

Early Estimation of the Reproduction Number in the Presence of Imported Cases: Pandemic Influenza H1N1-2009 in New Zealand

Michael George Roberts^{1*}, Hiroshi Nishiura^{2,3}

1 Centre for Mathematical Biology, Institute of Information and Mathematical Sciences, and New Zealand Institute for Advanced Study, Massey University, Auckland, New Zealand, **2** PRESTO, Japan Science and Technology Agency, Saitama, Japan, **3** Theoretical Epidemiology, University of Utrecht, Utrecht, The Netherlands

Abstract

We analyse data from the early epidemic of H1N1-2009 in New Zealand, and estimate the reproduction number \mathcal{R} . We employ a renewal process which accounts for imported cases, illustrate some technical pitfalls, and propose a novel estimation method to address these pitfalls. Explicitly accounting for the infection-age distribution of imported cases and for the delay in transmission dynamics due to international travel, \mathcal{R} was estimated to be 1.25 (95% confidence interval: 1.07, 1.47). Hence we show that a previous study, which did not account for these factors, overestimated \mathcal{R} . Our approach also permitted us to examine the infection-age at which secondary transmission occurs as a function of calendar time, demonstrating the downward bias during the beginning of the epidemic. These technical issues may compromise the usefulness of a well-known estimator of \mathcal{R} - the inverse of the moment-generating function of the generation time given the intrinsic growth rate. Explicit modelling of the infection-age distribution among imported cases and the examination of the time dependency of the generation time play key roles in avoiding a biased estimate of \mathcal{R} , especially when one only has data covering a short time interval during the early growth phase of the epidemic.

Citation: Roberts MG, Nishiura H (2011) Early Estimation of the Reproduction Number in the Presence of Imported Cases: Pandemic Influenza H1N1-2009 in New Zealand. PLoS ONE 6(5): e17835. doi:10.1371/journal.pone.0017835

Editor: Maciej F. Boni, University of Oxford, Viet Nam

Received: October 15, 2010; **Accepted:** February 16, 2011; **Published:** May 26, 2011

Copyright: © 2011 Roberts, Nishiura. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: MGR wishes to acknowledge funding from the Health Research Council under contract number 10/754 (www.hrc.govt.nz/index.html). HN receives support from the Japan Science and Technology Agency, Saitama, Japan, through the PRESTO programme (www.jst.go.jp/kisoken/presto/index_e.html). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: m.g.roberts@massey.ac.nz

Introduction

Influenza A (H1N1-2009) emerged in 2009 [1]. The ensuing pandemic precipitated an international effort to quantify epidemiological parameters, as a necessary first step to assessing its potential impact [2]. Among epidemiological quantities, the most commonly used determinant of the transmission potential has been the basic reproduction number (\mathcal{R}_0), defined as the expected number of secondary cases arising from a typical primary case throughout its entire course of infection in a fully susceptible population. The value of \mathcal{R}_0 is a primary determinant of the size of an epidemic and the effort required to contain it [3,4]. Given that \mathcal{R}_0 has been theoretically defined for a fully susceptible population, we (in common with other authors, e.g. [5,6]) refer to the reproduction number \mathcal{R} , which we estimate from the initial growth phase of the epidemic [7,8]. Such an estimate can aid public health decision-making in real-time during the course of a pandemic [9,10].

The emergence of H1N1-2009 was first detected in North America in March 2009, and initial estimates of its reproduction number, ranging from 1.4 to 1.6 [6], and from 2.2 to 3.1 [5], were published in May of the same year, and derived from Mexican data. Because the emergence in Mexico was at the same time as the beginning of the winter influenza season for Southern Hemisphere countries, it was important to determine if there was a higher transmission potential under winter conditions. A

preliminary study in New Zealand estimated \mathcal{R} to be in the range 1.8–2.2 [11], determined from the exponential growth rate of locally transmitted cases and the assumption that the generation time was known, with a mean of 2.8 days. Another study in Victoria, Australia, estimated the reproduction number to be in the range 2.1–2.6 [12]. Later analyses of the same datasets from New Zealand and Australia, which distinguished imported cases from local cases, estimated the instantaneous (effective) reproduction number as a function of time, and the highest estimate appeared to be smaller than those published in the earlier studies (1.2–1.9 for New Zealand and 1.4–1.6 for Victoria) [13,14]. Because the H1N1-2009 pandemic in these countries (and all the countries other than Mexico) involved repeated introductions of imported cases, it is essential to explicitly account for this aspect in order to appropriately model the transmission dynamics.

Despite the recognition of the role of imported cases in New Zealand, we have yet to clarify the reasons behind the overestimation of \mathcal{R} in the above-mentioned study [11,15]. The purposes of the present study are to illustrate two technical pitfalls in estimating \mathcal{R} during the early epidemic growth phase, and to offer a novel estimation method for \mathcal{R} in the presence of imported cases. Because one should know the best method of obtaining an unbiased estimate of \mathcal{R} in a similar setting, in order to give appropriate feedback to the public health authorities, we compare different modelling strategies for estimating \mathcal{R} in the presence of imported cases. In the next section, we describe the H1N1-2009

epidemic in New Zealand, and illustrate of the estimation method for \mathcal{R} as used in the earlier study. We then explore the underlying reasons for the overestimation of \mathcal{R} . Although a potentially important source of error is heterogeneous mixing (e.g. age-related heterogeneity and other social contact structures), we will not discuss this. Heterogeneous mixing was important in Japan [16], but there was no strong signature of clustering of cases among children during the containment phase in New Zealand. The proportion of children among local confirmed cases by 22 June was as small as 51.4%, and the mean and median ages of local confirmed cases were 22.4 and 19.0 years, respectively. Rather than age-related heterogeneity, we describe two critical factors, one of which is concerned with an explicit modelling approach to imported cases.

In the next section we describe the epidemic in New Zealand, and reexamine the data for the incidence of infection. We then propose a model for the epidemic, based on a renewal process with immigration. The proposed model is used as the basis for a statistical estimation of \mathcal{R} , and we conclude with some remarks concerning the infection-age distribution.

Methods

H1N1-2009 in New Zealand

The daily incidence of confirmed cases of H1N1-2009 in New Zealand is shown in Figure 1. The first cases were in a group of students who had visited Mexico and returned on April 25 [11]. The infection was declared notifiable shortly afterwards, and cases were recorded in the *EpiSurv* database. The date of incidence in Figure 1 is assumed to be the earliest date provided on the database, which may be either the date of symptom onset, hospitalisation, death or reporting. Since the data do not offer further information, we hereafter regard the earliest recorded date as the date of infection (see Discussion). As was adopted elsewhere [17,18], cases with a history of overseas travel within 10 days preceding the onset of illness tend to be defined as imported cases. Since we examine only early epidemic phase without obvious clustering among locally-acquired cases, we assume that no misclassification has occurred in distinguishing between imported and locally-acquired cases.

The initial outbreak declined to extinction in early May and the major epidemic began in June. Until June 22, all reported cases were subject to confirmatory diagnosis and were consistently recorded. On June 22 the health authorities switched the control policy from a *containment* to a *management* phase. During the latter phase not all cases were confirmed, hence the reporting coverage must have been incomplete. The last confirmed case in New Zealand in 2009 was recorded in the database on December 29. A total of 3210 confirmed cases were recorded. Because we estimate \mathcal{R} from the early epidemic growth phase (when the cases should ideally be recorded consistently over time), we limit our analyses to the containment phase before June 22.

Let $i(t)$ be the incidence (i.e. the number of new cases) at calendar time t . During the early growth phase, each primary case generates on average \mathcal{R} secondary cases. The relative frequency of secondary transmission with respect to the time since infection of a primary case is denoted by $g(\tau)$, which is referred to as the generation time (and τ is referred to as infection-age). The expected number of new cases $\mathbb{E}(i(t))$ in the absence of imported cases is written (e.g. [3,19–21])

$$\mathbb{E}(i(t)) = \mathcal{R} \int_0^\infty i(t-\tau)g(\tau)d\tau. \tag{1}$$

When the incidence grows exponentially with growth rate r , we have $i(t) = i_0 e^{rt}$ where $i_0 = i(0)$, a constant. Replacing $i(t)$ on the both sides of (1):

$$\hat{\mathcal{R}} = \frac{1}{\int_0^\infty \exp(-r\tau)g(\tau)d\tau} = \frac{1}{M(-r)}, \tag{2}$$

where $M(-r)$ is the moment-generating function of the generation time, given the intrinsic growth rate r [8]. Hence \mathcal{R} can be estimated, given an estimate of r and if the generation time distribution $g(\tau)$ is assumed to be known [7,8].

We should not ignore demographic stochasticity during the early growth phase of an outbreak, hence the following pure birth process is useful when estimating r [22,23]. Let $I(t)$ be the cumulative incidence at time t . Then

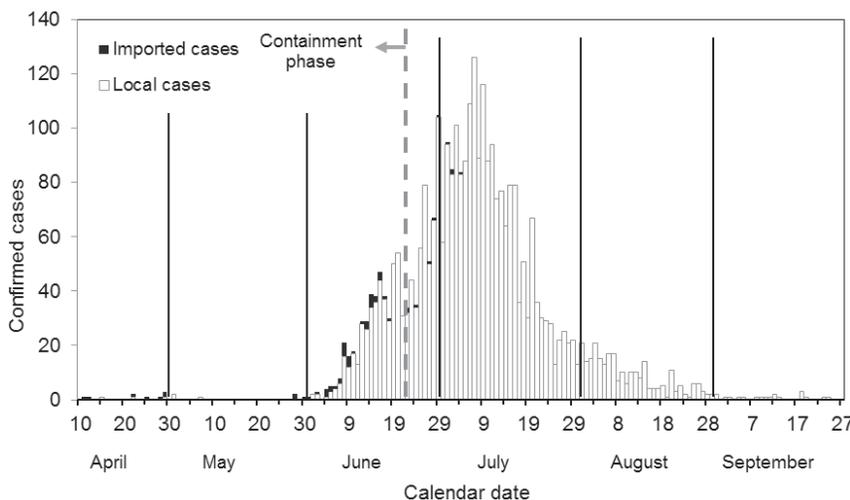


Figure 1. The daily incidence of H1N1-2009 in New Zealand from April to September 2009. Only confirmed cases are shown. White bars represent local cases (i.e. locally transmitted cases without overseas travel), black bars represent imported cases. Vertical solid lines indicate the last calendar date of each month. The vertical dashed line is at June 22, the date on which the control policy switched from a containment to a management phase. doi:10.1371/journal.pone.0017835.g001

$$\begin{aligned} \Pr(I(t+h)=n|I(t)=n) &= 1 - rh + o(h), \\ \Pr(I(t+h)=n+1|I(t)=n) &= rh + o(h), \\ \Pr(I(t+h) > n+1|I(t)=n) &= o(h). \end{aligned} \tag{3}$$

For the analytic solution of equations (3) see [24]. Given our observations of the cumulative number of cases, we have

$$\Pr(I(t)=n+m|I(t-1)=n) = \binom{n+m-1}{n-1} e^{-m}(1-e^{-r})^m \tag{4}$$

which can be used as a conditional likelihood function to estimate r .

The observed and predicted cumulative numbers of local confirmed cases are shown in Figure 2. Although the earliest dates of incidence in Figure 1 have been refined and are different from those analysed in an earlier study [11], the estimated growth rate from 2–13 June is $r=0.258 \text{ day}^{-1}$ (95% confidence interval (CI): 0.219, 0.302), which is consistent with the estimate in [11]. The mean μ and variance σ^2 of the generation time have been estimated from contact tracing in the Netherlands to be 2.70 days and 1.21 days², respectively [25]. Assuming that the generation time follows a gamma distribution, the estimator of \mathcal{R} based on equation (2) is $\hat{\mathcal{R}} = (1+r\sigma^2/\mu)^{\mu^2/\sigma^2}$, leading to $\hat{\mathcal{R}}=1.93$ (95% CI: 1.76, 2.15). This is high compared with published estimates from other countries (e.g. [5,6,12,16,26,27]), and is likely to be an overestimate.

A general renewal process with imported cases

When analysing data for the initial growth of an epidemic it is important to account for imported cases correctly. In equation (2) a stochastic pure birth process was fitted to local cases alone. In discarding imported cases, we correctly removed cases that would otherwise be counted as secondary cases, but at the same time removed some primary cases. This could artificially elevate the estimate of r , and thus \mathcal{R} . In the presence of imported cases with incidence $c(t)$ at time t , the renewal process (1) could read

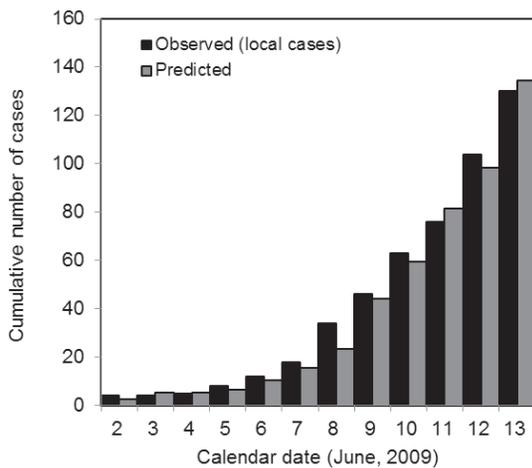


Figure 2. Observed (black) and predicted (grey) cumulative numbers of confirmed locally transmitted cases. Predicted values represent conditional expectations given by $\mathbb{E}(I_t|I_{t-1}) = I_{t-1}e^r$ where I_t is the cumulative number of cases at day t , and $\hat{r}=0.258 \text{ day}^{-1}$ is the maximum likelihood estimate of the growth rate. doi:10.1371/journal.pone.0017835.g002

$$\mathbb{E}(I(t)) = \mathcal{R} \left(\int_0^{\infty} I(t-\tau)g(\tau)d\tau + \int_0^{\infty} c(t-\tau)g(\tau)d\tau \right). \tag{5}$$

Equation (5) is a general form of the age-dependent branching process with immigration [28]. Although equation (5) does not account for different infection-age distribution among imported cases (as compared to local cases), models of this type have been applied to data sets for H1N1-2009 in several published studies [13,14,26] If we have $I(t)=I_0e^{rt}$, the estimator of \mathcal{R} is

$$\hat{\mathcal{R}} = \frac{1}{M(-r) + I_0^{-1}e^{-rI_0} \int_0^{\infty} c(t-\tau)g(\tau)d\tau}. \tag{6}$$

The denominator on the right-hand side includes the imported cases, $c(t)$, and hence a solution requires an approximation to $c(t)$. At the very least, equation (6) highlights that the estimate of \mathcal{R} based on equation (2) results in an overestimate in the presence of imported cases.

Equation (5) requires further modification to capture the underlying dynamics of the epidemic. Before being diagnosed in New Zealand, imported cases were infected overseas, hence there was a time-lag from their infection to their involvement in local transmission. To approximate this, we introduce a constant delay in the involvement of imported cases, i.e.

$$\mathbb{E}(I(t)) = \mathcal{R} \left(\int_0^{\infty} I(t-\tau)g(\tau)d\tau + \int_{\tau_0}^{\infty} j(t,s)g(s)ds \right), \tag{7}$$

where τ_0 represents the time taken from infection to importation (for example the time taken for an international flight). The shortest connecting flight from Mexico City to Auckland is 20 hours 30 minutes, hence we assume that $\tau_0=1$ day. The importance in capturing this delay has been emphasised elsewhere [15]. We have ignored possible transmission during transit. In reality, the infection-age distribution among imported cases is influenced by the transmission dynamics at the origin of importation. However the second integral contains the term $j(t,s)$, the number of new imported cases arriving at time t and infection-age s (instead of $c(t-\tau)$). The times of infection among imported cases are seldom known, hence we postulate an epidemic process at the origin of importation. First, assume that $j(t,s)$ may be expressed by the convolution

$$j(t,s) = \int_{-\infty}^t c(u)q(s-u)du$$

where $c(t)$ corresponds to the observed counts of new imported cases at time t , and $q(s)$ is the density function of the infection-age of imported cases. Second, assume the incidence of infection at the origin of importation to be approximated by an exponential with the same growth rate r as that in New Zealand. That is, we assume that the epidemic is in an early phase at the origin of importation and the growth of cases is sufficiently approximated by deterministic exponential growth. The density function of the infection-age of imported cases, $q(s)$ is then given by

$$q(s) = \frac{\exp(-rs)(1-G(s))}{\int_{\tau_0}^{\infty} \exp(-ru)(1-G(u))du}$$

for $s > \tau_0$ [29] and $q(s)=0$ otherwise. It should be noted that

dynamics other than exponential growth would require us to examine additional data (e.g. epidemic data at the origin of importation), but our assumption permits us to account for the infection-age of imported cases by using local epidemic data only (i.e. the data set in New Zealand). Consequently, the time- and age-dependent number of imported cases is modelled as

$$j(t,s) = \frac{\int_{-\infty}^t c(u)e^{-r(s-u)}(1-G(s-u))du}{\int_{\tau_0}^{\infty} e^{-ru}(1-G(u))du} \quad (8)$$

in equation (7). The growth rate r in the right-hand side of (8) can be replaced by $\mu(\mathcal{R}\sigma^2/\mu^2 - 1)/\sigma^2$ as in the estimator of \mathcal{R} described above.

In summary, we have devised a modelling approach to early epidemic processes with imported cases that accounts for two issues. The first is a constant delay (τ_0) in transmission dynamics involving imported cases, which corresponds to the time in transit. The second is a distributed delay. The imported cases are likely to have been infected shortly before departure, but their infection-age distribution should take account of transmission dynamics at their origin. There are other factors that could add further detail: for example the relative contribution of imported cases to secondary transmission in comparison with local cases (see [15] for a description), and heterogeneous mixing. However, we have insufficient data to account for these. It should be noted that the

time-dependency of imported cases in the proposed model (7) does not lead to an exponential growth of local cases. Simpler age-dependent branching process models with immigration have been examined elsewhere to find the analytical solutions to describe the growth of local cases [30,31].

Statistical estimation of \mathcal{R}

We now estimate \mathcal{R} using the modelling approaches described above. Figure 3A shows the daily incidence of confirmed imported (black) and local (grey) cases from 28 May to 22 June 2009. Since we failed to jointly estimate \mathcal{R} and the generation time distribution (see below), we assume that the generation time distribution $g(\tau)$ is known, with the mean 2.70 days and the variance 1.21 days² [25], but we examine the sensitivity of \mathcal{R} to the mean generation time. While \mathcal{R} may vary by location, over time and according to the level of public health interventions, we generally expect the generation time to be consistent between locations, unless extrinsic measures significantly influence epidemiological patterns of transmission [32]. Since the observed data are provided as daily reports, we discretize the distribution,

$$g_{\tau} \Rightarrow \frac{G(\tau) - G(\tau - 1)}{G(\tau_{\max})}$$

for $\tau > 0$, with $g_0 = 0$, and where τ_{\max} is the longest infection-age causing secondary transmission. In the following we set $\tau_{\max} = 9$,

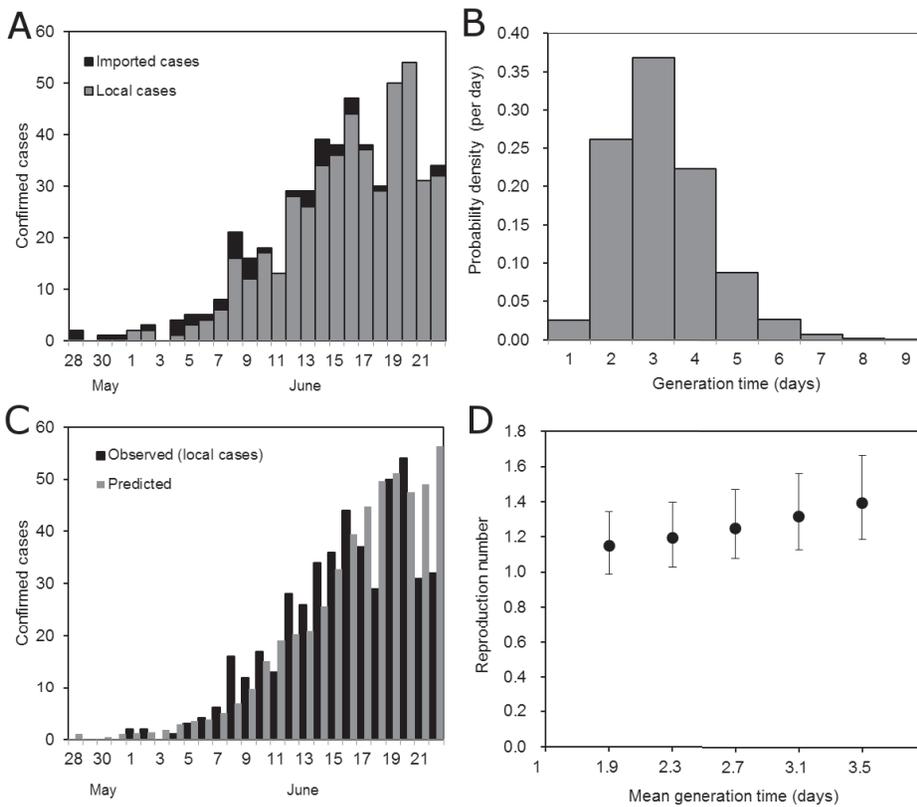


Figure 3. Transmission dynamics of H1N1-2009 in New Zealand. A: Observed daily incidence of imported (black) and local (grey) cases from 28 May to 22 June 2009. We examine only confirmed cases during the containment phase. B: Discretised distribution of the generation time. Mean and variance are assumed to be 2.70 days and 1.21 days², respectively. C: Observed (black) and predicted (grey) numbers of local confirmed cases. Predicted values represent conditional expectations derived from our proposed model, which includes adopting a negative binomial offspring distribution. D: Sensitivity of the estimated reproduction number to the mean generation time, over the range 1.9–3.5 days. Whiskers extend to the upper and lower 95% confidence intervals based on the profile likelihood. doi:10.1371/journal.pone.0017835.g003

because the frequency of secondary transmission after infection-age 9 days is negligible.

We examine three different models to illustrate the impact of underlying assumptions with regard to imported cases on the estimate of \mathcal{R} , and to determine the best modelling strategy. Let I_t and C_t be the incidence of local and imported cases on day t , respectively. We denote the history of both series of cases up to day t by H_t . The first model we examine is the renewal equation with imported cases, but without an adjustment of infection-age distribution and without a delay. Given H_{t-1} , the conditional expected incidence of local cases on day t is

$$\mathbb{E}(I_t|H_{t-1}) = \mathcal{R} \left(\sum_{s=1}^{\infty} I_{t-s}g_s + \sum_{s=1}^{\infty} C_{t-s}g_s \right). \tag{9}$$

The second model accounts for a constant delay in imported cases, but without adjustment for their infection-age distribution, i.e.

$$\mathbb{E}(I_t|H_{t-1}) = \mathcal{R} \left(\sum_{s=1}^{\infty} I_{t-s}g_s + \sum_{s=1+\tau_0}^{\infty} C_{t-s}g_s \right), \tag{10}$$

where τ_0 is the time taken for transit, assumed to be one day. In the third model, we incorporate the adjustment of infection-age distribution and a constant delay in the transmission dynamics:

$$\mathbb{E}(I_t|H_{t-1}) = \mathcal{R} \left(\sum_{s=1}^{\infty} I_{t-s}g_s + \sum_{s=1+\tau_0}^{\infty} \sum_{u=-\infty}^t C_u q_{s-u}g_s \right), \tag{11}$$

where

$$q_{s-u} = \frac{\exp(-r(s-u))(1-G(s-u))}{\int_1^{\infty} \exp(-rw)(1-G(w))dw}$$

and $r = \mu(\mathcal{R}^{\sigma^2/\mu^2} - 1)/\sigma^2$.

We examine two conditional likelihood functions for the estimation of \mathcal{R} . First, if we regard \mathcal{R} as a (deterministic) parameter and ignore individual heterogeneity in the number of secondary transmissions, then the infection process is Poisson [3]. Assuming that the discrete generation time follows a multinomial distribution, a thinned Poisson is obtained [33] which is known to be useful for the joint estimation of \mathcal{R} and g_{τ} [26,33]:

$$L(\mathcal{R}|H_{t-1}) = \prod_{t=1}^T \frac{\exp(-\mathbb{E}(I_t|H_{t-1}))\mathbb{E}(I_t|H_{t-1})^{N_t}}{N_t!}, \tag{12}$$

where T is the last date of observation (equivalent to 22 June 2009) and N_t represents the observed number of local cases on day t . As an alternative, we incorporate a gamma-distributed individual heterogeneity for the infection process, which results in a negative binomial distribution [6,34]:

$$L(\mathcal{R}|H_{t-1}) = \prod_{t=1}^T \frac{\Gamma(N_t+k)}{N_t! \Gamma(k)} \left(\frac{k}{k+\mathbb{E}(I_t|H_{t-1})} \right)^k \left(\frac{\mathbb{E}(I_t|H_{t-1})}{k+\mathbb{E}(I_t|H_{t-1})} \right)^{N_t}. \tag{13}$$

The dispersion parameter k has to be jointly estimated when employing equation (13). The Poisson distribution is obtained as

$k \rightarrow \infty$, and the logarithmic series distribution is obtained as $k \rightarrow 0$. A maximum likelihood estimate of \mathcal{R} (and additionally, k for the negative binomial likelihood distribution when applicable) is obtained by minimizing the negative logarithm of either (12) or (13), and the 95% CI is derived from the profile likelihood. To compare model fit we employ Akaike's Information Criterion, $AIC = -2LL_{\max} + 2m$, where LL_{\max} is the maximum value of the loglikelihood function and m is the number of parameters estimated.

Assessment of the infection-age distribution

In equations (9–11), the right-hand side inside parenthesis (i.e. other than the factor \mathcal{R}) may be interpreted as the expected number of cases who have a potential to cause transmission at time t (We refer to these as primary cases). For example, using the best-fit model (11), the expected number of primary cases is

$$\sum_{s=1}^{\infty} I_{t-s}g_s + \sum_{s=1+\tau_0}^{\infty} \sum_{u=-\infty}^t C_u q_{s-u}g_s.$$

The mean infection-age at which secondary transmission has occurred is

$$\bar{a}_t = \frac{\sum_{s=1}^{\infty} s I_{t-s}g_s + \sum_{s=1+\tau_0}^{\infty} s \sum_{u=-\infty}^t C_u q_{s-u}g_s}{\sum_{s=1}^{\infty} I_{t-s}g_s + \sum_{s=1+\tau_0}^{\infty} \sum_{u=-\infty}^t C_u q_{s-u}g_s}, \tag{14}$$

and the variance is

$$v_t^2 = \frac{\sum_{s=1}^{\infty} (s-\bar{a}_t)^2 I_{t-s}g_s + \sum_{s=1+\tau_0}^{\infty} (s-\bar{a}_t)^2 \sum_{u=-\infty}^t C_u q_{s-u}g_s}{\sum_{s=1}^{\infty} I_{t-s}g_s + \sum_{s=1+\tau_0}^{\infty} \sum_{u=-\infty}^t C_u q_{s-u}g_s}. \tag{15}$$

The time required for the generation time to converge to a stable distribution has attracted the recent attention of epidemic modellers [35], but this has been preceded by discussions in the mathematical demography literature for more than 30 years. The population entropy, proposed by Lloyd Demetrius, is defined by

$$Q = \frac{-\int_0^{\infty} g(\tau)\log g(\tau)d\tau}{\int_0^{\infty} sg(s)ds},$$

and has been shown to measure the rate of convergence of a population to a stable age distribution [36]. Further theoretical accounts of Q , and insights into its interpretation employing a Leslie model, are described in [37].

Results

Estimates of the reproduction number

The maximum likelihood estimates of \mathcal{R} ranged from 1.18 to 1.37, see Table 1. Where different offspring distributions were used with the same type of model, the negative binomial distribution resulted in a better fit than the Poisson distribution (lower AIC). In addition, the negative binomial distribution always led to a greater \mathcal{R} than the Poisson distribution, but with wider uncertainty bounds reflecting its fatter tail. When different models were compared, the model that accounted for the infection-age distribution and a constant delay for imported cases (AD) was the best fit, and resulted in the estimate $\mathcal{R} = 1.25$ (95% CI: 1.07, 1.47). Differences in the estimates of \mathcal{R} were very small between models

Table 1. Comparison of parameter estimates and model fit.

Model ¹	Offspring distribution ²	\mathcal{R} (95% CI) ³	AIC ⁴	Dispersion parameter ⁵
RP	P	1.22 (1.11, 1.33)	168.9	
RP	NB	1.36 (1.13, 1.66)	149.9	10.2 (3.8, 30.1)
RD	P	1.22 (1.11, 1.33)	169.6	
RD	NB	1.37 (1.14, 1.68)	150.1	10.0 (3.7, 29.1)
AD	P	1.18 (1.08, 1.28)	157.2	
AD	NB	1.25 (1.07, 1.47)	144.3	14.4 (5.1, 48.2)

¹RP: renewal process, equation (9); RD: RP plus a constant delay in imported cases, equation (10); AD: RD plus statistical adjustment of infection-age distribution among imported cases, equation (11).

²P: Poisson distribution, equation (12); NB: negative binomial distribution, equation (13).

³ \mathcal{R} : reproduction number, CI: confidence intervals derived from profile likelihood.

⁴AIC: Akaike Information Criterion.

⁵Dispersion parameter of negative binomial distribution.

doi:10.1371/journal.pone.0017835.t001

with and without a constant delay for imported cases (RP & RD). This is because g_1 of the assumed generation time distribution is small, but if the generation time were shorter than assumed, its influence would be greater [15].

In Figure 3C the observed and expected (based on the best-fit model) numbers of locally transmitted confirmed cases are compared as a function of time. In Figure 3D the sensitivity of \mathcal{R} to different mean generation times, ranging from 1.9 days to 4.0 days (assumed maximum), is examined. As in previous studies [7,8], as the mean generation time increases the estimate of \mathcal{R} also increases (ranging from 1.15–1.39). This illustrates the importance of having a reliable estimate of the generation time distribution if one is to obtain a precise estimate of \mathcal{R} .

In addition to the results shown in Table 1, we attempted to jointly estimate \mathcal{R} and the generation time distribution using a Poisson-distributed likelihood function. Employing a model with a one day delay for importation, and limiting the maximum generation time to 4 days, we obtained $g_1=0.448$, $g_2=0.083$, $g_3=0.105$ and $g_4=0.364$. Thus, the mean of the jointly estimated generation time was 2.38 days. Attempts to estimate with a greater maximum generation time did not result in successful convergence. We know of no explanation for the implied bimodal distribution, so regard this as a failure to implement a joint estimation. We discuss this outcome in the next section.

Infection-age at which secondary transmission occurs

In Figure 4A the mean generation time (i.e. the mean infection-age at which secondary transmission occurs) is shown as a function of time, as derived from equation (14). Initially, the mean infection-age of secondary transmission is small, and is shorter than the assumed mean generation time, 2.70 days. As the epidemic progresses, the mean generation time increases and converges to the assumed mean. Despite its convergence to 2.70 days, the mean generation time tends to be short during the first 2 weeks of the epidemic. In Figure 4B the variance of the generation time is shown to fluctuate as a function of time (from equation (15)), before converging to the assumed variance.

The assessment of the time-dependent generation time distribution plays a key role in interpreting the reason behind the overestimation of \mathcal{R} when employing the exponential growth rate r based on a pure birth process (4). It must be noted that the well-known estimator $\mathcal{R}=1/M(-r)$ depends on the assumption that the infection-age distribution is stable. If not stable, the direct application of the estimator could yield a biased estimate of \mathcal{R} . Even provided that r is estimated to be as large as 0.258 day^{-1}

during the initial phase of the epidemic, the estimate reflects transmission that occurs at earlier infection-ages than the mean generation time. The illustrated time-dependency of the generation time distribution also partly explains the failure of the joint estimation of \mathcal{R} and the generation time reported above. Given that the majority of observed transmission events take place at early infection-ages, and given that the variance has also fluctuated, a precise estimate of the generation time distribution is not possible. In fact, the jointly estimated generation time would be shorter than an unbiased estimate of the generation time. A successful joint estimation would require a longer time series of data than we examined. In addition, a recent study has shown that the joint estimation involves several technical difficulties during the early exponential growth phase of an epidemic, especially in the presence of heterogeneous transmission [38].

In Figure 4C the sensitivity of Q to the mean generation time is examined. It is evident that the time taken for convergence is longer when the mean generation time is longer. In addition, it is important to examine the influence of the variance of the generation time on Q (Figure 4D). If the variance were zero (i.e. for a delta function), the infection-age distribution would not converge to a stable distribution. As the variance increases, convergence improves. Since the reporting interval for influenza (i.e. daily data) is similar to the mean generation time, the time taken for convergence is less likely to be a significant problem than it would be for slower diseases (e.g. HIV/AIDS). Nevertheless, this issue cannot be ignored when we estimate \mathcal{R} from a dataset covering a short period of time during the early growth phase. Indeed, the time-dependent infection-age distribution is a plausible explanation for an overestimation using the growth rate r . In addition to the issue of precise estimation of \mathcal{R} from early epidemic growth data, this highlights the critical importance of quantifying the generation time distribution, and especially its variance, if we are to understand the underlying epidemic dynamics.

Discussion

We have estimated the reproduction number \mathcal{R} for H1N1-2009 in New Zealand, by reanalysing the early epidemic growth data. We explored two modelling issues: taking account of imported cases; and the infection-age distribution at which secondary transmission occurs during the early growth phase of the epidemic. We believe these provide at least part of the underlying reasons for a previous overestimate of \mathcal{R} [11]. Explicitly accounting for the

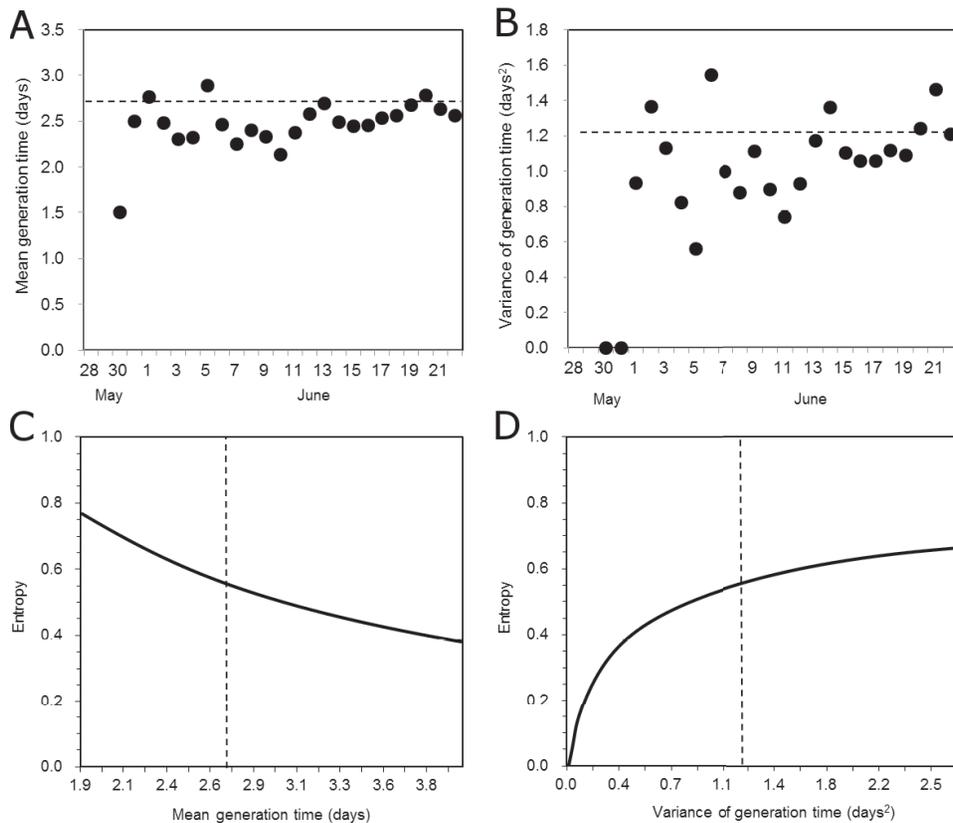


Figure 4. Assessment of the distribution of generation time. A: The mean generation time as a function of calendar time. B: The variance of the generation time as a function of time. C: Sensitivity of population entropy to mean generation time. D: Sensitivity of population entropy to the variance of the generation time. The horizontal dashed line in A, and the vertical dashed line in C, represents the assumed mean generation time, 2.70 days; which is fixed in B and D. The horizontal dashed line in B, and the vertical dashed line in D, represents the assumed variance, 1.21 days²; which is fixed in A and C.

doi:10.1371/journal.pone.0017835.g004

infection-age distribution of imported case, and the delay due to transit, \mathcal{R} was estimated to be 1.25 (95% CI: 1.07, 1.47). Despite wide uncertainty, the upper 95% CI is smaller than the lower 95% CI of the published preliminary estimate [11]. Moreover, our modelling approach permitted us to examine the generation time as a function of calendar time, demonstrating that generation time is biased downwards during the beginning of the epidemic. Both points illustrate important technical pitfalls in the use of the exponential growth rate r and the estimator $\mathcal{R} = 1/M(-r)$ for early growth data. To avoid a biased estimate of \mathcal{R} , we propose investigation of both of these issues, especially when one has to measure \mathcal{R} from data collected over a short period of time during the early growth phase.

We have shown that explicitly accounting for imported cases would be a key factor in avoiding an overestimation of \mathcal{R} . We have also emphasised the importance of addressing the infection-age distribution for imported cases, which will be different to that for local cases. When modelling transmission from imported cases, one should account for the time-lag from infection to importation, and account for the transmission dynamics at the origin of importation. The former can be approximated by a fixed delay, the average time required for international travel. The latter requires an assumption concerning the transmission dynamics at the origin. In addition, the use of a negative binomial offspring distribution was favoured for the three models we examined. Demographic stochasticity during the early growth phase is not negligible, and it appears that the stochastic early epidemic process

in New Zealand was better captured by the negative binomial distribution than the Poisson distribution, indicating the presence of individual heterogeneity in the transmission process.

One implication of the proposed model is that the generation time was yet to converge to a stable distribution in New Zealand at June 22, 2009. In particular, the mean infection-age at which secondary transmission occurred appeared to be short, partially explaining the reason for the overestimation of \mathcal{R} . It must be remembered that the estimator $\mathcal{R} = 1/M(-r)$ is based on the assumption that the generation time is stable, and this is frequently not the case early in the epidemic. One should then employ a renewal process (equations (1) and (11) in the absence and presence of imported cases, respectively) and estimate \mathcal{R} as a parameter. The population entropy Q indicates the rate of convergence to a stable distribution.

Four limitations of this analysis should be noted. First, our estimate of \mathcal{R} is based on the daily incidences of confirmed cases, which are recorded when an infection is classified as *notifiable*. As with any data set there could be issues with classification and interpretation, but these are the best items of information available at the time. In particular, the earliest date recorded has been taken as a proxy for the date of infection for locally transmitted cases. Further in-depth investigation of each case (e.g. taking account of the incidence and reporting delay) could potentially produce a more accurate data set, but our objective is to produce an estimate based on the information to hand. Second, although we recognise the crucial role of the generation time distribution, we have based

our distribution g_i on the published result of a contact tracing exercise [25]. Despite the existing method for estimating g_i in real time [33], we have yet to invent a method for its unbiased estimation [38–40]. Third, we have ignored heterogeneity (other than infection-age) and adopted the homogeneous mixing assumption. As has been discussed elsewhere [6,12,22,27], age-related heterogeneity is likely to provide additional insights into the transmission dynamics, and estimation of the relative contribution of imported cases to secondary transmission (compared with local cases) should be possible through examining additional epidemiological information. Fourth, although unlikely to vary the results of the present study (because the flow of cases can be assumed to be unidirectional from North America to New Zealand), we focused on the mobility of primary cases and did not discuss that of secondary cases. Where emigration would influence the growth estimate of cases (e.g. in Mexico or for a meta-population model), models with bidirectional mobility would be called for.

In conclusion, the early epidemic data in New Zealand did not suggest that the transmission potential of H1N1-2009 was higher

than in Northern Hemisphere countries [6,16,26,27]. The present study has highlighted the importance of modelling the transmission dynamics of imported cases and examining the infection-age distribution of primary cases during the early stage of an epidemic, and we believe that these aspects explain some of the reasons for the overestimation of \mathcal{R} in an earlier study [11]. When it is necessary to obtain an estimate of the transmission potential for a novel emerging disease, we suggest the use of equation (11) with a negative binomial offspring distribution, and equation (14) for the assessment of the mean infection-age of primary cases.

Acknowledgments

The authors are grateful to ESR Ltd for access to the EpiSurve Database.

Author Contributions

Conceived and designed the experiments: MGR HN. Performed the experiments: HN. Analyzed the data: MGR HN. Contributed reagents/materials/analysis tools: MGR HN. Wrote the paper: MGR HN.

References

- Neumann G, Noda T, Kawaoka Y (2009) Emergence and pandemic potential of swine-origin H1N1 influenza virus. *Nature* 459: 931–939.
- van Kerkhove MD, Asikainen T, Becker NG, Borge S, Desenclos JC, et al. (2010) Studies needed to address public health challenges of the 2009 H1N1 influenza pandemic: insights from modeling. *PLoS Medicine* 7: e1000275.
- Diekmann O, Heesterbeek JAP (2000) *Mathematical epidemiology of infectious diseases: model building, analysis and interpretation*. Chichester: Wiley.
- Roberts MG (2007) The pluses and minuses of R_0 . *Journal of the Royal Society Interface* 4: 949–961.
- Boëlle PY, Bernillon P, Desenclos JC (2009) A preliminary estimation of the reproduction ratio for new influenza A(H1N1) from the outbreak in Mexico, March–April 2009. *Eurosurveillance* 14: pii:19205.
- Fraser C, Donnelly CA, Cauchemez S, Hanage WP, van Kerkhove MD, et al. (2009) Pandemic potential of a strain of influenza A (H1N1): Early findings. *Science* 324: 1557–1561.
- Roberts MG, Heesterbeek JAP (2007) Model-consistent estimation of the basic reproduction number from the incidence of an emerging infection. *Journal of Mathematical Biology* 55: 803–816.
- Wallinga J, Lipsitch M (2007) How generation intervals shape the relationship between growth rates and reproductive numbers. *Proceedings of the Royal Society of London Series B* 274: 599–604.
- Ong JBS, Chen MC, Cook AR, Lee HC, Lee VJ, et al. (2010) Real-time epidemic monitoring and forecasting of H1N1-2009 using influenza-like illness from general practice and family doctor clinics in Singapore. *PLoS ONE* 5: e10036.
- Wallinga J, van Boven M, Lipsitch M (2010) Optimizing infectious disease interventions during an emerging epidemic. *Proceedings of the National Academy of Science, USA* 107: 923–928.
- Nishiura H, Wilson N, Baker MG (2009) Estimating the reproduction number of the novel influenza A virus (H1N1) in a southern hemisphere setting: preliminary estimate in New Zealand. *The New Zealand Medical Journal* 122: 73–77.
- McBryde ES, Bergeri I, van Gemert C, Rotty J, Headley EJ, et al. (2009) Early transmission characteristics of influenza A(H1N1)v in Australia: Victorian state. *Eurosurveillance* 14: pii:19363.
- Kelly HA, Mercer GN, Fielding JE, Dowse GK, Glass K, et al. (2010) Pandemic (H1N1) 2009 influenza community transmission was established in one Australian state when the virus was first identified in North America. *PLoS ONE* 5: e11341.
- Paine S, Mercer GN, Kelly PM, Bandaranayake D, Baker MG, et al. (2010) Transmissibility of 2009 pandemic influenza A(H1N1) in New Zealand: effective reproduction number and influence of age, ethnicity and importations. *Eurosurveillance* 15: pii:19591.
- Nishiura H, Roberts MG (2010) Estimation of the reproduction number for 2009 pandemic influenza A(H1N1) in the presence of imported cases. *Eurosurveillance* 15: pii:19622.
- Nishiura H, Castillo-Chavez C, Safan M, Chowell G (2009) Transmission potential of the new influenza A(H1N1) virus and its age-specificity in Japan. *Eurosurveillance* 14: pii:19227.
- Nishiura H (2010) Travel and age of influenza A (H1N1) 2009 virus infection. *Journal of Travel Medicine* 17: 269–270.
- Mukherjee P, Lim PL, Chow A, Barkham T, Seow E, et al. (2010) Epidemiology of travel-associated pandemic (H1N1) 2009 infection in 116 patients, Singapore. *Emerging Infectious Diseases* 16: 21–26.
- Kermack WO, McKendrick AG (1927) Contributions to the mathematical theory of epidemics. I. *Proceedings of the Royal Society, Series A* 115: 700–721. (reprinted in *Bull Math Biol* 53 (1991) 33–118.
- Roberts MG (2006) Modeling strategies for containing an invading infection. *Mathematical Population Studies* 13: 205–214.
- Roberts MG, Baker K, Jennings LC, Sertsov G, Wilson N (2007) A model for the spread and control of pandemic influenza in an isolated geographical region. *Journal of the Royal Society Interface* 4: 325–330.
- Nishiura H, Chowell G, Safan M, Castillo-Chavez C (2010) Pros and cons of estimating the reproduction number from early epidemic growth rate of influenza A (H1N1) 2009. *Theoretical Biology and Medical Modelling* 7: 1.
- Nishiura H, Klinkenberg D, Roberts MG, Heesterbeek JAP (2009) Early epidemiological assessment of the virulence of emerging infectious diseases: a case study of an influenza pandemic. *PLoS ONE* 4: e6852.
- Bailey NTJ (1964) *The elements of stochastic processes with applications to the natural sciences*. New York: Wiley.
- Hahne S, Donker T, Meijer A, Timen A, van Steenberghe J, et al. (2009) Epidemiology and control of influenza A(H1N1)v in the Netherlands: the first 115 cases. *Eurosurveillance* 14: pii:19267.
- White LF, Wallinga J, Finelli L, Reed C, Riley S, et al. (2010) Estimation of the reproductive number and the serial interval in early phase of the 2009 influenza A/H1N1 pandemic in the USA. *Influenza and Other Respiratory Viruses* 3: 267–276.
- Yang Y, Sugimoto JD, Basta NE, Chao DL, Matrajt L, et al. (2009) The transmissibility and control of pandemic influenza A (H1N1) virus. *Science* 326: 729–733.
- Jagers P (1968) Age-dependent branching processes allowing immigration. *Theory of Probability and its Applications* 13: 225–236.
- Keyfitz BL, Keyfitz N (1997) The McKendrick partial differential equation and its uses in epidemiology and population study. *Mathematical and Computer Modelling* 26: 1–9.
- Grey DR (1976) On possible rates of growth of age-dependent branching processes with immigration. *Journal of Applied Probability* 13: 138–143.
- Parthasarathy PR, Sharafali M (1991) Age-dependent branching process with birth process immigration. *International Journal of Mathematical Education in Science and Technology* 22: 317–320.
- Lipsitch M, Bergstrom CT (2004) Invited commentary: real-time tracking of control measures for emerging infections. *American Journal of Epidemiology* 160: 517–519.
- White LF, Pagano M (2008) A likelihood-based method for real-time estimation of the serial interval and reproductive number of an epidemic. *Statistics in Medicine* 27: 2999–3016.
- Lloyd-Smith JD, Schreiber SJ, Kopp PE, Getz WM (2005) Superspreading and the effect of individual variation on disease emergence. *Nature* 438: 355–359.
- Miller JC, Davoudi B, Meza R, Slim AC, Pourboholoul B (2010) Epidemics with general generation interval distributions. *Journal of Theoretical Biology* 262: 107–115.
- Demetrius L (1977) Measures of fitness and demographic stability. *Proceedings of the National Academy of Science, USA* 74: 384–386.
- Tuljapurkar SD (1982) Why use population entropy? it determines the rate of convergence. *Journal of Mathematical Biology* 13: 325–337.
- Griffin JT, Garske T, Ghani AC, Clarke PS (2011) Joint estimation of the basic reproduction number and generation time parameters for infectious disease outbreaks. *Biostatistics*; in press (doi: 10.1093/biostatistics/kxq058).

39. Scalia Tomba G, Svensson A, Asikainen T, Giesecke J (2010) Some model based considerations on observing generation times for communicable diseases. *Mathematical Biosciences* 223: 24–31.
40. Nishiura H (2010) Time variations in the generation time of an infectious disease: implications for sampling to appropriately quantify transmission potential. *Mathematical Biosciences and Engineering* 7: 851–869.