Travel Medicine and Infectious Disease xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Travel Medicine and Infectious Disease

journal homepage: www.elsevier.com/locate/tmaid

Approximate Bayesian algorithm to estimate the basic reproduction number in an influenza pandemic using arrival times of imported cases

Ka Chun Chong[a,b], Benny Chung Ying Zee[a,b], Maggie Haitian Wang[a,b,*]

[a] Division of Biostatistics, JC School of Public Health and Primary Care, The Chinese University of Hong Kong, Hong Kong, China
[b] Clinical Trials and Biostatistics Laboratory, Shenzhen Research Institute, The Chinese University of Hong Kong, Hong Kong, China

ARTICLE INFO

Keywords:
Basic reproduction number
Epidemic models
Travel data
SIR model
Influenza pandemic
Surveillance

ABSTRACT

Background: In an influenza pandemic, arrival times of cases are a proxy of the epidemic size and disease transmissibility. Because of intense surveillance of travelers from infected countries, detection is more rapid and complete than on local surveillance. Travel information can provide a more reliable estimation of transmission parameters.

Method: We developed an Approximate Bayesian Computation algorithm to estimate the basic reproduction number (R_0) in addition to the reporting rate and unobserved epidemic start time, utilizing travel, and routine surveillance data in an influenza pandemic. A simulation was conducted to assess the sampling uncertainty. The estimation approach was further applied to the 2009 influenza A/H1N1 pandemic in Mexico as a case study.

Results: In the simulations, we showed that the estimation approach was valid and reliable in different simulation settings. We also found estimates of R_0 and the reporting rate to be 1.37 (95% Credible Interval [CI]: 1.26–1.42) and 4.9% (95% CI: 0.1%–18%), respectively, in the 2009 influenza pandemic in Mexico, which were robust to variations in the fixed parameters. The estimated R_0 was consistent with that in the literature.

Conclusions: This method is useful for officials to obtain reliable estimates of disease transmissibility for strategic planning. We suggest that improvements to the flow of reporting for confirmed cases among patients arriving at different countries are required.

1. Introduction

Basic reproduction number (R_0) is an epidemiological metric to measure the number of secondary infections generated on average by an infected patient in a whole susceptible population. It is useful in summarizing the transmissibility of an infectious disease in a population. If $R_0 > 1$, an infection will persist in a population and become endemic because each infected person is expected to have more than one transmission. In contrast, if $R_0 < 1$, the disease transmission cannot be sustained. An underestimation of R_0 could lead to unpreparedness among officials on disease mitigation.

The estimated R_0 can be fitted through feeding syndromic, serological data and laboratory-confirmed counts into simple statistical models (e.g., exponential growth curve) or traditional Susceptible-Infectious-Recovered (SIR) models [1–4]. Nevertheless, common estimation approaches using such data required an assumption of no underreporting, although several approaches were developed to adjust for this problem [5]. Recently, syndromic data were commonly used for influenza prediction and R_0 estimation. For example, Ginsberg et al. demonstrated that Google search queries could track weekly influenza activity [6]. Conversely, serological data could be used to infer influenza transmissibility, although in this case, the time for diagnostic confirmation is longer [7,8]. Compared with other kinds of data, serological data could be used to infer asymptomatic infections without being affected by under-reporting.

To overcome the underreporting problem in R_0 estimation using surveillance data, travel data from the exported cases can provide additional information. The arrival times of infected cases from the originating country are a metric of the expansion of the epidemic and the interaction thereof with international transportation networks [9–11]. In addition to using routine surveillance data, the arrival times of exported cases could help reduce the errors incurred by undetected local cases. Compared with serological data, this information is usually more readily accessible.

In this study, we developed an approach to estimate R_0 and the reporting rate for a new influenza pandemic using an Approximate Bayesian Computation (ABC). The ABC algorithm adopted the use of routine surveillance data as well as information on exported cases, i.e.,

* Corresponding author. Division of Biostatistics, JC School of Public Health and Primary Care, The Chinese University of Hong Kong, Hong Kong, China.
E-mail addresses: marc@cuhk.edu.hk (K.C. Chong), bzee@cuhk.edu.hk (B.C.Y. Zee), maggiew@cuhk.edu.hk (M.H. Wang).

https://doi.org/10.1016/j.tmaid.2018.04.004
Received 27 July 2015; Received in revised form 6 April 2018; Accepted 9 April 2018

the arrival time of the first introduction of an infection to different countries, originating from the source country. We applied this method to the influenza A/H1N1 pandemic in Mexico in mid-March 2009.

2. Materials and methods

2.1. Generation of an epidemic

To fit the influenza epidemic, we used a simple SIR model to describe the transmission dynamics of the infectious disease [1]. In this model, a population is partitioned into three different sub-populations ('compartments'): susceptible ($S(t)$), infectious ($I(t)$), and recovered ($R(t)$) in each of the time points, t ($t = 0, 1, 2, 3 \ldots$). Individuals in the susceptible compartment ($S(t)$) can be infected with a disease at a specific transmission rate (β) and $h_t = (\beta/N)SI$ is the number of new infections per unit time. Infected individuals will become part of the infectious compartment ($I(t)$) and recover at rate γ. The infectious duration is equal to $1/\gamma$, given an exponential assumption. By assuming the pandemic is started with a wholly susceptible population, R_0 is equal to β/γ [2]. Based on some numeric calculations (Appendix A), we can approximate the prevalence of infections $I(t)$ in the population as

$$I(t) = I(0)\exp[(\beta - \gamma)t] = \exp[(R_0 - 1)\gamma t] \quad (1)$$

2.2. Generation of cases seeded by originating country

Assuming that the visitors have the same exposure risk to the disease as the local residents, we can determine the number of infected visitors generated to country i as

$$\text{Number of susceptible visitors in the country} \times \text{Force of infection} = m_i(t) \times d \times \frac{\beta I(t)}{N} \quad (2)$$

where $m_i(t)$ is the number of visitors from country i to the country with the pandemic outbreak (originating country) at day t ($i = 1, \ldots, n$). d is the mean duration of visitor stay in the outbreak country, so the product of $m_i(t)$ and d is the total number of visitors staying the originating country. By adapting the approximation of the prevalence of infections ($I(t)$), the expectation day of the first detection of an imported infection is

$$E(T_i) = \sum_{t=0,1,2,\ldots} (1 - p_{i,t})^t pt \quad (3)$$

where $p_{i,t}$ is the time-varying probability of detecting at least one infected visitor in country i (Appendix B). Given the values of the different parameters, the expected arrival day of the imported cases seeded by originating country can be generated and fitted against the observed values. Fig. 1 shows a simple expression of the generation of exported cases from the infected country.

2.3. Estimation algorithm

Consistent with the literature, solely fitting surveillance data to an SIR model for R_0 estimation may subject to the problem of underreporting. In this study, we employed an ABC algorithm to estimate multiple parameters [12]. The ABC algorithm allows for a Bayesian inference on drawing posterior distributions for the parameters (Θ):

$$f(\Theta|data) \propto f(data|\Theta)\pi(\Theta) \quad (4)$$

where $f(data|\Theta)$ is the data model and $\pi(\Theta)$ is the prior distribution of the parameters. By using typical Monte Carlo simulation methods, the posterior distribution can be iteratively generated from the random draws of the data model and the prior distributions. However, these methods would produce high autocorrelations in the chains of estimates, especially in SIR-type models, and ABC can be used to avoid these problems [34]. In the ABC algorithm, we assumed that the epidemic started at an unobserved time, with interval t^{unobs}, because of underreporting (a constant reporting rate [r] is assumed), so the observed incidence by time (h'_t) is $rh_{t+t^{unobs}}$, which is adjusted from the SIR model. Although the actual epidemic cannot be unobserved, a growing outbreak would undoubtedly affect the visitors, and thus export the cases to other countries (Fig. 1). On the basis of this assumption, we utilized the surveillance data (i.e. number of reported local cases by time) and travel data (i.e. daily travelling rate from country i to the country with disease outbreak and the arrival date of the first imported case bringing the disease to country i) in the estimation algorithm.

In the ABC algorithm, we set the parameter space Θ as t^{unobs}, R_0, r and started by randomly drawing the parameter values from the prior distributions (i.e. i.e. $R_0 \sim Uniform(1, 3)$, $r \sim Uniform(0.01\%, 50\%)$, and $t^{unobs} \sim Uniform(0, 12$ weeks before the observed epidemic start time)). Given the daily travelling rates ($m_i(t)$) and fixing N, φ, d, and γ, the data of the disease incidence and expected arrival day of the imported cases can be generated through adapting the prior parameter values in SIR model, and equation (3) respectively. The generated data are then compared with the observed data using a distance metric, i.e., d(*observed data, generated data*). A common L_1-norm is employed for the distance metric, i.e., least absolute deviations. The simulated data will only be accepted as draws from the posterior distribution if the distance is less than the tolerance level (ε) in which $\varepsilon > 0$. Because of two different generated data sets, two tolerance levels (ε_1 and ε_2) were respectively applied. The computation algorithm is noted in Appendix C.

2.4. Sampling uncertainty

To assess the validity and reliability of the estimation algorithm, we conducted a simulation exercise for the following three scenario settings:

- Scenario 1: $R_0 = 1.2$, $r = 1\%$, and $t^{unobs} = 6$ weeks before the observed epidemic start time
- Scenario 2: $R_0 = 1.7$, $r = 10\%$, and $t^{unobs} = 2$ weeks before the observed epidemic start time
- Scenario 3: $R_0 = 2.2$, $r = 20\%$, and $t^{unobs} = 3$ days before the observed epidemic start time

We fixed the infectious duration as 3 days, the average duration of stay in the originating country as 3 days, and φ as 30% in the simulation. The population was fixed at 1,000,000 individuals. In each of the iteration of simulations, the daily rates of travel to n countries ($i = 1, \ldots, n$) were randomly drawn from a uniform distribution (200, 2000) and were constant over time, i.e., $m_i(t) = m_i$. On the basis of these values, the arrival days of imported cases were randomly drawn from a geometric distribution using the probability by time in equation (3). The medians and 2.5% and 97.5% percentiles (i.e. 95% credible interval [CI]) of the ABC estimates were obtained over 1000 realizations in each of the scenarios.

2.5. Numeric application

The estimation algorithm was applied to the influenza A/H1N1 pandemic in 2009. The influenza A/H1N1 pandemic originated in Mexico on March 14, 2009. During the start of the pandemic, officials and public lacked the knowledge about this communicable disease and as time passed, an unexpected in increase in influenza infections were successively reported. Within a short period of time, the virus had spread to many other countries worldwide. In June 2009, the World Health Organization (WHO) raised the alert level from epidemic to pandemic. From April 2009 to May 2010, more than 18,000 laboratory-confirmed deaths from H1N1 infection were reported. Although the influenza A/H1N1 pandemic ended, the disease tends to occur seasonally.

In the numeric application, we fixed the population of Mexico (N) as