

Transmission parameters and real-time estimation methods

Simon Cauchemez¹², Fabrice Carrat¹²³, Alain-Jacques Valleron¹²³,
Neil Ferguson⁴ and Pierre-Yves Boëlle¹²³

¹ INSERM U707, Paris, France

² Université Pierre et Marie Curie, Paris, France

³ AP-HP, Paris, France

⁴ Imperial College, Department of Infectious Disease Epidemiology, London, Great Britain

Up to now, there is no clear evidence of transmission of the H5N1 influenza virus and the upper admissible value for the reproduction number R (the mean number of cases generated by a single infectious) may be presently set to 0.1 (1), far from the threshold value required for a pandemic to occur ($R=1$). The value that R could attain in case of a human pandemic strain is the object of much speculation. This ignorance also concerns the natural history of the H5N1 pandemic strain infection.

In the present pre-pandemic context, most models designed for planning and presented in this book assume that the H5N1 pandemic flu will have roughly the same transmission characteristics as past seasonal or pandemic flu. Therefore, sound pandemic planning requires consistent estimates of current flu transmission parameters.

Of course, if the H5N1 pandemic was to start, more information would become available on the disease and it would be critical to estimate transmission of this new strain as soon as possible, so that models outputs may be updated adequately.

Here, we show how transmission parameters can be derived from household transmission data, or epidemic curves plus subsets of traced cases. We discuss the critical issue of real-time estimation and describe a statistical framework that can be used for a large set of emerging diseases.

Characterizing flu transmission

Transmission probabilities

In most models, the person-to-person transmission probability must be specified. This probability would for example depend on social or geographic distance in the pair (probability that an infectious transmits the virus to a member of his/her household, school, neighbourhood...).

It is generally difficult to estimate these probabilities because observations of the chain of transmission _ who is infected by whom _ are usually incomplete. Statistical modeling is therefore needed to derive these probabilities.

Household studies provide a convenient setting to derive these parameters. In this context, Longini and al. (2, 3) have shown that it is possible to estimate household and community probabilities of transmission from the distribution of the final number of cases in households at the end of the epidemic.

Natural history of the disease

Detailed description of the natural history of the disease is also required. Depending on the disease, this includes: the incubation period (delay between infection and beginning of symptoms); the latent period (delay between infection and beginning of infectiousness); the infectious period (for which the case may transmit the virus).

The incubation period may be derived from challenges on volunteers (4-7) or observation of epidemics where times of infection are known (8). Derivation of latent and infectious periods is more problematic. Indeed, infectiousness is closely related to viral load _ a subject that does not shed the virus can not be infectious _; but the exact relationship between viral load and transmission risk is poorly known. If infectiousness is considered dichotomous (a subject is “infectious” or not), the viral load cut off above which the case is infectious is unknown. Furthermore, if transmission risk is proportional to viral load, the scale that is used (log-scale or not) may have important impact on infectiousness profile. Eventually, viral load is only one among many factors (symptoms, age, social behaviour...) that should be taken into account to determine infectiousness.

Statistical modeling provides a way to integrate all these factors and confront them with field data. Infectiousness profiles _ risk of transmission according to time since infection _ may for example be derived.

Analysis of symptoms follow-up in households

In two previous works (9, 10), we analyzed a follow-up of influenza symptoms in 334 households during 15 days after an index case visited a general practitioner with virologically confirmed influenza (11).

The main difficulty of these analyses was that the transmission process was only very partially observed: we observed the time when symptoms started for a case; but, we did not know where, by whom and when the case had been infected. Data augmentation has been used to deal with this missing data problem. All information needed to compute easily the likelihood (probability of the observation given the transmission parameters), like times of infection or beginning/end of infectiousness, were made up. Then a systematic exploration of the augmented data was carried out to average out their influence.

Key durations may not be all derived from such analyses, which therefore partly rest on biological observations (4-8). For example, incubation period must be specified to link infectiousness to observation on symptoms. Besides, derivation of both latent and infectious period is difficult from such data; in practice, we made an assumption on the latent period and estimated the corresponding infectiousness profile.

In the two analyses, different assumptions were made for the incubation period, the latent period and the representation of infectiousness, as summarized in table 1. Importantly, the results remained remarkably robust to these various assumptions. For example, Figure 1 presents infectiousness profiles (panel a) and probabilities to escape transmission in a household of size 3 with 1 infectious (panel b) derived from the 2 analyses. In the first analysis (9), in the absence of latent period, infectiousness starts at infection and decreases with time. In the second analysis (10), the absence of transmission during the day following infection is balanced by an increase of infectiousness during the second day, with a peak of infectiousness observed 2 days after infection. The average transmission interval (12) _ mean time lag between infection in the primary case and infection in the secondary cases _ is nevertheless unaffected by these different profiles : it is equal to 2.4 days in the first approach and to 2.6 days in the second.. Eventually, after 2 days, both characterisations of transmission exhibit very similar patterns (see panel b).

Real-time estimation of the reproduction number

Real-time estimation of the reproduction number R is an important part of real-time modeling (13). It would also allow to evaluate in real-time the efficacy of control measures and to determine whether they should be reinforced (if $R > 1$). Such task is however burdened by:

- The difficulty to obtain detailed data in real-time: at best, we can expect availability of the epidemic curve, plus a small subset of traced cases, nearly in real time;
- A poor knowledge of the natural history of the diseases, especially if the disease is emergent;
- The yet unrecorded secondary cases with long incubation periods (censorship);
- The difficulty to analyze the data as they arrive: the method that consists in fitting a mathematical transmission model to the data is disease-dependant and may require time delay to be implemented; generic estimation methods, available for a large set of diseases, are expected.

Here, we present tools for real-time estimation of R from the epidemic curve. For the sake of clarity, we consider the simple example of an epidemic with 3 cases (#1, #2 and #3). Denoting R_1 the number of cases infected by case #1 (in this example, R_1 is equal to 0, 1 or 2), we show how R_1 can be estimated from the observation of times t_1, t_2, t_3 of symptoms onset.

Retrospective estimation when the serial interval is known

The (clinical onset) serial interval (12) is the time lag from symptoms onset in primary case to symptoms onset in secondary cases. Wallinga and Teunis (14) have shown that, if the serial interval is known, the sequence of symptoms onsets t_1, t_2, t_3 can be used to calculate the probability that case #1 has infected case #2 and that he/she has infected case #3. The expected value of R_1 is then simply the sum of these probabilities.

Real-time estimation when the serial interval is known

Let now assume that the epidemic is ongoing, observed up to time T , and that cases #1, #2 and #3 are the first cases of the epidemic. Assume that a fourth case (#4) has already been infected by case #1, but has not yet been detected because his incubation period was long. Here, the approach of Wallinga and Teunis would underestimate R_1 , because it would dismiss future cases, like #4, that may already have been infected by case #1. In (15), we proposed a correction for censoring.

Real-time estimation from the epidemic curve + subset of traced cases

In practice, the serial interval may be derived from a subset of traced cases for which we observe i) the time of symptoms onset of the case and ii) the time of symptoms onset of his/her primary case. At the beginning of an emerging disease outbreak, the subset is empty or very small, so that uncertainty on serial interval is large. While the epidemic goes on, the subset of traced cases increases, and leads to more precise estimates. In an ongoing work, we design a statistical framework that makes it possible to take this uncertainty into account. We use the case tracing dataset to derive of the serial interval. This information is used to make inference on complete case tracing up to present time T from the epidemic curve. Eventually, we correct for censorship at time T .

Monitoring of a SARS-like outbreak

To evaluate the performance of the method to monitor an epidemic, a SARS-like outbreak was simulated. In the simulation, the serial interval had mean 8.4 days and standard deviation 3.8 days (16); the reproduction number had mean 3 before control measures and 0.7 after (14); 5% of cases were traced.

Figure 2 shows the simulated epidemic curve (panel a) and the reproduction number estimated for the last 10 days of observation _ at time t , estimate of R for time period $[t-10, t]$

with data available up to time t_* (panel b). During the first 2 weeks of the outbreak, it is not possible to make inference on the reproduction number because too few cases are traced; after what the method gives accurate estimates of the reproduction number.

Conclusion

Forecasting the eventual impact of epidemics relies strongly on dynamical models for the spread of diseases. To reinforce the credibility of these predictions, it is desirable to put greater detail in the model, but it is also essential that key parameters are given values reflecting reality. We have shown that with modern statistical methods, this may be achieved by the analysis of past epidemics, even if these have been only partially observed. Furthermore, the statistical techniques summarized here allow adjusting the estimates through real-time monitoring of the parameters of an ongoing epidemic. This set of promising techniques will be a strong asset for helping public health authorities in case of an emerging communicable disease outbreak.

References

1. Gay N. Inter-pandemic surveillance, international spread & impact of control measures. GHSAG Flu Modeling Workshop. London, 2005.
2. Longini IM, Jr., Koopman JS, Monto AS, Fox JP. Estimating household and community transmission parameters for influenza. *Am J Epidemiol* 1982;115:736-51.
3. Longini IM, Jr., Koopman JS. Household and community transmission parameters from final distributions of infections in households. *Biometrics* 1982;38:115-26.
4. Fritz RS, Hayden FG, Calfee DP, Cass LM, Peng AW, Alvord WG, et al. Nasal cytokine and chemokine responses in experimental influenza A virus infection: results of a placebo-controlled trial of intravenous zanamivir treatment. *J Infect Dis* 1999;180:586-93.
5. Reuman PD, Bernstein DI, Keefer MC, Young EC, Sherwood JR, Schiff GM. Efficacy and safety of low dosage amantadine hydrochloride as prophylaxis for influenza A. *Antiviral Res* 1989;11:27-40.
6. Hayden FG, Treanor JJ, Fritz RS, Lobo M, Betts RF, Miller M, et al. Use of the oral neuraminidase inhibitor oseltamivir in experimental human influenza: randomized controlled trials for prevention and treatment. *Jama* 1999;282:1240-6.
7. Hayden FG, Treanor JJ, Betts RF, Lobo M, Esinhart JD, Hussey EK. Safety and efficacy of the neuraminidase inhibitor GG167 in experimental human influenza. *Jama* 1996;275:295-9.
8. Moser MR, Bender TR, Margolis HS, Noble GR, Kendal AP, Ritter DG. An outbreak of influenza aboard a commercial airliner. *Am J Epidemiol* 1979;110:1-6.
9. Cauchemez S, Carrat F, Viboud C, Valleron AJ, Boelle PY. A Bayesian MCMC approach to study transmission of influenza: application to household longitudinal data. *Statist Med* 2004;23:3469-87.
10. Ferguson NM, Cummings DAT, Cauchemez S, Fraser C, Riley S, Meeyai A, et al. Strategies for containing an emerging influenza pandemic in Southeast Asia. 2005;advanced online publication.
11. Carrat F, Sahler C, Rogez S, Leruez-Ville M, Freymuth F, Le Gales C, et al. Influenza burden of illness: estimates from a national prospective survey of household contacts in France. *Arch Intern Med* 2002;162:1842-8.
12. Fine PE. The interval between successive cases of an infectious disease. *Am J Epidemiol* 2003;158:1039-47.

13. Gani R. Real Time Modelling of Pandemic Influenza. GHSAG Flu Modeling Workshop. London, 2005.
14. Wallinga J, Teunis P. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *Am J Epidemiol* 2004;160:509-16.
15. Cauchemez S, Boëlle P-Y, Donnelly CA, Ferguson NM, Thomas G, Leung GM, et al. Real-Time Estimation of the Reproduction Number for Early Detection of Control Measures Efficacy: The Example of SARS. *Emerg Infect Dis*;in press.
16. Lipsitch M, Cohen T, Cooper B, Robins JM, Ma S, James L, et al. Transmission dynamics and control of severe acute respiratory syndrome.[see comment]. *Science* 2003;300:1966-70.

Table 1: Assumptions made in Stat Med (9) and Nature (10) models for the incubation period, the latent period and the representation of infectiousness.

Assumptions	Stat Med (9)	Nature (10)
Incubation period	Relatively vague - uniformly drawn in 1-3 days; based on (4-7).	Precise - with mean 1.5 days and standard deviation 0.5 days; based on (8).
Latent period	None - based on challenges on volunteers (4-7) in which influenza virus is detected shortly after virus inoculation (24h).	Equal to incubation period - based on the observation that start of symptoms is coincident with a sharp increase in viral shedding (6).
Representation of infectiousness	Dichotomous - a case is “infectious” or “not infectious”, with a variable duration of infectiousness.	Continuous - the risk of transmission is a continuous function of time since infectiousness has started.

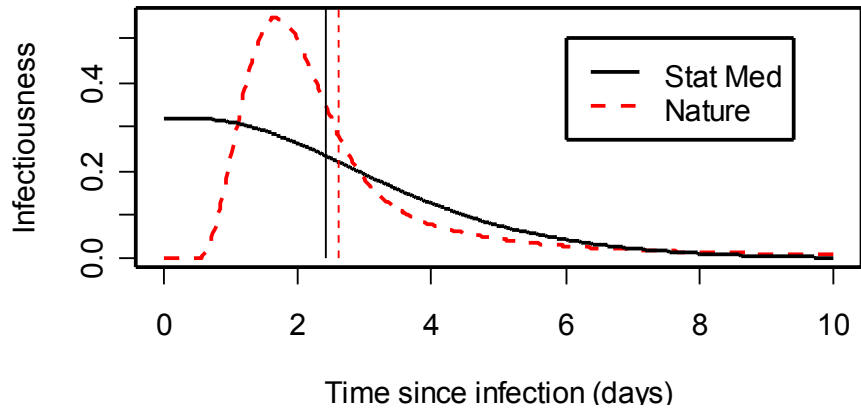
Figure Captions

Figure 1: Flu characterizations of transmission according to time since infection, derived from household transmission data. Two different analyses, published respectively in Stat Med (9) and Nature (10), are compared. Panel a: infectiousness profile; vertical lines indicate estimates of average transmission interval _ mean time lag between infection in the primary case and infection in the secondary cases _ . Panel b: probabilities to escape transmission in a household of size 3 with 1 infectious.

Figure 2: Real-time estimation of the reproduction number R for a SARS-like simulated outbreak. Panel a: daily number of symptoms onsets in the simulated data; in the simulation, the reproduction number was 3 before day 20, and 0.7 after; case tracing was available for 5% of the cases. Panel b: Expectation (solid line) and 95% credible interval (dotted line) of the reproduction number R calculated for the last 10 days of observation, according to the time of last observation; simulation value is also indicated (dashed line). Estimates of the reproduction number are only available once enough cases have been traced, i.e. after 14 days.

Figure 1

(a)



(b)

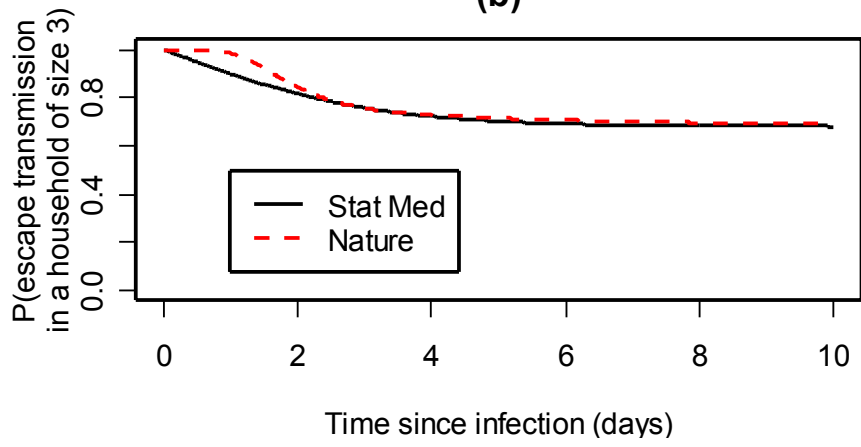


Figure 2

