American Journal of Epidemiology

Vol. 178, No. 9
DOI: 10.1093/aje/kwt133
Advance Access publication:
September 15, 2013

Practice of Epidemiology

A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics

Anne Cori*, Neil M. Ferguson, Christophe Fraser, and Simon Cauchemez

* Correspondence to Dr. Anne Cori, Department of Infectious Disease Epidemiology, MRC Centre for Outbreak Analysis and Modelling, Imperial College London, St Mary's Campus, Norfolk Place, London W2 1PG, United Kingdom (e-mail: a.cori@imperial.ac.uk).

Initially submitted November 26, 2012; accepted for publication May 23, 2013.

The quantification of transmissibility during epidemics is essential to designing and adjusting public health responses. Transmissibility can be measured by the reproduction number *R*, the average number of secondary cases caused by an infected individual. Several methods have been proposed to estimate *R* over the course of an epidemic; however, they are usually difficult to implement for people without a strong background in statistical modeling. Here, we present a ready-to-use tool for estimating *R* from incidence time series, which is implemented in popular software including Microsoft Excel (Microsoft Corporation, Redmond, Washington). This tool produces novel, statistically robust analytical estimates of *R* and incorporates uncertainty in the distribution of the serial interval (the time between the onset of symptoms in a primary case and the onset of symptoms in secondary cases). We applied the method to 5 historical outbreaks; the resulting estimates of *R* are consistent with those presented in the literature. This tool should help epidemiologists quantify temporal changes in the transmission intensity of future epidemics by using surveillance data.

incidence; influenza; measles; reproduction number; SARS; smallpox; software

Abbreviations: CI, credible interval; SARS, severe acute respiratory syndrome.

The reproduction number, *R*, is the average number of secondary cases of disease caused by a single infected individual over his or her infectious period. This statistic, which is time and situation specific, is commonly used to characterize pathogen transmissibility during an epidemic. The monitoring of *R* over time provides feedback on the effectiveness of interventions and on the need to intensify control efforts (1–4), given that the goal of control efforts is to reduce *R* below the threshold value of 1 and as close to 0 as possible, thus bringing an epidemic under control.

A wide range of methods have been proposed to estimate *R* from surveillance data (5–12). However, methods based on fitting mechanistic transmission models to incidence data are often difficult to generalize because of the context-specific assumptions often made (e.g., presence/absence of a latency period or size of the population studied). Recently, a simpler statistical approach was proposed, which addressed this issue. The Wallinga and Teunis method (13) is generic and requires only case incidence data and the distribution of the serial interval (the time between the onset of symptoms in a primary case and the onset of symptoms of secondary cases) to estimate *R* over the course of an epidemic. It is based on the probabilistic reconstruction of transmission trees and on counting the number of secondary cases per infected individual. The method estimates 1 value of *R* per time step of incidence (typically, per day).

However, the approach has several drawbacks. First, estimates are right censored, because the estimate of *R* at time *t* requires incidence data from times later than *t*. Approaches to correct for this issue have been developed (14).

When the data aggregation time step is small (e.g., daily data), estimates of *R* can vary considerably over short time periods, producing substantial negative autocorrelation. Other studies have developed methods to achieve smoother estimates, but the results can be sensitive to the selected time step or to smoothing parameters (15–18).

The implementation of these methods requires time and expertise, especially to produce confidence or credible intervals

for R. Hence, although there are many methods to quantify transmissibility during an epidemic, none currently comes as a ready-to-use tool for nonmodelers.

The aim of our study was to develop a generic and robust tool for estimating the time-varying reproduction number, similar in spirit to earlier methods, but implemented with ready-to-use software and without the drawbacks mentioned above. We provide Microsoft Excel (Microsoft Corporation, Redmond, Washington) and R software (R Foundation for Statistical Computing, Vienna, Austria) versions of this tool, and a user-friendly web interface will soon be available, as well. (The use of the letter R to denote both the reproduction number and the software package is coincidental).

After describing our approach, we apply it to data from selected historical outbreaks of pandemic influenza, severe acute respiratory syndrome (SARS), measles, and smallpox.

MATERIALS AND METHODS

A more detailed description of our methods can be found in the Web Appendices 1–13 available at http://aje.oxfordjournals.org/.

We assume that, once infected, individuals have an infectivity profile given by a probability distribution w_s, dependent on time since infection of the case, s, but independent of calendar time, t. For example, an individual will be most infectious at time s when w_s is the largest. The distribution w_s typically depends on individual biological factors such as pathogen shedding or symptom severity.

The instantaneous reproduction number (19), R_t, can be estimated by the ratio of the number of new infections generated at time step t, I_t, to the total infectiousness of infected individuals at time t, given by $\sum_{s=1}^{t} I_{t-s}w_s$, the sum of infection incidence up to time step $t-1$, weighted by the infectivity function w_s. R_t is the average number of secondary cases that each infected individual would infect if the conditions remained as they were at time t.

In practice, contact rates and transmissibility can change over time, particularly when control measures are initiated. This affects the number of secondary cases that a given individual infected at time step t will actually infect. The case reproduction number at time step t, R_t^c, takes into account those changes. It is the average number of secondary cases that a case infected at time step t will eventually infect (19). It is sometimes called the cohort reproduction number because it counts the average number of secondary transmissions caused by a cohort infected at time step t. However, estimation of R_t^c can be undertaken only in retrospect, once the secondary cases generated by cases infected at t have been infected. R_t^c is the quantity estimated in Wallinga and Teunis–type approaches (although the Wallinga and Teunis method considers cohorts of individuals with symptom onset at time t rather than infection at time t) (13).

The distinction between R_t^c and R_t is similar to the distinction between the actual life span of individuals born in 2013, which we can measure only retrospectively after all individuals have died (i.e., in a century), and life expectancy in 2013, estimated now by assuming that death rates in the future will be similar to those in 2013.

R_t is the only reproduction number easily estimated in real time. Moreover, effective control measures undertaken at time t are expected to result in a sudden decrease in R_t and a smoother decrease in R_t^c (19). Hence, assessing the efficiency of control measures is easier by using estimates of R_t. For these reasons, we focus on estimating the instantaneous reproduction number R_t in this article. (See Wallinga and Teunis (13), and Cauchemez et al. (14, 15) for methods used to estimate the case reproduction number).

Given the definition of R_t stated above, the incidence of cases at time step t is, on average, $\mathbf{E}[I_t] = R_t \sum_{s=1}^{t} I_{t-s}w_s$, where $\mathbf{E}[X]$ denotes the expectation of a random variable X, and I_{t-s} is the incidence at time step $t-s$ (19). Bayesian statistical inference based on this transmission model leads to a simple analytical expression of the posterior distribution of R_t if we assume a gamma prior distribution for R_t. This makes obtaining any desired characteristic of this posterior distribution (e.g., the median, the variance, or the 95% credible interval) straightforward (Web Appendix 1).

However, the resulting R_t estimates can be highly variable and hence difficult to interpret when the time step of data is small (20). We therefore calculate estimates over longer time windows, under the assumption that the instantaneous reproduction number is constant within that time window. At each time step t, we calculate the reproduction number over a time window of size τ ending at time t. These estimates, denoted $R_{t,\tau}$, quantify the average transmissibility over a time window of length τ ending at time t. They are expected to be less variable as the window size τ increases, because 2 successive time windows will then have increasing overlap. As τ increases, the estimates of $R_{t,\tau}$ will also be more precise. In fact, we show in Web Appendix 2 that the precision of these estimates depends directly on the number of incident cases in the time window $[t-\tau+1; t]$. This allows us to control the precision by adjusting the window size.

We also provide estimates of $R_{t,\tau}$ that take into account the uncertainty in the serial interval distribution parameters by integrating over a range of means and standard deviations of the serial interval (Web Appendix 4).

The estimation method presented above is developed for the ideal situation in which times of infection are known and the infectivity profile w_s may be approximated by the distribution of the generation time (i.e., time from the infection of a primary case to infection of the cases he/she generates) (19). However, times of infection are rarely observed, and the generation time distribution is therefore difficult to measure. On the other hand, the timing of onset of symptoms is usually known, and such data collected in closed settings where transmission can reliably be ascertained (e.g., households) can be used to estimate the distribution of the serial interval (time between onset of symptoms of a case and onset of symptoms of his/her secondary cases). Therefore, in practice, we apply our method to data consisting of daily counts of onset of symptoms where the infectivity profile w_s is approximated by the distribution of the serial interval.

For many diseases, including influenza (21), SARS (22), measles (23), and smallpox (24), it is expected that infectiousness starts only around the time of symptom onset. In such diseases, and when the infectiousness profile after symptoms is independent of the incubation period, the distributions

of the serial interval and the generation time are identical (Web Appendix 9), and our estimates are exact (albeit with t defined as the time of symptom onset of a primary case and a time lag in our estimates of R_t equal to the incubation period).

We provide a Microsoft Excel spreadsheet (available at http://tools.epidemiology.net/EpiEstim.xls) that implements the estimation method described above. Documentation on how to use the Microsoft Excel file is provided in Web Appendix 14. We have also developed an R package, EpiEstim, which can be downloaded at http://cran.r-project.org/web/packages/EpiEstim/index.html, in which both our method and the Wallinga and Teunis method (13) are implemented to facilitate comparison. We also developed a user-friendly web interface that will soon be available at http://shiny.epidemiology.net/EpiEstim.

RESULTS

To illustrate the insights that our method can provide, we applied it to 5 historical epidemics that varied in terms of transmissibility, serial interval, and population size. For each epidemic, we retrieved the epidemic curve, as well as the mean and standard deviation of the serial interval, from the literature (Table 1). The discrete distribution of the serial interval, w_s, was then obtained by assuming a gamma distribution (Web Appendix 11). For each day t of each epidemic, we estimated the reproduction number for the weekly window ending on that day ($R_{t,\tau=7}$, now denoted R for simplicity). The 5 epidemic curves, serial interval distributions, and R estimates are presented in Figure 1. Estimates are not shown from the very beginning of each epidemic because precise estimation is not possible in this period (Web Appendix 3). The estimated case reproduction numbers for those 5 epidemics are also shown in Figure 1 for comparison.

Measles in Hagelloch, Germany, 1861

R initially decreased from an initial median value of 4.3 (95% credible interval (CI): 2.0–8.2) in the middle of the third week to 3.0 (95% CI: 1.3–5.9) at the end of the same week, and then increased to 11.5 (95% CI: 8.3–15.3) in the middle of week 4, and finally decreased again until the end of the epidemic, falling below 1 at the beginning of week 7.

The increase in R from weeks 3 to 4 suggests increasing transmissibility. Previous studies have highlighted the importance of the structure (by classroom and household) of the contact network in this epidemic and have suggested the existence of early "superspreaders" (25). These characteristics could explain the increase in R. Interestingly, just after the first peak of incidence, R was still above 1, indicating that the epidemic was not yet over; and indeed, a second peak was still to come.

Pandemic influenza in Baltimore, Maryland, 1918

This epidemic curve was characterized by 2 days with unusually high incidence, on the 1st and the 15th of October 1918 (days 31 and 45). This might be related to a recollection bias, because the data were collected after the epidemic. Although we used a 1-week ($\tau = 7$) time window to calculate R, estimates still fluctuated. We found an initial median estimate of R of 1.4 (95% CI: 1.0–1.9) at the end of week 2. Estimates were then quite stable until R peaked at 2.4 (95% CI: 2.2–2.6) in the middle of week 5 (coincident with the second highest peak in incidence). Estimates then decreased, with R falling below 1 early in week 7, before the largest peak in incidence (though around that peak, R estimates just exceed 1 for a few days). At the very end of the epidemic, the credible intervals widen because of low case numbers.

Table 1. Description of the 5 Data Sets Analyzed, Corresponding to 5 Epidemics Between 1861 and 2009

First Author, Year (Reference No.)	Disease	Location	Year of Epidemic	Incidence of[a]	Mean (SD) Serial Interval, days	Reference for Mean (SD) Serial Interval
Groendyke, 2011 (37)	Measles	Hagelloch, Germany	1861	Onset of early symptoms	14.9 (3.9)	Derived from Groendyke et al. (37)[b]
White, 2008 (17); Fraser, 2011 (20); Frost, 1919 (38); Vynnycky, 2007 (39)	Pandemic influenza	Baltimore, Maryland	1918	Onset of symptoms	2.6 (1.5)	Ferguson et al. (40); Boelle et al. (41)
Fenner, 1988 (26); Gani, 2001 (42)	Smallpox	Kosovo	1972	Onset of symptoms	22.4 (6.1)	Derived from Riley and Ferguson (43)[b]
Cori, 2009 (16)	SARS	Hong Kong	2003	Onset of symptoms	8.4 (3.8)	Lipsitch et al. (44)
Cauchemez, 2011 (31)	Pandemic influenza	School in Pennsylvania	2009	Onset of acute respiratory illness among children attending the school	2.6 (1.5)	Ferguson et al. (40); Boelle et al. (41)

Abbreviations: SARS, severe acute respiratory syndrome; SD, standard deviation.

[a] Clinical characteristic considered to define the incidence. The incidence at time step t is the number of individuals showing this clinical characteristic at time step t.

[b] Estimates of the mean and standard deviation of the generation time for measles and smallpox were not available directly from the literature. Instead, we derived them indirectly from published estimates of the latency and infectious periods. Technical details about this derivation are described in Web Appendix 13.

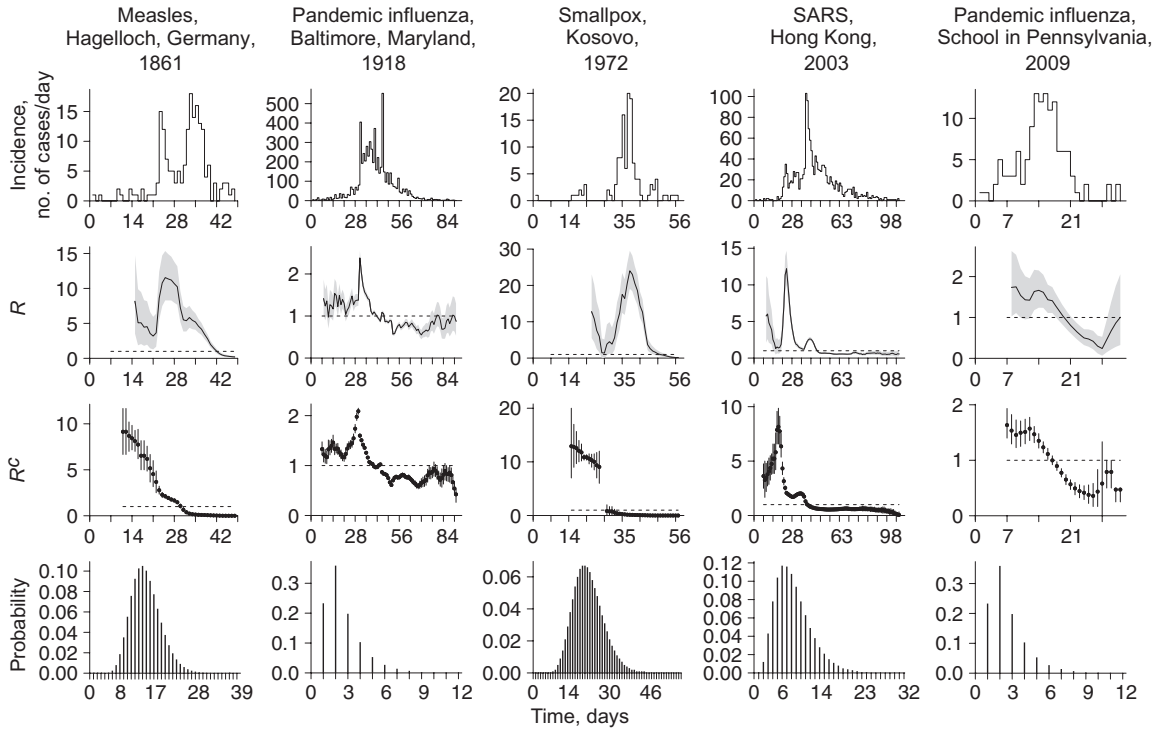

Figure 1. The first row shows daily epidemic curves (from left to right) for measles in Hagelloch, Germany, October 1861–January 1862; pandemic influenza in Baltimore, Maryland, September–November 1918; smallpox in Kosovo, February–April 1972; severe acute respiratory syndrome (SARS) in Hong Kong, February–June 2003; and pandemic influenza in a school in Pennsylvania, April–May 2009. The second row shows daily estimates of the instantaneous reproduction numbers R over sliding weekly windows; the black lines show the posterior medians and the grey zones show the 95% credible intervals; the horizontal dashed lines indicate the threshold value $R = 1$. The third row shows daily estimates of the case reproduction numbers R^c over sliding weekly windows; the black dots show the mean estimates, and the bars show the 95% confidence intervals; the horizontal dashed lines indicate the threshold value $R^c = 1$. The fourth row shows the serial interval distributions used for estimation of R and R^c.

Fraser et al. (20) found a similar temporal trend for R and attributed the decrease in R to social distancing measures that were undertaken around October 10 (day 40). This is consistent with our analysis (given that we were looking at R estimates over the past week) in which R fell below 1 on day 43.

Because the serial interval distribution for the 1918 pandemic is poorly documented, we explored a range of means and standard deviations for the serial interval and derived estimates of R by integrating over all these values (Web Appendix 4). The results are shown in Figure 2. The median estimated R in the presence of this uncertainty differed by 3% or less from the estimates obtained with a fixed serial interval distribution. However, the credible intervals were wider, reflecting the increased uncertainty.

We also examined the choice of the assumed time window width used to estimate R for this data set. Figure 3 shows daily estimates of R for 1-day, 1-week, 2-week, and 4-week windows, assuming a known serial interval distribution (as in Table 1). The estimates varied substantially according to the window size chosen. The 1-day window estimates were so variable that it was hard to derive any trend from them. As the window size grew, the median estimates were smoother, and the credible intervals were narrower, as expected. For 4-week windows, the upper credible interval was below 1 at the end of the epidemic. However, longer intervals delayed the time at which the median estimated R fell below 1. Overall, for this data set, a 1-week window represents a good compromise.

Smallpox in Kosovo, 1972

The analysis of the smallpox outbreak in Kosovo in 1972 illustrates the potentially long delay between the first case and the time when it is reasonable to start estimating R. Here, for a small epidemic with a long mean serial interval, that delay was as long as 4 weeks. We found that R increased from a median value of 3.4 (95% CI: 0.8–9.3) early in the fourth week to 23.9 (95% CI: 19.0–29.5) in the middle of week 6 and then decreased again until the end of the epidemic, falling below 1 only in the beginning of week 8.

The initial increase in R is consistent with a report that identified that transmission during the "second generation of cases" was unusually high, which the authors assumed to be "associated with inadequate protection from vaccination" (26). Estimates stayed above 1 until very late in the epidemic, indicating the limited success of control measures.

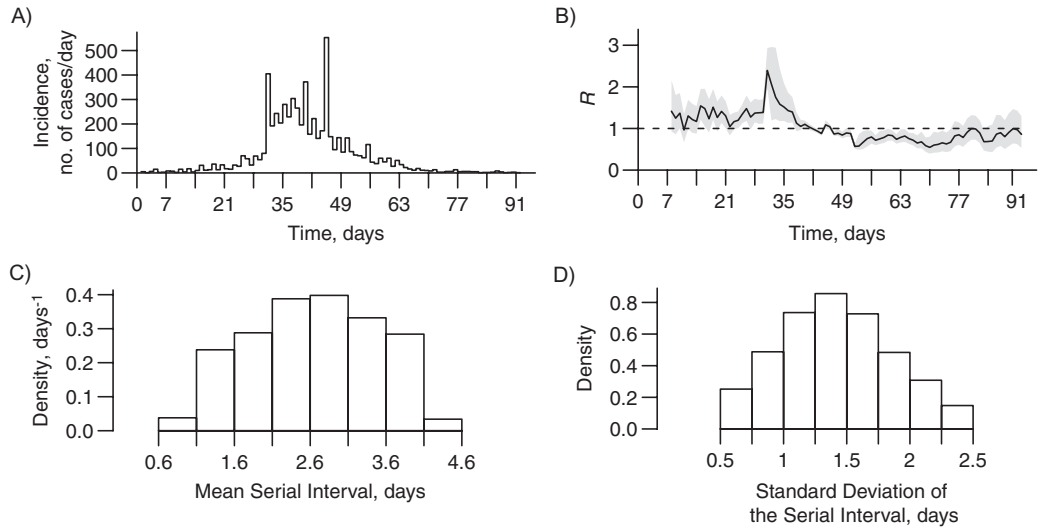

Figure 2. Estimated reproduction number for pandemic influenza in Baltimore, Maryland, September–November 1918. A) Daily epidemic curve; B) daily estimates of the reproduction number *R* over sliding weekly windows (the black line shows the posterior medians and the grey zones show the 95% credible intervals; the horizontal dashed line indicates the threshold value $R = 1$); C) histogram of the mean serial intervals explored; and D) histogram of the standard deviations of the serial interval explored.

Vaccination, which started on March 16 (day 31) was slow (95% coverage was achieved only by the end of April, around day 70) and sometimes ineffective (26).

SARS in Hong Kong, 2003

For the SARS outbreak in Hong Kong in 2003, we find 2 successive peaks in *R*. The first occurred in the middle of week 3 with a median estimate of 12.2 (95% CI: 10.0–14.7), and the second occurred at the end of week 6 with a median estimate of 2.6 (95% CI: 2.4–2.9). *R* then fell below 1 by the end of week 7.

These 2 peaks coincide with the occurrence of known superspreading events, the first occurring in weeks 3 and 4, and the second occurring between weeks 5 and 6 (16, 27, 28). It is notable that *R* falls below 1 very quickly after the epidemic peak, while incidence is still quite high. Similar trends were found in previous analyses of this epidemic (13, 14).

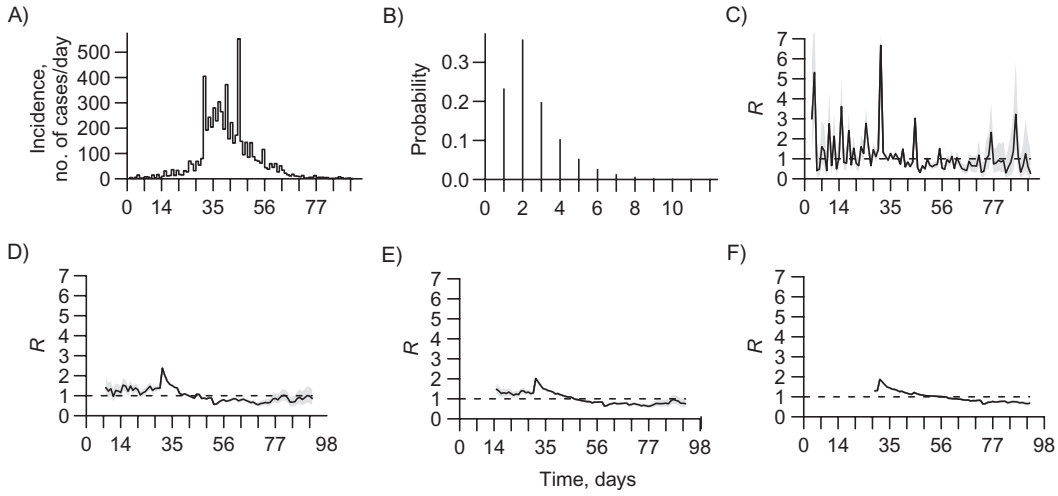

Figure 3. Estimated reproduction number for pandemic influenza in Baltimore, Maryland, September–November 1918, with several time windows. A) Daily epidemic curve; B) serial interval distribution; C–F) daily estimates of the reproduction numbers *R* over 1-day windows; C), over sliding weekly windows; D), over sliding 2-week windows; E) and over sliding 4-week windows; F) black lines show the posterior medians, and grey zones show the 95% credible intervals; the horizontal dashed lines indicate the threshold value $R = 1$.

Pandemic influenza in a school in Pennsylvania, 2009

We estimated that R was relatively constant over the whole second week of the epidemic, with a median around 1.7 (early in the week, 95% CI: 1.0–2.6; late in the week, 95% CI: 1.2–2.2). R then decreased, falling below 1 early in week 4. This could reflect the impact of control measures or could be due to the depletion of susceptibles in the school population. In the last days of the epidemic, 2 new cases appeared, probably as a result of reintroduction of infection from outside the school, resulting in estimates of R increasing again from a minimum value of 0.2 (95% CI: 0.1–0.5) to 0.9 (95% CI: 0.3–2.0).

Comparison between instantaneous reproduction number R and case reproduction number R^c

Figure 1 shows the instantaneous (R) and case (R^c) reproduction numbers estimated for the 5 epidemics. R^c was estimated by using the Wallinga and Teunis method (13), but on weekly windows (Web Appendix 5).

The estimates of R^c on weekly windows are smoother than the estimates of R on weekly windows. Moreover, they are ahead of the estimates of R by a mean serial interval. When the serial interval is short (e.g., for influenza or SARS), this delay is small, and the smoothing effect is not very strong. However, when the serial interval is long (e.g., for measles or smallpox), both effects are more dramatic, and the curves have very different interpretations. For the measles outbreak in Hallegoch, Germany, the highest estimate of R^c is early in the epidemic, on the week ending on day 11, which reflects high transmissibility 2 weeks later (i.e., on the week ending on day 25), coinciding exactly with the peak in the estimated R. This means that the high transmissibility around day 25 is due to cases who have shown symptoms, on average, 2 weeks earlier. Similarly, for the smallpox epidemic in Kosovo, the peak in R^c is on the week ending on day 15, and the peak in R is on the week ending on day 38 (i.e., 23 days later), which is the mean serial interval we have assumed. Transmissibility, measured by the instantaneous reproduction number, was very high during the second generation of cases, around day 38, but this was caused by the first generation of cases, who had symptoms around day 15.

DISCUSSION

We have developed a simple and generic method to estimate time-varying instantaneous reproduction numbers from incidence time series. A simulation study presented in Web Appendix 6 shows that our method is able to detect changes in the reproduction number, for instance, following a control measure. We applied this method to analyze the time course of transmissibility for 5 historical outbreaks. Our estimates of the instantaneous reproduction number are consistent with estimates of the case reproduction number despite considerable differences in interpretation. Our analyses are also in agreement with previously published results obtained with generally more complicated and less general methods.

For instance, although our estimates of the reproduction number for the 1918 influenza pandemic in Baltimore, Maryland, are similar to the maximum likelihood estimates obtained by Fraser et al. (20), it is much easier to produce credible intervals with our method than to produce confidence intervals with the previously used maximum likelihood estimates approach. Our method is also easier to implement and more flexible than the parametric estimation used by White and Pagano (17) on the same data set.

Similarly, for the 2003 SARS epidemic in Hong Kong, Wallinga and Teunis (13), as well as Cauchemez et al. (14), found temporal trends for the case reproduction number similar to our estimates of the instantaneous reproduction number but with lower peak values. The case reproduction number (which is the quantity derived in those studies) is estimated over a generation of infection (i.e., over 8.4 days on average for SARS). When the instantaneous reproduction number is estimated on time windows shorter than the average generation time (which is the case for our weekly windows), we expect the case reproduction to be smoother than the instantaneous reproduction number, which could explain why we find higher peaks. Again, it is more straightforward to produce credible intervals with our method than it is with those approaches.

Robust estimation of R provides important insights into temporal changes in transmission during an epidemic. However, interpreting the temporal trends is not always straightforward. Changes in R can be due to changes in underlying transmissibility (e.g., due to seasonality), changes in contact patterns in the population affected, the impact of control measures, or the depletion of the size of the susceptible population. For instance, we found that R decreased very early during the SARS epidemic in Hong Kong. However, because many control measures were put in place at different times during the SARS epidemic (29, 30), it is difficult to relate the decrease seen directly to a specific control measure from this analysis alone.

Likewise, we estimated that, for the outbreak of 2009 pandemic influenza in a school in Pennsylvania, there was again an early decrease in R. But just estimating R does not allow us to determine whether this reflects a true reduction in transmissibility, possibly due to the school closure between May 14–20 (days 17–23) or the depletion of susceptibles. By using a more complex analysis, Cauchemez et al. (31) showed that the second of these explanations was more likely.

The method developed here relies on knowledge of the serial interval distribution but is able to directly incorporate uncertainty in serial interval distribution estimates**.** Allowing the mean and variance of the serial interval distribution to vary round average values affects median R estimates to a limited extent but increases the credible intervals around those estimates.

The estimates of R obtained with our method are quite sensitive to the size of the sliding window over which the estimates are calculated. Small windows can lead to highly variable estimates with wide credible intervals, whereas longer windows lead to smoothed estimates with narrower credible intervals. In Web Appendix 2, we discuss an interesting result on the minimum number of cases that need to be included in a time window to achieve a given precision in the estimate of R. Because the beginning of an epidemic has few incident cases, we used this result to provide guidance on when it is reasonable to start estimating R.

Finally, our method makes several assumptions that would benefit from reiteration. First, we applied our method to time series of onset of symptoms, and we used the serial interval

distribution as an approximation for the infectivity profile w_s. We showed in Web Appendix 12 that for diseases for which infectiousness starts only at the time of symptom onset, with an infectiousness profile after symptom onset independent of the incubation period, this leads to exact, but time-lagged, estimates of the reproduction number. Although for many diseases, including those studied here (21–24), this assumption is sensible, it does not hold for pathogens such as human immunodeficiency virus, for which infectiousness precedes symptoms (32).

In such cases, and if data on the incubation period (delay between infection and symptom onset) are available, a possible strategy would be to use the incubation period distribution to back-calculate the incidence of infections from the incidence of symptoms and then apply our method to estimate the reproduction number from those inferred data (19). However, this approach may lead to oversmoothed incidence time series compared with the true infection incidence (19).

Second, we assumed that all cases were detected. We showed in a simulation study that this should not dramatically affect estimates as long as the proportion of asymptomatic cases and the reporting rate are constant through time (Web Appendix 8). As an example, an early precursor of the method applied here was used to analyze time series of polio disease incidence, in which approximately 1 in 200 infections was symptomatic (33). In some situations, the reporting rate is likely to change over the course of an epidemic, for instance, as a result of improved case ascertainment or case definitions or changes in health care–seeking behavior over time. If data on reporting are available, it is possible to extend our method to take variable reporting rates into account (6).

We assumed that there were no imported cases, so that each incident case could be attributed to a previous case in the incidence time series. However, if imported cases were identified as such, our method could easily be adapted to account for them, as was done in previous studies by using other estimation methods (12, 34–36).

Finally, we assumed that the serial interval distribution was constant throughout the epidemic. However, if there were independent data (e.g., from contact tracing studies) suggesting evidence of changes in the serial interval distribution, our method could be applied with different serial interval distributions for different time periods of the epidemic.

Despite these assumptions, we feel the simplicity of the method we have presented here outweighs the limitations highlighted above. We hope our method will be adopted by epidemiologists and public health organizations. This will be facilitated by the R package and, more importantly, the simple Microsoft Excel tool and the web interface we have developed and released with this paper. These software programs should allow rapid analysis of incidence time series of any infectious disease within the scope described above and should be valuable tools for future outbreak investigations.

ACKNOWLEDGMENTS

Author affiliation: Department of Infectious Disease Epidemiology, MRC Centre for Outbreak Analysis and Modelling, Imperial College London, London, United Kingdom (Anne Cori, Neil M. Ferguson, Christophe Fraser, Simon Cauchemez).

This work was supported by the United Kingdom Medical Research Council methodology project G0800596, the FP7 European Management Platform for Emerging and Re-emerging Infectious Disease Entities project, and the National Institute of General Medical Sciences Models of Infectious Disease Agent Study program for support. S.C. was supported by the Research Council United Kingdom. C.F. holds a fellowship supported by the Royal Society.

We thank Dr. Thibaut Jombart for his help in developing the R package EpiEstim, and Dr. David Aanensen for his help in setting up the web interface.

Professor Christophe Fraser and Dr. Simon Cauchemez contributed equally to this work.

Conflict of interest: none declared.

REFERENCES

1. Anderson R, May R. *Infectious Diseases of Humans: Dynamics and Control*. Oxford, United Kingdom: Oxford University Press; 1991.
2. Ferguson NM, Cummings DA, Fraser C, et al. Strategies for mitigating an influenza pandemic. *Nature*. 2006;442(7101):448–452.
3. Fraser C, Riley S, Anderson RM, et al. Factors that make an infectious disease outbreak controllable. *Proc Natl Acad Sci U S A*. 2004;101(16):6146–6151.
4. Anderson RM, May RM. Directly transmitted infectious diseases: control by vaccination. *Science*. 1982;215(4536):1053–1060.
5. Riley S, Fraser C, Donnelly CA, et al. Transmission dynamics of the etiological agent of SARS in Hong Kong: impact of public health interventions. *Science*. 2003;300(5627):1961–1966.
6. Fraser C, Donnelly CA, Cauchemez S, et al. Pandemic potential of a strain of influenza A (H1N1): early findings. *Science*. 2009;324(5934):1557–1561.
7. Ferguson NM, Donnelly CA, Anderson RM. Transmission intensity and impact of control policies on the foot and mouth epidemic in Great Britain. *Nature*. 2001;413(6855):542–548.
8. Amundsen EJ, Stigum H, Rottingen JA, et al. Definition and estimation of an actual reproduction number describing past infectious disease transmission: application to HIV epidemics among homosexual men in Denmark, Norway and Sweden. *Epidemiol Infect*. 2004;132(6):1139–1149.
9. Bettencourt LM, Ribeiro RM. Real time Bayesian estimation of the epidemic potential of emerging infectious diseases. *PLoS One*. 2008;3(5):e2185.
10. Cintron-Arias A, Castillo-Chavez C, Bettencourt LM, et al. The estimation of the effective reproductive number from disease outbreak data. *Math Biosci Eng*. 2009;6(2):261–282.
11. Howard SC, Donnelly CA. Estimation of a time-varying force of infection and basic reproduction number with application to an outbreak of classical swine fever. *J Epidemiol Biostat*. 2000;5(3):161–168.
12. Kelly HA, Mercer GN, Fielding JE, et al. Pandemic (H1N1) 2009 influenza community transmission was established in one Australian state when the virus was first identified in North America. *PLoS One*. 2010;5(6):e11341.

13. Wallinga J, Teunis P. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *Am J Epidemiol*. 2004;160(6):509–516.
14. Cauchemez S, Boelle PY, Donnelly CA, et al. Real-time estimates in early detection of SARS. *Emerg Infect Dis*. 2006;12(1):110–113.
15. Cauchemez S, Boelle PY, Thomas G, et al. Estimating in real time the efficacy of measures to control emerging communicable diseases. *Am J Epidemiol*. 2006;164(6): 591–597.
16. Cori A, Boelle PY, Thomas G, et al. Temporal variability and social heterogeneity in disease transmission: the case of SARS in Hong Kong. *PLoS Comput Biol*. 2009;5(8):e1000471.
17. White LF, Pagano M. Transmissibility of the influenza virus in the 1918 pandemic. *PLoS One*. 2008;3(1):e1498.
18. Hens N, Van Ranst M, Aerts M, et al. Estimating the effective reproduction number for pandemic influenza from notification data made publicly available in real time: a multi-country analysis for influenza A/H1N1v 2009. *Vaccine*. 2011;29(5): 896–904.
19. Fraser C. Estimating individual and household reproduction numbers in an emerging epidemic. *PLoS One*. 2007;2(1):e758.
20. Fraser C, Cummings DA, Klinkenberg D, et al. Influenza transmission in households during the 1918 pandemic. *Am J Epidemiol*. 2011;174(5):505–514.
21. Lau LL, Cowling BJ, Fang VJ, et al. Viral shedding and clinical illness in naturally acquired influenza virus infections. *J Infect Dis*. 2010;201(10):1509–1516.
22. Peiris JS, Chu CM, Cheng VC, et al. Clinical progression and viral load in a community outbreak of coronavirus-associated SARS pneumonia: a prospective study. *Lancet*. 2003; 361(9371):1767–1772.
23. Simpson REH. Infectiousness of communicable diseases in the household (measles, chickenpox, and mumps). *Lancet*. 1952;2(6734):549–554.
24. Eichner M, Dietz K. Transmission potential of smallpox: estimates based on detailed data from an outbreak. *Am J Epidemiol*. 2003;158(2):110–117.
25. Groendyke C, Welch D, Hunter DR. *A Network-based Analysis of the 1861 Hagelloch Measles Data*. University Park, PA: Department of Statistics, Pennsylvania State University; 2011.
26. Fenner F, Henderson DA, Arita I, et al. *Smallpox and its Eradication*. Geneva, Switzerland: World Health Organization; 1988.
27. Lee N, Hui D, Wu A, et al. A major outbreak of severe acute respiratory syndrome in Hong Kong. *N Engl J Med*. 2003; 348(20):1986–1994.
28. Leung GM, Hedley AJ, Ho LM, et al. The epidemiology of severe acute respiratory syndrome in the 2003 Hong Kong epidemic: an analysis of all 1755 patients. *Ann Intern Med*. 2004;141(9):662–673.
29. SARS Expert Committee of HKSAR Government. Chronology of the SARS epidemic in Hong Kong. 2003. Hong Kong: SARS Expert Committee. (http://www.sars-expertcom.gov.hk/english/reports/reports/files/e_app3.pdf). (Accessed May 3, 2013).
30. World Health Organization. SARS: chronology of a serial killer. 2003. Geneva, Switzerland: World Health Organization. (http://www.who.int/csr/don/2003_07_04/en/print.html). (Accessed May 3, 2013).
31. Cauchemez S, Bhattarai A, Marchbanks TL, et al. Role of social networks in shaping disease transmission during a community outbreak of 2009 H1N1 pandemic influenza. *Proc Natl Acad Sci U S A*. 2011;108(7):2825–2830.
32. Babiker A, Darby S, De Angelis D, et al. Time from HIV-1 seroconversion to AIDS and death before widespread use of highly-active antiretroviral therapy: a collaborative re-analysis. *Lancet*. 2000;355(9210):1131–1137.
33. Grassly NC, Fraser C, Wenger J, et al. New strategies for the elimination of polio from India. *Science*. 2006;314(5802): 1150–1153.
34. Cowling BJ, Lau MS, Ho LM, et al. The effective reproduction number of pandemic influenza: prospective estimation. *Epidemiology*. 2010;21(6):842–846.
35. Nishiura H, Roberts MG. "Estimation of the reproduction number for 2009 pandemic influenza A(H1N1) in the presence of imported cases" [letter]. *Eurosurveillance*. 2010;15(29): 19–20.
36. Paine S, Mercer GN, Kelly PM, et al. Transmissibility of 2009 pandemic influenza A(H1N1) in New Zealand: effective reproduction number and influence of age, ethnicity and importations. *Eurosurveillance*. 2010;15(24): 9–17.
37. Groendyke C, Welch D, Hunter DR. Bayesian inference for contact networks given epidemic data. *Scand J Stat*. 2011; 38(3):600–616.
38. Frost WH, Sydenstricker E. Influenza in Maryland: preliminary statistics of certain localities. *Public Health Rep*. 1919;34(11):491–504.
39. Vynnycky E, Trindall A, Mangtani P. Estimates of the reproduction numbers of Spanish influenza using morbidity data. *Int J Epidemiol*. 2007;36(4):881–889.
40. Ferguson NM, Cummings DA, Cauchemez S, et al. Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature*. 2005;437(7056):209–214.
41. Boelle PY, Ansart S, Cori A, et al. Transmission parameters of the A/H1N1 (2009) influenza virus pandemic: a review. *Influenza Other Respi Viruses*. 2011;5(5):306–316.
42. Gani R, Leach S. Transmission potential of smallpox in contemporary populations. *Nature*. 2001;414(6865): 748–751.
43. Riley S, Ferguson NM. Smallpox transmission and control: spatial dynamics in Great Britain. *Proc Natl Acad Sci U S A*. 2006;103(33):12637–12642.
44. Lipsitch M, Cohen T, Cooper B, et al. Transmission dynamics and control of severe acute respiratory syndrome. *Science*. 2003;300(5627):1966–1970.