Laboratory and Epidemiology Communications

The Relationship between the Cumulative Numbers of Cases and Deaths Reveals the Confirmed Case Fatality Ratio of a Novel Influenza A (H1N1) Virus

Hiroshi Nishiura1,2*
1PRESTO, Japan Science and Technology Agency, Saitama 332-0012, Japan; and 2Theoretical Epidemiology, University of Utrecht, Utrecht, The Netherlands

Communicated by Ichiro Kurane
(Accepted February 17, 2010)

When a novel influenza virus emerges, it is crucial to gain an understanding of the virulence at the very early stage of a pandemic. If a high mortality is predicted, then corresponding countermeasures can be chosen and implemented, and the extent of the pandemic in the presence and absence of interventions is subsequently estimated. In the ongoing pandemic involving the influenza A (H1N1) virus 2009, such an early assessment was initially made based on confirmed cases whose risk of death was estimated at up to 0.5% (1–3). Although it later appeared that the risk of death among all symptomatic cases with medical attendance (i.e., symptomatic case fatality ratio [sCFR]) ranged from approximately 0.02–0.05% (4,5), a ratio which is more useful for the assessment of virulence and prediction of the mortality impact, epidemiological interpretation of minimally available data during the very early stages of the pandemic in the present study involves confirmed cases alone.

Yoshikura (6) recently reported a linear relationship between the logarithms of the cumulative numbers of confirmed cases and deaths. In general, a log-log plot enables the relationship between two items to be interpreted and the relevant parameters estimated, given that an underlying power law mechanism exists (such as that observed for the individual contact heterogeneity of infectious diseases (7)). To facilitate epidemiological understanding of the underlying mechanisms behind the empirically observed linear relationship in a bottom-up fashion, a mechanistic statistical model is needed. This article aims to explicitly clarify the relationship between the cumulative numbers of confirmed cases and deaths, to apply this relationship to estimate the confirmed case fatality ratio of the 2009 H1N1 pandemic virus, and to compare the estimates between different countries.

Let the cumulative numbers of confirmed cases and deaths by calendar time $t$ be $C_t$ and $D_t$, respectively. Based on a well-known mechanistic relationship (1), the expectation of $D_t$ given $C_t$ is written as

$$E(D_t) = pu_t C_t$$  \[1\]

where $p$ is the confirmed case fatality ratio (cCFR), defined as an unbiased risk of death among confirmed cases. $u_t$ is referred to as a factor of underestimation (1), which reflects the proportion of confirmed cases whose risk of death has been observed by calendar time $t$, i.e.,

$$u_t = \frac{\sum_{r=0}^{\infty} C_{t-r} F_r}{C_t}$$  \[2\]

where $F_r$ is the conditional cumulative distribution function of the time from onset to death (given a fatal outcome). Usually, $u_t$ varies as a function of calendar time during the non-linear epidemic phase, but is independent of time during the initial exponential growth phase, resulting in

$$u = M(-r)$$  \[3\]

where $M(-r)$ is the moment-generating function of the time from onset to death, given an exponential growth rate of confirmed cases $r$. For instance, if $F_r$ follows a Weibull distribution with the shape and scale parameters $\alpha$ and $\beta$, respectively, we have

$$M(-r) = \sum_{n=0}^{\infty} \frac{(-r)^n \alpha^n}{n!} \Gamma \left(1 + \frac{n}{\beta} \right)$$  \[4\]

From equation [1], the log-log relationship between $C_t$ and $D_t$ is given by

$$\ln D_t = \ln C_t + \ln pu_t.$$  \[5\]

It should be noted that during the exponential growth phase, equation [1] yields a linear relationship between $C_t$ and $D_t$, even without logarithmic transformation. In other words, the slope $k$ described in Yoshikura (6) is always 1 as long as an epidemic follows an exponential growth phase, and the intercept $L$ is the product of two strictly interpretable quantities, i.e., $p$ and $u$, during that growth phase (it should be noted that $L$ varies with calendar time during the non-linear phase).

Figure 1 compares the early growth phases in five different countries that reported at least 10 deaths by 1 July 2009 (8). Consistent with equation [1], the relationship between $C_t$ and $D_t$ appears to be linear (Fig. 1A). However, the slopes are different. Argentina yielded the largest estimate (0.034) followed by Mexico (0.014), while the slopes of the remaining three countries, the United States, Canada, and Chile, ranged from 0.002–0.003. Figure 1B shows the exponential growth phase of confirmed cases during the corresponding period of observation, and compares the observed and predicted exponential growth of $C_t$. While the growth rate in Mexico alone was higher than the other growth rates (0.093 per day), the remaining four countries yielded similar rates to one another (0.05–0.06 per day), and thus experienced invasion of the H1N1 virus with similar transmission poten-

*Corresponding author: Mailing address: Theoretical Epidemiology, University of Utrecht, Yalelaan 7, Utrecht, 3584 CL, The Netherlands. Tel: +31-30-253-1233, Fax: +31-30-252-1887, E-mail: h.nishiura@uu.nl

---

**Corresponding author:** Mailing address: Theoretical Epidemiology, University of Utrecht, Yalelaan 7, Utrecht, 3584 CL, The Netherlands. Tel: +31-30-253-1233, Fax: +31-30-252-1887, E-mail: h.nishiura@uu.nl
where Obst is the observed cumulative number of deaths at time. The mean (standard deviation) was estimated to be 10.72 and 2.13, respectively, while the Cramer-von Mises goodness-of-fit test revealed no significant deviation between the observed data and the fitted Weibull distribution were estimated to be 9.5 (4.7) days. The parameters \( \alpha \) and \( \beta \) (Fig. 2).

Table 1. Estimates of the exponential growth rate, factor of underestimation and confirmed case fatality ratio of the novel influenza A (H1N1) virus in five different countries

<table>
<thead>
<tr>
<th>Country</th>
<th>Period</th>
<th>( r ) (^{1)}</th>
<th>( u ) (^{2)}</th>
<th>cCFR (95% CI) (^{3)}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argentina</td>
<td>17 June–6 July</td>
<td>0.059</td>
<td>59.1%</td>
<td>2.33 (1.97, 2.72)</td>
</tr>
<tr>
<td>Canada</td>
<td>29 May–1 July</td>
<td>0.060</td>
<td>58.6%</td>
<td>0.40 (0.33, 0.47)</td>
</tr>
<tr>
<td>Chile</td>
<td>10 June–3 July</td>
<td>0.063</td>
<td>57.5%</td>
<td>0.25 (0.18, 0.33)</td>
</tr>
<tr>
<td>Mexico</td>
<td>5 May–21 May</td>
<td>0.093</td>
<td>45.0%</td>
<td>5.58 (5.15, 6.02)</td>
</tr>
<tr>
<td>USA</td>
<td>9 May–15 June</td>
<td>0.052</td>
<td>62.6%</td>
<td>0.25 (0.22, 0.29)</td>
</tr>
</tbody>
</table>

\(^1)\) The exponential growth rate of confirmed cases (per day).

\(^2)\) The factor of underestimation (1), which reflects the proportion of confirmed cases whose risk of death has already been observed by the end of the observation period.

\(^3)\) cCFR, the unbiased confirmed case fatality ratio, regarded as the risk of death among confirmed cases. The 95\% confidence interval (CI) is derived from the profile likelihood.

Fig. 1. Epidemiological dynamics of the novel influenza A (H1N1) virus in five different countries. (A) Linear relationship between the cumulative numbers of confirmed cases and deaths. Each line represents the best approximation of the relationship during the exponential growth phase of confirmed cases. (B) Exponential growth of confirmed cases as a function of time. The time (horizontal axis) is counted from 24 April 2009 onward. Each line represents the predicted cumulative number of confirmed cases assuming exponential growth (i.e., \( C(t) = \alpha \exp(\beta t) \)).

Given that a linear relationship between \( C_t \) and \( D_t \), the times from onset to death \( n = 83 \) were analyzed to calculate \( u \) using equations [3] and [4] (Fig. 2). The mean (standard deviation) was estimated to be 9.5 (4.7) days. The parameters \( \alpha \) and \( \beta \) for the Weibull distribution were estimated to be 10.72 and 2.13, respectively, while the Cramer-von Mises goodness-of-fit test revealed no significant deviation between the observed data and the fitted Weibull distribution. The Weibull plot compares the quantiles of the observed and predicted distributions, with the Weibull quantiles on the vertical axis and the observed values on the horizontal axis. The straight diagonal reference line is a theoretical line of fit. As the observed data fit the Weibull distribution well, the values fall approximately on the reference line. The dashed lines represent 95\% confidence limits.

To further clarify the information that can be extracted from the relationship between \( C_t \) and \( D_t \), the times from onset to death \( n = 83 \) were analyzed to calculate \( u \) using equations [3] and [4] (Fig. 2). The mean (standard deviation) was estimated to be 9.5 (4.7) days. The parameters \( \alpha \) and \( \beta \) for the Weibull distribution were estimated to be 10.72 and 2.13, respectively, while the Cramer-von Mises goodness-of-fit test revealed no significant deviation between the observed data and the fitted Weibull distribution. Subsequently, the exponential growth rate in Argentina was similar to that estimated in the United States, Canada, and Chile, the abovementioned points (A) differential efficacy of medical treatment, (B) clustering of cases among high-risk persons (e.g., asthmatic and chronic renal failure patients and pregnant women), (C) differential susceptibility (e.g., ethnicity and genetic susceptibility), (D) differential efficacy of medical treatment, and so on. Given that the exponential growth rate in Argentina was similar to that estimated in the United States, Canada, and Chile, the abovementioned points (A)-(D) may be the case. However for Mexico, it should be noted that the data were reported much later than the actual early exponential growth phase, reflecting the dynamics near the peak of the first epidemic curve in Mexico. Understanding the epidemiological dynamics in Mexico requires reconstruction of the early epidemic curve, which could imply a substantial number of unreported (unobserved) cases.
Lastly, why was the slope $k$ of the log-log plot in Yoshikura (6) maintained with $k > 1$ even during the non-linear phase of the pandemic? Given that the product of $p$ and $u_t$ is small (and thus, its negative logarithm is large), its influence on the right-hand side of equation [5] is not negligible. Since $u_t$ tends to decrease as a function of the epidemic time, the decrease in $u_t$ could have resulted in a considerable decrease in the right-hand side of equation [5], thereby resulting in a need for $k > 1$ to model $D_t$ as a function of $C_t$ alone.

In summary, the explicit relationship in equation [1] not only explains the underlying mechanisms behind the linear extrapolation in Yoshikura (6), but also permitted us to implement a comparative estimation study of the cCFRs between different countries.

The author is grateful to Dr. Hiroshi Yoshikura for useful discussions and comments. This work was supported by the JST PRESTO program.

REFERENCES