official control measures. Although reductions in transmission in the absence of official control measures have been inferred from the dynamics of influenza in U.S. cities by a previous study (18), this finding for Philadelphia is surprising, given that the parade is often interpreted as a sign that Philadelphia residents had not yet recognized the seriousness of the pandemic, and, moreover, that the parade is often seen as an opportunity for large-scale transmission. If we have accurately estimated the timing of the decline in infectivity ratios, then the parade was probably not in fact a major venue for transmission, perhaps because it took place in the open air, and we must infer that behavior changes made a contribution to the slowing of the epidemic prior to official control measures. We have considered the possibility that the timing is estimated incorrectly. For example, it is possible that the time-to-death distribution we have employed is too long, resulting in artificially early changes in the incidence curve to reflect changes in the death curve. We believe that a shorter time-to-death distribution is unlikely because we are unaware of any reliable data suggesting that the time-to-death distribution is substantially shorter than the one we have adopted from ref. 2; in fact, the distribution we used is shorter than the one obtained from military data and used previously in refs. 4 and 18 (see *SI Appendix*). A likely source of error in estimating the exact timing of declines in infectivity ratios might be the nature of the deconvolution process, as seen in simulations in the *SI Appendix*. Briefly, deconvolved incidence curves have an exponential growth period ending about two days before the original one.

One may wonder if it is possible to use the death curve directly to infer epidemiological parameters of interest, such as R_0. One can try the exponential growth rate approach (see the *SI Appendix*). In theory, given a long exponential growth period for incidence (long in comparison to the time-to-death distribution), the death curve would have the same exponential growth rate. However for the Philadelphia data, deconvolution shows that the death

curve growth rate is somewhat smaller than the incidence curve growth rate, and the estimate of R_0 resulting from the death curve growth rate ($\simeq 2 - 2.02$) is biased downward. Alternatively, one can consider the Wallinga and Teunis approach for estimating the reproductive numbers by using the "death-to-death" distribution. The latter can be understood as follows: For people who died on day t, one can look at the times of death of the individuals they've infected. However this distribution depends on the state of the epidemic around time t, and thus changes with t. Moreover, assuming that the serial interval distribution is independent for different pairs of cases, the death-to-death distribution will not be independent for successive pairs of cases (e.g., A who infects B who infects C), violating the assumptions of the approach. At the same time, we found that having information about the incidence curve can be used for a number of purposes besides estimating R_0. Our estimates of the daily reproductive numbers and assessment of the rapid decline of infectivity ratios around the peak of the epidemic in Philadelphia may not be accessible directly from the death curve.

It is also worthwhile to compare the approach to estimating R_t here to an approach which assumes that future opportunities for infection for each individual, regardless of when they are infected, will follow a fixed distribution going forward in time. One might expect that the relation between the reproductive numbers, the serial interval (w_i) (see *Methods*), and new infections is that for an average person infected on day t, the expected number $E_{t,i}$ of people s/he will infect on day $t + i$ is

$$E_{t,i} = R_t \cdot w_i. \qquad [1]$$

The relation above is known to hold during periods when there is little to no depletion of susceptibles and behavior change (23)—during that period, R_t is in fact constant. However this may be untenable for the whole duration of an epidemic, where the number of susceptibles may decline over the course of a single individual's infectious period (24); similarly, measures to control the epidemic and behavior change may take place during that period. To put the issue a bit more precisely, consider the ratio $\frac{E_{t,i_1}}{E_{t,i_2}}$. If Eq. 1 was correct, this ratio would equal $\frac{w_{i1}}{w_{i2}}$ and would be independent of the day t of the infection. However, when the epidemic progresses very rapidly, populations of susceptibles (and their possible behavior) can be quite different on days $t + i_1$ and $t + i_2$. Thus, it is unreasonable to expect that this ratio is independent of t. The formulation proposed here, and in ref. 9, avoids this pitfall.

Methods

Richardson–Lucy-Type Deconvolution. The deconvolution problem is to assess the daily incidence curve (I_t) given the daily death curve ($D_1, \ldots, D_N$) and time from-infection-to-death distribution ($d_1, \ldots, d_l$). Here, D_j is the number of deaths recorded on day j; and d_k represents the probability that an infected person who will eventually die, will die on day k after his/her infection. The daily death curves for Philadelphia and New York State (excluding city) are recorded in refs. 7 and 8. The time from-infection-to-death distribution is obtained as a convolution of two distributions: time from symptom onset to death, taken from ref. 2, based on 599 hospital autopsy reports and the influenza latent period distribution, taken from ref. 25. It is plotted in Fig. 5.

To assess incidence I_t on day t, let p be the probability that an infected individual will die from influenza and pneumonia; in the *SI Appendix*, we allow for time dependence of p due to potential shifts in demographics of the infected. The number of people who got infected on day t and later died is binomial $Bin(I_t, p)$; because I_t is large and p is small, it is well approximated by a Poisson variable $Pois(\lambda_t)$ with (an unknown) mean $\lambda_t = p \cdot I_t$. Those $Pois(\lambda_t)$ deaths are binned over the subsequent days according to the time-to-death distribution.

We wish to estimate the daily incidence curve during some time period $(t_1, \ldots, t_2)$, which overlaps with the period $(1, \ldots, N)$ for which the daily counts for the number of deaths are available ($N = 92$ for Philadelphia and $N = 75$ for New York State). Because over 95% of deaths happen within three weeks of infection, we want to estimate incidence from day $t_1 = -20$ (three weeks prior to the available death data); we also take $t_2 = N - 2$ as $d_1 = 0$ for
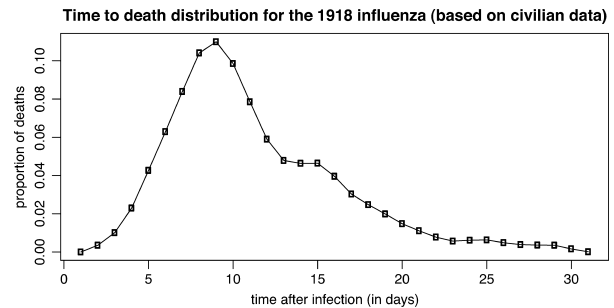
Fig. 5. Assumed proportions of deaths on each day since infection. Based on the Cook County Hospital data from ref. 2 for time from symptom onset to death and the latent period distribution from ref. 25.

our time-to-death distribution. Thus, the number of unknown parameters is $N + 19$, exceeding the number N of observations.

To assess the vector of unknown parameters ($\lambda_{t_1}, \ldots, \lambda_{t_2}$) (thus estimating the incidence curve up to a multiple), we iterate in the space of parameters by using the expectation maximization algorithm (26); this procedure is called the RL iteration. The initial guess $\lambda^0 = (\lambda_{t_1}^0, \ldots, \lambda_{t_2}^0)$ for the parameters is the death curve, shifted back by nine days, as the time-to-death distribution ($d_1, \ldots, d_{31}$) peaks on day nine after infection (see Fig. 4). RL iterations produce a sequence $\lambda^n = (\lambda_{t_1}^n, \ldots, \lambda_{t_2}^n)$ of the Poisson parameters. Explicitly, let $q_j = \sum_{-j+1 \leq i \leq N-j} d_i$ be the probability that a death resulting from incidence on day j will be observed during the interval $1, \ldots, N$; here $-20 \leq j \leq N - 2$ and $d_i = 0$ for $i \leq 0$ or $i \geq 32$. Let $D_i^n = \sum_{j<i} d_{i-j} \lambda_j^n$ be the expected number of deaths to occur on day i, conditional on the parameters λ^n. Then,

$$\lambda_j^{n+1} = \frac{\lambda_j^n}{q_j} \cdot \sum_{i>j} \frac{d_{i-j} D_i}{D_i^n}. \qquad [2]$$

The probability of observations $D_1, \ldots, D_N$, conditional on the Poisson parameters λ^n and the time-to-death distribution, increases with each iteration. While iterating, we do not seek convergence to a maximum likelihood solution; rather, we use a criterion to end iterations and produce our estimate of the Poisson parameters. To understand this, for the nth iteration, let $E_i^n = \sum_{j<i} \lambda_j^n d_{i-j}$ be the expected number of deaths on day i. If λ^n were the true parameters, D_i would be Poisson distributed with mean E_i^n; thus, the expectation $E(\frac{(E_i^n - D_i)^2}{E_i^n})$ would be 1. This suggests to iterate until the normalized χ^2 statistic

$$\chi^2 = \frac{1}{N} \sum_i \frac{(E_i^n - D_i)^2}{E_i^n} \qquad [3]$$

descends below 1 for the first time.

Bootsrapping simulations in the *SI Appendix* address the question of how closely a deconvolved curve resembles the original incidence curve.

Estimating Reproductive Numbers and Infectivity Ratios. In this section, we show how to estimate the daily reproductive numbers for the 1918 influenza from incidence curves and the infectiousness profile distribution. We remark that the words "infectiousness profile" and "serial interval" are often used interchangeably, though this is not quite accurate (see ref. 27, where "infectiousness profile distribution" is called "infectious contact distribution"). We prefer the term "infectiousness profile distribution" because it reflects upon average individual infectivity, which depends only on the time since one became infected, unlike the proportion of infectious contacts on each day after infection or the time between one's infection and the infection of one's infector.

We set a cutoff of 10 days for the infectivity process. The infectiousness profile distribution ($w_1, \ldots, w_{10}$), essentially taken from ref. 25, is plotted in the *SI Appendix*. Here each number w_i represents the proportion of the cumulative infectiousness which falls between days ($i - 0.5, i + 0.5$) for an average person. Note that, conveniently, the latent period in ref. 25 has an offset of 0.5 days.

By using the infectiousness profile distribution and the incidence curve, we can describe some key epidemiological parameters of interest. The infectivity ratio IR_t on day t measures the number of people infected by an "average" infector on day t:

$$IR_t = \frac{I_t}{\sum_{i<t} I_i \cdot w_{t-i}}. \qquad [4]$$

This concept was already used in ref. 10, where it was denoted by $\phi_1(t)$, and in ref. 11, where it was denoted by R_t. If the population of susceptibles and their behavior does not change much over a reasonable interval (covering the time of one individual's infectiousness history), the infectivity ratio would equal the reproductive number. This is not the case for the epidemics in question. However, because the number of infected is large, we can compute the reproductive number R_t on day t in a "forward-looking" way, by adding up the numbers of infections caused in subsequent days by an average person who got infected on day t:

$$R_t = \sum_{i>0} \text{IR}_{t+i} \cdot w_i = \sum_{i>0} \frac{w_i I_{t+i}}{\sum_{l=i-10}^{i-1} w_{i-l} I_{t+l}}. \quad [5]$$

This is the Wallinga–Teunis formula, and its derivation essentially follows ref. 11. Note that this derivation relies on the fact that the number of infected is large, hence the notion of an average infectiousness profile makes sense; this is different from the original setting of an emerging epidemic in ref. 9,

where it is assumed that all the infected have the same infectiousness profile. The two derivations help to clarify the notion that the reproductive number R_t is forward-looking, taking account of all the events that may happen after time t that will affect how many secondary cases an individual infected at t will infect. By contrast, the infectivity ratio looks backward from the day on which infection occurs. In the case of exponential growth (where opportunities for transmission remain constant over time), the two measures are equivalent.

1. Brookmeyer R, Gail M (1994) *AIDS Epidemiology: A Quantitative Approach*. (Oxford Univ Press, Oxford).
2. Brundage JF, Shanks D (2008) Deaths from bacterial pneumonia during 1918–19 influenza pandemic. *Emerging Infectious Dis* 14:1193–1199.
3. Donnelly CA, et al. (2003) Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong. *Lancet* 361:1761–1766.
4. Mills CE, Robins JM, Lipsitch M (2004) Transmissibility of 1918 pandemic influenza. *Nature* 432:904–906.
5. Richardson WH (1972) Bayesian-based iterative method of image restoration. *J Opt Soc Am* 62:55–59.
6. Lucy LB (1974) An iterative technique for the rectification of observed distributions. *Astronom J* 79:745–754.
7. Rogers SL (1920) *Special Tables of Mortality from Influenza and Pneumonia, in Indiana, Kansas, and Philadelphia, PA*. (U.S. Dept Commerce, Washington, DC).
8. Eichel OR (1923) *A Special Report on the Mortality from Influenza in New York State During the Epidemic of 1918–19*. (NY State Dept Health, Albany).
9. Wallinga J, Teunis P (2004) Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *Am J Epidemiol* 160:509–516.
10. Wallinga J, Lipsitch M (2007) How generation intervals shape the relationship between growth rates and reproductive numbers. *Proc Biol Sci* 274:599–604.
11. Fraser C (2007) Estimating individual and household reproduction numbers in an emerging epidemic. *PLoS ONE* 2:e758.
12. Anonymous (Sept 27, 1918) Medical men are told of Influenza's spread. *Philadelphia Record*.
13. Barry JM (2004) *The Great Influenza: The Epic Story of the Deadliest Plague in History*. (Viking, New York).
14. Anonymous (Oct 2, 1918) Spread of Influenza grows serious in city. *Philadelphia Record*.
15. Crosby AW (2003) *America's Forgotten Pandemic: The Influenza of 1918*. (Cambridge Univ Press, New York).
16. Frost WH (1920) Statistics of influenza morbidity. With special reference to certain factors in case incidence and case-fatality. *Public Health Rep* 35:584–597.
17. Britten RH (1932) The incidence of epidemic influenza, 1918–19. *Public Health Rep* 47:303–339.
18. Bootsma M, Ferguson N (2007) The effect of public health measures on the 1918 influenza pandemic in U.S. cities. *Proc Natl Acad Sci USA* 104:7588–7593.
19. Hatchett RJ, Mecher CE, Lipsitch M (2007) Public health interventions and epidemic intensity during the 1918 influenza pandemic. *Proc of Nat Acad Sci USA* 104:7582–7587.
20. Clarke SK, Heath RB, Sutton RN, Stuart–Harris CH (1958) Serolgical studies with the Asian strain of influenza A. *Lancet* 1:814–818.
21. Davis LE, Caldwell GG, Lynch RE, Bailey RE, Chin TD (1970) Hong Kong influenza: The epidemiologic features of a high school family study analyzed and compared with a similar study during the 1957 Asian influenza epidemic. *Am J Epidemiol*, 92:240–247.
22. Carrat F, et al. (2008) Time lines of infection and disease in human influenza: A review of volunteer challenge studies. *Am J Epidem* 167:775–785.
23. White LF, Pagano M (2008) Transmissibility of the influenza virus in the 1918 pandemic. *PLoS ONE* 3:e1498.
24. Svensson A (2007) A note on generation times in epidemic models. *Math BioSci*, 208:390–394.
25. Ferguson NM, et al. (2005) Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature* 437:209–214.
26. Dempster AP, Laird NM, Rubin DB (1997) Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc Ser B* 39:1–38.
27. Kenah E, Lipsitch M, Robins JM (2008) Generation interval contraction and epidemic data analysis. *Math BioSci* 213:71–79.
28. Lipsitch M, et al. (2003) Transmission dynamics and control of severe acute respiratory syndrome. *Science* 300:1966–1970.
29. Shepp LA, Vardi Y (1982) Maximum likelihood reconstruction for emission tomography. *IEEE Trans Med Imaging* 113:113–122.

MEDICAL SCIENCES