JOINT QUANTIFICATION OF TRANSMISSION DYNAMICS AND DIAGNOSTIC ACCURACY APPLIED TO INFLUENZA

Hiroshi Nishiura

[1] PRESTO, Japan Science and Technology Agency (JST)
4-1-8 Honcho Kawaguchi, Saitama 332-0012, Japan
[2] Theoretical Epidemiology, University of Utrecht
Yalelaan 7, 3584 CL, Utrecht, The Netherlands

ABSTRACT. The influenza A (H1N1) pandemic 2009 posed an epidemiological challenge in ascertaining all cases. Although the counting of all influenza cases in real time is often not feasible, empirical observations always involve diagnostic test procedures. This offers an opportunity to jointly quantify transmission dynamics and diagnostic accuracy. We have developed a joint estimation procedure that exploits parsimonious models to describe the epidemic dynamics and that parameterizes the number of test positives and test negatives as a function of time. Our analyses of simulated data and data from the empirical observation of interpandemic influenza A (H1N1) from 2007-08 in Japan indicate that the proposed approach permits a more precise quantification of the transmission dynamics compared to methods that rely on test positive cases alone. The analysis of entry screening data for the H1N1 pandemic 2009 at Tokyo-Narita airport helped us quantify the very limited specificity of influenza-like illness in detecting actual influenza cases in the passengers. The joint quantification does not require us to condition diagnostic accuracy on any pre-defined study population. Our study suggests that by consistently reporting both test positive and test negative cases, the usefulness of extractable information from routine surveillance record of infectious diseases would be maximized.

1. **Introduction.** The pandemic influenza A (H1N1) 2009 virus has spread worldwide. After it was first identified in early 2009, mathematical models have been widely used to answer policy relevant questions [15, 37]. Although scientific approaches are not exclusively limited to epidemiological modeling, early modeling efforts gave a critically important indication of the *pandemic potential*, namely, the transmission potential and the severity of the disease. The transmission potential was measured by estimating the reproduction number; the average number of secondary cases generated by a typical primary case [7, 27, 33, 40]. The severity of the disease was assessed mainly from the case fatality ratio (CFR); the conditional risk of death for patients with a disease or infection [7, 8, 23, 28, 33]. By combining these two indicators a theoretical estimate of the magnitude of the pandemic, the expected proportion of influenza-associated deaths among the total population, is obtained [22].

However, there are two technical difficulties that prevent a precise estimation of

---

the pandemic potential during the very early stages of the pandemic. One is the heterogeneity in transmission and in the conditional risk of death from the disease [9]. In one specific case with a high degree of assortative mixing, an epidemic starting with school clusters greatly complicated early quantification of the next-generation matrix [22] (the square matrix which describes within- and between-group transmission). The other technical issue is the unobservability of the infection event. Except for humans infected with the highly pathogenic avian influenza virus, influenza is generally a mild disease involving asymptomatic infections. Because of this, before the 2009 pandemic, it was recognized that counting in real time all the individuals infected with influenza was impractical [32]. Indeed, the difficulty in ascertaining cases of influenza has had a profound impact on the statistical estimation of CFR [18]. For H1N1-2009 infections, the CFR conditioned on confirmed cases was estimated to be 0.5% [7, 8, 23], but the estimate conditioned on all symptomatic cases with medical attendance was approximately 0.05%, a tenth of the estimate that relied on confirmatory diagnosis [28]. Because the above mentioned issues that affect the estimation of CFR have been discussed elsewhere [13, 19], the present study focuses on the difficulty in ascertaining cases and on the estimation of the transmission potential.

A direct method to empirically observe the number (or the proportion) of all infected individuals in a population is to conduct a serological survey. Such a population-wide survey is expensive, and is not logistically feasible for epidemiological assessment in real time (especially if the population size is large). The modeling community has yet to develop a real-time method for indirectly estimating the total number of infected individuals in a population.

Even provided that the real time counting of all infected individuals is not technically feasible, it should be remembered that the empirical observation of an epidemic always involves diagnostic testing procedures. Even when an influenza *case* does not undergo any laboratory testing, the case must still meet some clinical criteria indicating *possible* influenza. For example, influenza-like illness (ILI) is usually defined as fever $> 38.0\ ^\circ$C and cough and/or sore throat. A more detailed surveillance might be based on the notification of cases with positive test results from rapid diagnostic testing (RDT). Although the diagnostic accuracy of ILI and RDT is not sufficiently high to include all influenza cases and to exclude all other causes [4, 12, 35], by using both positive and negative test results, diagnostic testing offers an opportunity to maximize the usefulness of surveillance data to estimate the incidence of cases among those that undergo testing.

Here we propose a way to maximize the use of the extractable information by jointly quantifying both the transmission dynamics and diagnostic accuracy. The present study offers a method for the joint estimation procedure with particular emphasis on parsimonious modeling approaches. In the next section, we discuss the epidemiological concept of clinical diagnosis with a binary outcome. In Section 3, the estimation framework is illustrated with simulated data. Subsequently, the proposed methods are applied to two empirical influenza datasets.

2. **The epidemiological concept of a diagnostic test.** While clinical performance of diagnostic tests is often assessed by several epidemiological measurements, e.g., accuracy, precision and reliability, in the present study we focus on the diagnostic accuracy because that is what determines the data generating process of our empirical data. Table 1 illustrates a two-by-two table of diagnostic test results with

TABLE 1. Data generating process for an infectious disease following a diagnostic test procedure.

	True positives	True negatives	Total
Test positives	αc_t	$(1-\beta)k_t$	P_t
Test negatives	$(1-\alpha)c_t$	βk_t	N_t

c_t and k_t are the incidence (true positives) and true negatives on day t, respectively.
α and β denote sensitivity and specificity, respectively.
P_t and N_t are the observed counts of test positives and test negatives on day t, respectively.

a binary classification (infection or non-infection). Diagnostic accuracy is expressed as *sensitivity* α and *specificity* β. Sensitivity is defined as the proportion of actual test positives among the total number of true positive individuals c_t on day t, while specificity is defined as the proportion of test negatives among the total number of true negative individuals, k_t on day t. The *test* can be anything that determines the presence of infection (e.g., ILI and RDT). Prior to the test, c_t and k_t, defined as *true positives* and *true negatives* respectively, are usually unknown. Rather, empirical data give the total numbers of *test positives* P_t and *test negatives* N_t on day t. Because we are considering epidemic dynamics, we assume that c_t, P_t and N_t depend on calendar time t, and that the measures of diagnostic accuracy, α and β, are independent of time.

If epidemic data rely on a diagnostic testing procedure, a simple and commonly used estimate of the incidence (or prevalence in the case of endemic equilibrium) is the number of test positive results, P_t. However, P_t depends not only on the actual incidence (true positives), but also on the diagnostic accuracy of the test. An improvement in incidence estimations is achieved by correcting P_t to account for α and β [30]. Table 1 gives

$$\alpha c_t + (1-\beta)k_t = P_t, \tag{1}$$

which can be rearranged as

$$\hat{c}_t = \frac{P_t + \hat{\beta} - 1}{\hat{\alpha} + \hat{\beta} - 1}. \tag{2}$$

Equation (2) suggests that the estimate of incidence (true positives) is obtained when α and β are known (prior to the surveillance).

However, the equation (2) is only applicable to observations on clearly defined populations that undergo diagnostic testing. In the literature, α and β are estimated for various diagnostic testing procedures, but the published estimates have often been conditioned on a pre-defined study population. For example, a study in Saudi Arabia which assessed the diagnostic accuracy of ILI in correctly diagnosing actual influenza estimated $\alpha = 0.67$ and $\beta = 0.64$. However, because the study population was limited to 555 pilgrims with upper respiratory symptoms who presented at the authors' clinics during the Hajj [29], the estimates are not strictly applicable to other settings. We, therefore, consider a way to jointly estimate α and β along with c_t and k_t by parameterizing c_t and k_t using the theoretical support derived from epidemic modeling.

It should be noted that our data are for individuals that have undergone diagnostic testing only once. For a more precise estimation of the incidence, we would ideally need test results from repeated measurements of a cohort population [39].
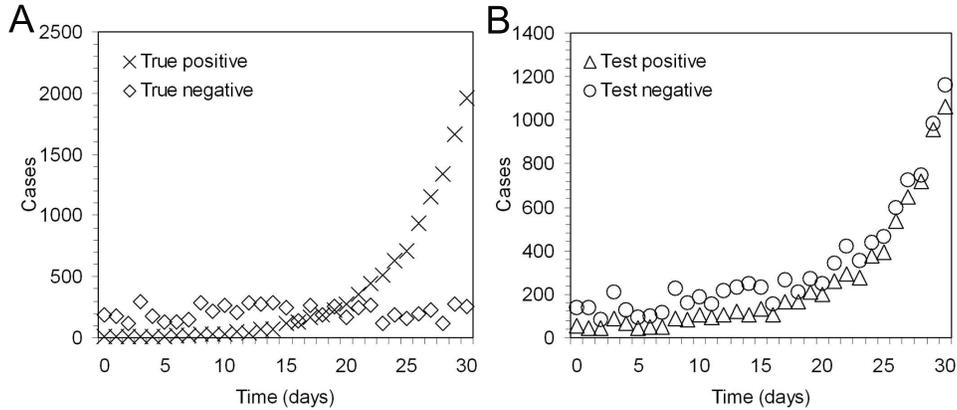
FIGURE 1. **Simulated data for early epidemic growth that account for the result of a diagnostic testing procedure. A**. The temporal distribution of infected and non-infected individuals. True positives represent the incidence (the number of new infections on day t) with exponential growth. True negatives are treated as noise. **B**. The time evolution of the observable empirical data. Cases testing positive and negative are readily observed on each day.

However, the most widely available data (routine surveillance based on passive notifications) do not include multiple (i.e. repeated) diagnostic test results for a single individual. Therefore, the present study also differs from the series of studies that employed a capture-recapture method to estimate the completeness of ascertainment of cases [38].

3. **An illustration of joint estimation.** The simulated data for the daily counts of true positives c_t and true negatives k_t are shown in Figure 1A. Because the simulated data are only for the first 1 month of an epidemic, we assume that the incidence (true positives) grows exponentially with an intrinsic growth rate $r = 0.2$ per day during this period. If the number of cases on day 0 is assumed to be $c_0 = 5$ cases, the expected value of the incidence on day t, $\mathrm{E}(c_t)$ is

$$\mathrm{E}(c_t) = c_0 \exp(rt). \tag{3}$$

For each day t, we randomly generate Poisson variates based on the value of $\mathrm{E}(c_t)$. For k_t, we assume that over time the true negatives simply act as noise, and we generate a uniform distribution of the random integers for between 100 and 300 persons (so that $\mathrm{E}(k_t) = 200$ with a variance-to-mean ratio far greater than 1).

Figure 1B shows the test results as a function of time, representing usual empirical observation. Sensitivity α and specificity β are assumed to be 0.5 and 0.7, respectively. These values are in concordance with published estimates of diagnostic accuracy of *documented fever* in diagnosing seasonal influenza in the USA [2]. Let p_t be a random variable which yields an estimator of test positive individuals, P_t. The diagnostic process of test positives, p_t is given by a sum of two random numbers,

$$p_t = \mathrm{Bin}(c_t, \alpha) + \mathrm{Bin}(k_t, 1 - \beta), \tag{4}$$

where $\mathrm{Bin}(n, m)$ represents a random number from a binomial distribution with the number of trials n and the probability of the event of interest m. Let n_t be a random variable representing the number of test negative individuals. Because the total number of cases undertaking diagnostic test on day t is determined by the underlying dynamics (i.e. c_t and k_t), and thus, is independent of the diagnostic process, n_t is calcualted as $c_t + k_t - p_t$.

As discussed in the previous section, one may rely only on the positive test results P_t in Figure 1B to analyze the transmission dynamics. However, true positives that are missed from P_t are included in N_t (and true negatives are included in P_t). For example, if we model the test positives to perfectly reflect exponential growth, i.e., $\mathrm{E}(Q_t) = q_0 \exp(r_q t)$, and assume that the likelihood of estimating q_0 and the growth rate r_q is proportional to

$$\prod_{t=0}^{30} \mathrm{E}(Q_t)^{P_t} \exp(-\mathrm{E}(Q_t)), \tag{5}$$

then, the maximum likelihood estimate (and the lower and upper 95% confidence intervals (CI) derived from profile likelihood) of r_q is 0.190 (95% CI: 0.190, 0.191), an underestimation of the actual incidence growth rate.

We, therefore, consider a data generating process for both P_t and N_t. Although passive surveillance usually does not count test negatives, we consider the case where both datasets are consistently recorded over time (see Discussion). From Table 1, the expected value of P_t, $\mathrm{E}(P_t)$ is

$$\mathrm{E}(P_t) = \mathrm{E}(c_t)\alpha + \mathrm{E}(k_t)(1 - \beta), \tag{6}$$

where

$$\mathrm{E}(c_t) = a \exp(bt), \tag{7}$$
$$\mathrm{E}(k_t) = d, \tag{8}$$

where a, b and d are the parameters. a represents the number of new cases on day 0, b is the exponential growth rate of true positives, and d is the number of true negatives which is assumed to be independent of time t. Similarly, the expected value of N_t, $\mathrm{E}(N_t)$ is modeled as

$$\mathrm{E}(N_t) = \mathrm{E}(c_t)(1 - \alpha) + \mathrm{E}(k_t)\beta. \tag{9}$$

As well as modeling P_t and N_t alone, the sum $S_t = P_t + N_t$ is independent of diagnostic process and so informative, i.e.,

$$\mathrm{E}(S_t) = a \exp(bt) + d. \tag{10}$$

In total, there are five unknown parameters, α, β, a, b and d. We naturally extend the likelihood in (5) to multivariate Poisson setting which allows the same covariance between all the pairs. The likelihood of estimating the five parameters is proportional to

$$\prod_t \mathrm{E}(P_t)^{P_t} \mathrm{E}(N_t)^{N_t} \mathrm{E}(S_t)^{P_t + N_t} \exp(-\mathrm{E}(P_t)) \exp(-\mathrm{E}(N_t)) \exp(-\mathrm{E}(S_t)). \tag{11}$$

Figure 2 shows the comparison of the simulated data and the expected values from the model mentioned above. Not only the observable counts (Figure 2B) but the daily numbers of true positives and true negatives are also precisely captured by the model. α and β are estimated to be 0.501 (95% CI: 0.500, 0.501) and 0.694 (95% CI: 0.687, 0.702), respectively. Maximum likelihood estimates of c_0 and r are 4.93
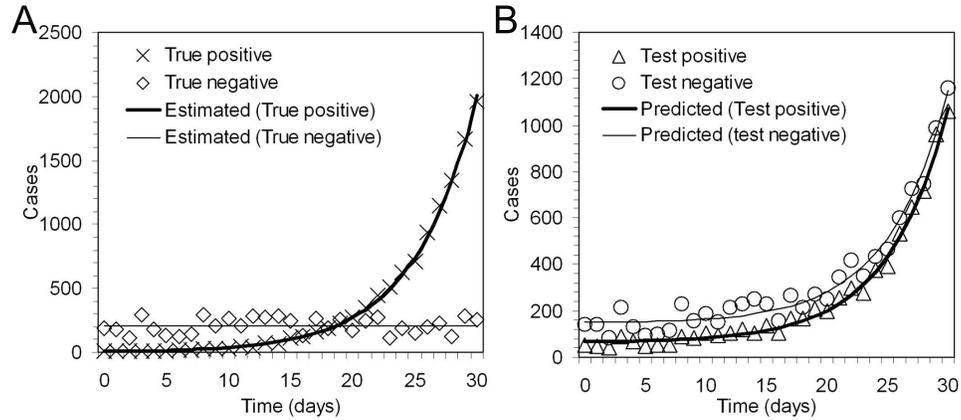
FIGURE 2. **Comparisons of the simulated and estimated early growth of an epidemic. A**. Comparison of the incidence and true negatives for the simulated and estimated data. **B**. Comparison of test positives and test negatives for the simulated and estimated data.

(95% CI: 4.88, 4.98) and 0.200 (95% CI: 0.200, 0.201), respectively. The intrinsic growth rate r when inserted into the Euler-Lotka equation, yields an estimate for the reproduction number, R (see next section) [20].

By exploiting the exponential growth of an epidemic during its early phase, our simple model permits joint quantification of the epidemic dynamics (the intrinsic growth rate in the illustrated example) and diagnostic accuracy. It should be noted that an important key data source for this framework is the counts of both test positive and test negative individuals. Although this framework is not applicable to all practical settings (e.g., the model may not converge if the incidence continues to be much smaller than the true negatives and if both the sensitivity and the specificity are extremely small), the proposed approach is theoretically valid for any infectious disease and any diagnostic test. Although this approach does not capture very mild cases without medical attendance (e.g., asymptomatic individuals), the proposed model can take into account all the individuals who undergo diagnostic testing as long as the assumed model appropriately captures actual transmission dynamics.

4. **The application of joint quantification to an entire epidemic curve.** We consider empirical datasets for interpandemic (seasonal) influenza A (H1N1) from October 2007 to August 2008 in the Mie prefecture, Japan [14]. During the local surveillance of interpandemic influenza in Mie, seven rapid diagnostic test kits were used, and the weekly counts of test positives and test negatives were reported from 73 hospitals in this prefecture throughout the epidemic season (week 42 in 2007 to week 35 in 2008; Figure 3). During this period, a total of 29,332 individuals underwent a rapid diagnostic test, and 13,055 (44.5 %) tested positive. This epidemic was exclusively caused by A/Brisbane/59/2007 (H1N1) virus, and those testing positive to type A virus accounted for 12,640 (96.8 %) of all positive cases. Although 382 of the remaining individuals tested positive to type B virus and the type-specificity of 33 others was unknown, we ignored the contributions of type B virus to the epidemic dynamics.
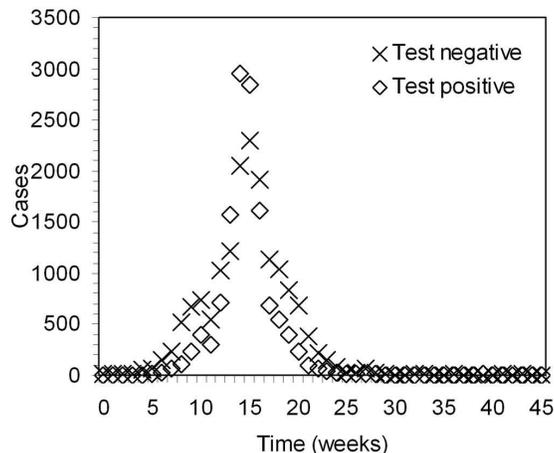
FIGURE 3. **Epidemic curves for interpandemic influenza A (H1N1) from October 2007 to August 2008, Mie prefecture, Japan.** The weekly numbers of influenza-like illness cases testing positive and negative to a rapid diagnostic test kit were recorded. The weeks were counted from week 42 (the week ending 21 October) 2007 onwards. The 2007/08 epidemic season was exclusively caused by A/Brisbane/59/2007 (H1N1). Although 3% of cases tested positive to the influenza B virus, the contributions of the type B virus to the epidemic dynamics are ignored.

It should be noted that the rapid diagnostic test (RDT) kits are known to be very specific, while the sensitivity tends to be relatively low [12]. Thus, while RDT is very useful for diagnosis by exclusion, it is not capable of detecting all influenza cases. Given that empirical estimates of the specificity for the majority of RDT are 100 %, to estimate the sensitivity α of the RDT and the reproduction number R, we fixed β at 1. Because of the unavailability of information for the test kit employed for each case, we assume that α does not vary between RDT kits [12].

The difference between this empirical example and the simulated data in the last section is that the entire epidemic curve is observed in Figure 3. Because we do not intend to fully investigate the underlying epidemic dynamics, we focus on the reproduction number R and some additional characteristics (e.g., the turning point of an epidemic) that constitute the information that we aim to extract from the dataset. Thus, we employ a parsimonious model and assume that the epidemic dynamics (the cumulative incidence $C(t)$ by the end of week t) are governed by the following general form of the logistic equation

$$C(t) = \frac{Z}{(1 + \exp(-r(t - t_m)))^{\frac{1}{g}}}, \tag{12}$$

where Z is the carrying capacity of the total number of cases, which is equivalent to the absolute final number of cases; r is the intrinsic growth rate; t_m satisfies $t_m = t_i + \ln(g)/r$ where t_i is the inflection point; and g is the exponent of deviation from the standard logistic curve. Equation (12) is referred to as the Richards model or the generalized logistic equation [3]. It is known to be very flexible and is
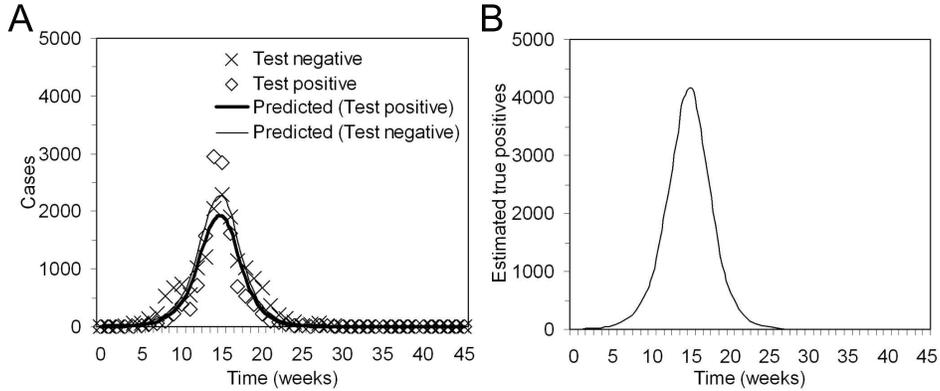
FIGURE 4. **Model-based estimates of interpandemic influenza A (H1N1) from October 2007 to August 2008 in the Mie prefecture, Japan** [14]. **A**. Comparisons of test positives and test negatives for the observed and estimated data. **B**. Estimated true positives (incidence) as a function of time. The weeks are counted from week 42 (the week ending 21 October) 2007 onwards.

useful for realizing the unimodal epidemic curve. This approach has been applied elsewhere to other infectious diseases [10, 11].

The expected value of true positives in week t is approximated by the difference between $C(t)$ and $C(t-1)$,

$$\mathrm{E}(c_t) = C(t) - C(t-1), \tag{13}$$

for $t \geq 1$ and the expected total number of reports in week t is

$$\mathrm{E}(S_t) = C(t) - C(t-1) + d, \tag{14}$$

where d is again the number of true negatives which is assumed to be independent of time t. The other assumptions are the same as those described in the last section and the likelihood equation is identical to (11). We estimate six parameters, α, d, r, Z, t_m and g, by minimizing the negative logarithm of (11).

Figure 4A compares observed and predicted weekly counts of test positives and test negatives. Although the peak incidence in test positives was not fully realized, the parsimonious model (12) described the observed pattern well. The sensitivity α of RDT was estimated to be 0.462 (95% CI: 0.461, 0.463), a little less than the published estimates based on pre-defined study populations. It should be noted that in the present case study, α is not conditioned on a pre-defined study population (clinical criteria are not considered for inclusion, as is the case in many empirical studies) but could be interpreted as being conditioned on all the individuals who undertook RDT in the 73 hospitals in Mie. The intrinsic growth rate r was estimated to be 0.621 (95% CI: 0.613, 0.628) per week. Therefore, if the generation time distribution follows an exponential distribution with a mean of $T_g = 3$ days, the reproduction number is $R = 1 + rT_g/7 = 1.27$. If the generation time is constant with a mean of $T_g = 3$ days, the reproduction number is $R = \exp(rT_g/7) = 1.30$. These estimates are consistent with other interpandemic settings in temperate zones [5]. The final size Z was 28,255 (95% CI: 28209, 28301) cases, and true negatives
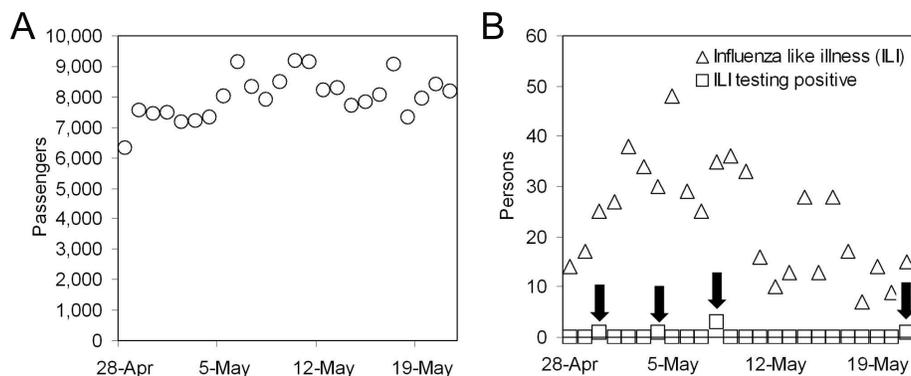
FIGURE 5. **Immigration and screening results at the Tokyo-Narita International Airport, April-May 2009** [17]. **A**. The daily number of passengers from Canada, Mexico and the USA from 28 April to 21 May 2009. The most strict entry screening measure was implemented during this period. Upon arrival and before unshipping, all the passengers were questioned about the presence of any suspicious symptoms as well as potential contacts and were screened for fever using a portable thermoscanner. A total of 191,969 passengers were screened. **B**. The daily number of passengers with influenza-like illness (ILI) and the number of ILI cases testing positive to a rapid diagnostic test. A total of 561 passengers were diagnosed as ILI and of these, 6 tested positive to rapid diagnostic testing. Five of the 6 cases were later confirmed to be influenza A (H1N1) 2009. Because the number of test positive ILI cases was small, non-zero counts are indicated by arrows.

were estimated to be 27.9 (95% CI: 26.8, 29.1) individuals in each week. t_m and g were estimated at 15.5 (15.4, 15.5) and 1.13 (1.11, 1.15), respectively.

Figure 4B shows the predicted number of true positives. It should be noted that the curve for true positives yields a higher peak incidence than that for test positives alone, indicating a critical need to address diagnostic accuracy and to understand actual epidemic dynamics. Indeed, understanding the height of peak incidence is crucial for estimating the burden of hospitalizations to, for example, calculate the required number of hospital beds during a future epidemic.

5. **The application of joint quantification to entry screening data.** We now consider the application of our model to a different dataset, the result of entry screening during the very early stages of the H1N1-2009 pandemic from late April to May 2009 in Japan [17]. Figure 5 illustrates the statistics of the entry screening results at Tokyo-Narita International airport which handles the majority of international passenger traffic to and from Japan. Epidemics in the USA and Mexico, caused by a novel influenza A (H1N1) virus, were officially announced on 24 April. This was followed by the recognition of the epidemic in Canada a few days later. Japan started a strict entry screening measure on 28 April 2009 targeting all passengers arriving from Canada, Mexico and the USA. The entry screening for all aircraft arriving from these three countries included *onboard quarantine* during

which medical specialists searched for febrile passengers and their contacts using portable thermoscanner and interviewed any passengers with suspicious symptoms before disembarkation. The entry screening measure was carried out until 18 June 2009 and involved more than two million passengers. The strictest onboard quarantine for all the aircrafts from Canada, Mexico and the USA was conducted from 28 April to 21 May. Figure 5A shows the daily number of passengers arriving at Narita airport from these three countries. During this period, the total number of passengers screened was 191,969.

During the entry screening, the definition of influenza-like illness (ILI) was a little broader than the usual criteria mentioned earlier. ILI during the entry screening was defined as the presence of fever $> 38.0\ ^\circ$C *or* any acute respiratory symptom. In addition, passengers with a history of probable contact with a case or with a travel history to Canada, Mexico or the USA were subjected to entry screening regardless of ILI. All ILI cases were subjected to rapid diagnostic testing and, if they tested positive to type A influenza virus, were identified as a suspected case to be further examined using RT-PCR for confirmatory diagnosis. Figure 5B shows the observed daily numbers of ILI cases and the number of ILI cases with positive test results. During the period of interest, 561 passengers (0.29 % among the total passengers) were identified as having ILI. Of these, six (1.07 %) tested positive to influenza A virus during rapid diagnostic testing and five of them were later confirmed to be influenza A (H1N1) 2009.

Apart from Japan, many other Asian countries carried out similar (or less strict) entry screening during the very early stages of the 2009 pandemic. Asian countries adopted screening as one of the countermeasures to be implemented to attain early containment (in anticipating the possible emergence of highly pathogenic avian influenza prior to the 2009 pandemic [34]). Nevertheless, even before the 2009 pandemic, it was recognized that entry screening might still allow substantial number of infected-and-incubating individuals to enter a new community [26]. Moreover, an empirical study of the 2009 pandemic has shown that entry screening was not associated with a substantial delay in the start of local transmission [6]. Indeed, without restricting the movement of a substantial fraction of the passengers at risk (corresponding to the etymological root of *quarantine* in restricting movement for 40 days) [25], the introduction of cases and local transmission are unavoidable [31] (although the effectiveness of entry screening is, of course, not zero).

In the present study we do not consider the overall protective effect of entry screening. Rather, we ignore the entrance of incubating individuals upon arrival and investigate how well diagnostic procedures were able to detect febrile passengers with influenza during the entry screening period. Specifically, the focus of the present study is the diagnostic accuracy of ILI and of the subsequent rapid diagnostic testing during the entry screening. Assessing the specificity β_1 of ILI using this dataset is critically important because the screened population is not conditioned on any specific medical states (except for a possible exposure in Canada, Mexico or the USA). The data, thus, provide an extremely rare opportunity for measuring the usefulness of ILI for screening a general population.

Let α_1 and β_1 denote the sensitivity and specificity of ILI, respectively, during the entry screening at Narita airport. Using an empirical result from Narita, α_1 was fixed at 0.25; from a total of 16 cases of influenza (including pandemic and interpandemic strains), 4 (25 %) exhibited high fever $> 38.0\ ^\circ$C upon arrival. Since ILI is non-specific, we estimated β_1 from the dataset in Figure 5. Similarly, let α_2

Table 2. Data generating process for an infectious disease following a two-step diagnostic procedure

	True positive	True negative	Total
Test positive	$\alpha_1 \alpha_2 c_t$	$(1 - \beta_1)(1 - \beta_2)k_t$	P_t
Test negative	$\alpha_1(1 - \alpha_2)c_t$	$(1 - \beta_1)\beta_2 k_t$	N_t

c_t and k_t are the incidence (true positives) and true negatives on day t, respectively.
α_1 and β_1 are the sensitivity and specificity, respectively, of the first screening test (ILI; influenza-like illness). Only those testing positive during the first screening were further examined.
α_2 and β_2 are the sensitivity and specificity, respectively, of second diagnostic testing.
P_t and N_t are the observed counts of test positives and test negatives (to rapid diagnostic testing) on day t, respectively.

and β_2 represent the sensitivity and specificity of rapid diagnostic test, respectively. As in the last section, β_2 was fixed at 1, and the sensitivity of the rapid diagnostic test was estimated. It should be noted that the measurements, α_1, α_2, β_1 and β_2 all represent the diagnostic accuracy in detecting febrile influenza cases upon arrival.

Let c_t and k_t be the incidence (true positives) and true negatives, respectively. Table 2 summarizes the data generating process. We assume that the diagnostic accuracy of the two tests is independent. Only cases that met the definition of ILI during entry screening were further examined by rapid diagnostic testing in Narita. Therefore, the total number of individuals who underwent both tests is represented by $c_t\alpha_1 + k_t(1 - \beta_1)$. Of these, $c_t\alpha_1\alpha_2 + k_t(1 - \beta_1)(1 - \beta_2)$ tested positive to rapid diagnostic testing and the remaining $c_t\alpha_1(1 - \alpha_2) + k_t(1 - \beta_1)\beta_2$ tested negative. Those individuals who did not undergo rapid diagnostic testing (because of non-ILI) are also informative, and are expressed as $c_t(1 - \alpha_1) + k_t\beta_1$.

Because we are considering the early stages of a pandemic, true positives and true negatives are modeled as

$$\mathrm{E}(c_t) \;=\; a \exp(bt), \tag{15}$$

$$\mathrm{E}(k_t) \;=\; d_1 + d_2 t, \tag{16}$$

describing c_t as an exponential function. As we described earlier, a represents the number of new cases on day 0 and b is the exponential growth rate of true positives. The true negatives are also allowed to vary as a function of time (with d_1 individuals on day 0 and linear coefficient d_2) because of variations in the total number of passengers with time (Figure 5A). Accordingly, the expected numbers of ILI positive and ILI negative individuals are

$$\mathrm{E}(P_{1t}) \;=\; \mathrm{E}(c_t)\alpha_1 + \mathrm{E}(k_t)(1 - \beta_1), \tag{17}$$

$$\mathrm{E}(N_{1t}) \;=\; \mathrm{E}(c_t)(1 - \alpha_1) + \mathrm{E}(k_t)\beta_1. \tag{18}$$

As before, the total number of passengers S_{1t} is informative:

$$\mathrm{E}(S_{1t}) = a \exp(bt) + d_1 + d_2 t. \tag{19}$$

Thus, the likelihood for the observation of the presence or absence of ILI, L_1, is proportional to

$$\prod_t \mathrm{E}(P_{1t})^{P_{1t}} \mathrm{E}(N_{1t})^{N_{1t}} \mathrm{E}(S_{1t})^{P_{1t}+N_{1t}} \exp(-\mathrm{E}(P_{1t})) \exp(-\mathrm{E}(N_{1t})) \exp(-\mathrm{E}(S_{1t})).$$
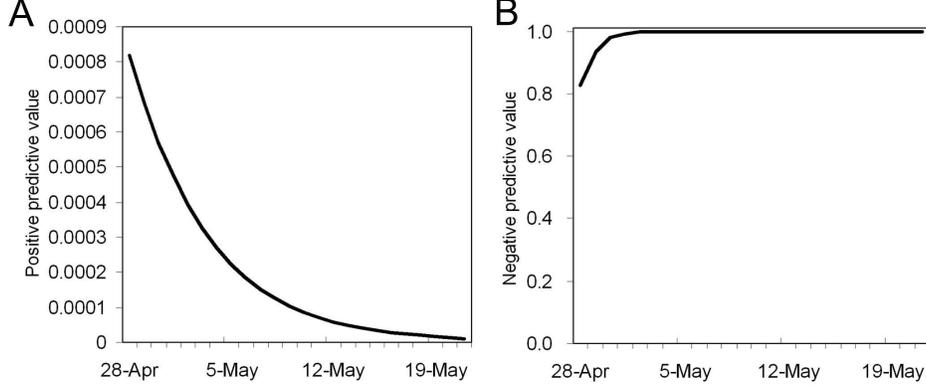
$$\tag{20}$$

FIGURE 6. **The diagnostic performance of fever screening upon arrival. A** shows the positive predictive value, the proportion of passengers with positive ILI results who were correctly diagnosed, and **B** shows the negative predictive value, the proportion of passengers with negative ILI results who were correctly diagnosed (as not having influenza).

Similarly, for the rapid diagnostic test (RDT) results, the expected numbers of RDT positive and RDT negative individuals are

$$\mathrm{E}(P_{2t}) = \mathrm{E}(c_t)\alpha_1\alpha_2 + \mathrm{E}(k_t)(1-\beta_1)(1-\beta_2), \qquad (21)$$
$$\mathrm{E}(N_{2t}) = \mathrm{E}(c_t)\alpha_1(1-\alpha_2) + \mathrm{E}(k_t)(1-\beta_1)\beta_2. \qquad (22)$$

Thus, the likelihood for the observation of RDT results, L_2, is proportional to

$$\prod_t \mathrm{E}(P_{2t})^{P_{2t}} \mathrm{E}(N_{2t})^{N_{2t}} \exp(-\mathrm{E}(P_{2t})) \exp(-\mathrm{E}(N_{2t})). \qquad (23)$$

Accordingly, the total likelihood L to estimate the six parameters (β_1, α_2, a, b, d_1 and d_2) is

$$L = L_1 L_2. \qquad (24)$$

It is not enough to focus on estimates of β_1 and α_2, because an understanding of the diagnostic performance of the entry screening is necessary to be able to evaluate the feasibility of an entry screening measure that relies on ILI as the first screening method. We have, therefore, measured the positive predictive value (PPV) and the negative predictive value (NPV) of employing ILI for screening influenza cases. PPV represents the proportion of cases with positive ILI results and NPV represents the proportion of cases with negative ILI results that were correctly diagnosed. In equations, they are written as

$$\mathrm{PPV} = \frac{\hat{c}_t\hat{\alpha}_1}{\hat{c}_t\hat{\alpha}_1 + \hat{k}_t(1-\hat{\beta}_1)}, \qquad (25)$$

$$\mathrm{NPV} = \frac{\hat{k}_t\hat{\beta}_1}{\hat{c}_t(1-\hat{\alpha}_1) + \hat{k}_t\hat{\beta}_1}. \qquad (26)$$

Based on the empirical data in Figure 5, the specificity of ILI β_1 was estimated to be 0.250 (95% CI: 0.249, 0.251). The sensitivity of RDT α_2 was estimated to

be 0.499 (95% CI: 0.099, 0.899). This indicates that ILI incorrectly identifies 75% of true negative individuals as possible influenza cases, and that the RDT which follows ILI screening appears to miss as many as 50% of true positive influenza cases. It should be noted that β_1 is even smaller than published estimates of ILI specificity for those with medical attendance [4, 35]. Both β_1 and α_2 are based on data for general populations with and without influenza and are not conditioned on any pre-defined medical state.

Figure 6 shows PPV and NPV based on equations (25) and (26). It is notable that during the course of strict entry screening at Narita airport, PPV remained smaller than 0.1% and NPV was greater than 82%. In other words, correct positive diagnoses were less than 0.1%, forcing 99.9% of ILI positive (but true negative) individuals to undergo unnecessary RDT. On the other hand, the proportion of correct negative results for the ILI negative individuals increased with time to close to 1. In practical terms, the use of ILI as the first screening option to identify febrile influenza cases is regarded as extremely costly.

6. **Discussion.** The data generating process for infectious diseases involves both the transmission dynamics and the assignment of infected individuals as either positive or negative by means of a clinical diagnostic procedure. In the present study, we propose a simple modeling approach to the joint quantification of these two aspects. As the starting point, we exploited exponential growth and a generalized logistic model to parameterize epidemic dynamics, thereby expressing both positive and negative test results as a function of calendar time. Our analysis of the interpandemic influenza A (H1N1) of 2007-08 gave a reproduction number, R, that appeared to be consistent with R for other interpandemic influenza in temperate zones. This framework has also helped quantify the very limited specificity of ILI and the low sensitivity of RDT in diagnosing influenza cases at an international border, suggesting that the screening procedure that begins with ILI is a very costly.

The biggest advantage of joint estimation is that the estimated diagnostic accuracy is not pre-conditioned on study populations that have some artificial inclusion criteria. Rather, the estimates are conditioned on population that has been included in the observed epidemic data. Given that the diagnostic accuracy of ILI is usually measured from epidemiological observation of a pre-defined population, the analysis of 2009 entry screening data is particularly highlighted as the methodological advantage. An additional advantage is that our method permits the precise quantification of the transmission dynamics. In Section 3 we have shown that relying only on test positive results tends to yield biased estimate of the epidemic growth rate, especially, if both the sensitivity and the specificity of a diagnostic test are low. With reference to interpandemic influenza in Mie, we have also shown that precise quantification of the epidemic dynamics appropriately captures the height of the epidemic peak.

An important practical implication drawn from the present study is that both test positives and test negatives (especially the latter) should be recorded during the empirical collection of data (routine surveillance). Previously, surveillance tended to record only those testing positive. We have shown that negative reporting facilitates the use of a simple statistical approach to extract much richer information. Although the proposed approach does not capture all influenza cases, our model allows the estimation of the *incidence* that diagnostic testing procedures aimed to detect (e.g., all the symptomatic influenza cases that underwent diagnostic testing).

An important disadvantage of our approach is in the parameterization of true negatives as a function of time. Although we assume that true negatives are independent of time, acting only as noise, this may not always be the case in all realistic settings. Two typical examples violate this assumption; (i) dual epidemics (an influenza epidemic coupled with another epidemic caused by a different pathogen) and (ii) an increase in people's awareness (and the increase in medical attendance as a function of time). To model both the true positives and true negatives as a function of time might make the model more vulnerable to their parametric assumptions. This limitation does not negate the proposed method but may be regarded purely as a data limitation. To address the involvement of a similar viral disease epidemic during an influenza epidemic, we have to account for the number of positive and negative test results for that viral disease in addition to the test results of influenza. Similarly, to account for time-dependency in medical attendance, we would have to measure the detailed time dependency of medical attendance in addition to the test results of influenza.

As another technical remark, it must be remembered that the successful joint estimation depends on the validity of assumed transmission dynamics. That is, the underlying transmission dynamics have to be correctly captured. For example, although we employed deterministic exponential growth (with an ad-hoc Poisson argument) in Section 3, that does not sufficiently account for demographic stochasticity, and thus, is only applicable to actual exponential growth phase with a large number of cases. In such a case, full stochastic model should replace the deterministic exponential formula. Instead of (7) and (11), we would have to use probability distribution of the number of cases on day t which is given as a solution of a pure birth process [1, 21] or birth-and-death process [36]. Since the likelihood function of the stochastic process calls for conditional measurement as a function of time, full likelihood may require us to account for more detailed relationship between the transmission dynamics and diagnostic accuracy [16]. Moreover, as discussed in Section 3, the successful convergence in such an initial phase will also depend on the relative size of true negative individuals as compared with the number of true positive individuals. Lastly, modeling practice of a pandemic during its very initial stage has posed a technical challenge in accounting for infection-age of imported cases [24].

As briefly mentioned in Section 2, there are other methods that use empirical data to estimate the incidence (or prevalence) of infectious diseases. Especially, the repeated measurement of the same individuals over time would greatly ease the relevant estimation framework. Another area to be explored is the use of multiple diagnostic testing procedures (e.g. ILI and RDT). Although our study of entry screening had only to account for the dependence of the series of test (and not for dependence between diagnostic accuracies), realistic applications frequently involve dependence in diagnostic accuracy between two or more tests. Regarding the quantification of the epidemic dynamics, the method may benefit from being extended to more rigorous approaches applied to a heterogeneously mixing population. We believe that, even without these future improvements, the present study convincingly emphasizes the usefulness of analyzing both test positive and test negative results, and offers an insight into the usefulness of joint quantification of the transmission dynamics and diagnostic accuracy.

REFERENCES

[1] N. T. J. Bailey, "The Elements of Stochastic Processes with Applications to the Natural Sciences," Wiley, New York, 1964.

[2] H. M. Babcock, L. R. Merz, E. R. Dubberke and V. J. Fraser, *Case-control study of clinical features of influenza in hospitalized patients,* Infect. Control. Hosp. Epidemiol., **29** (2008), 921–926.

[3] R. B. Banks, "Growth and Diffusion Phenomena: Mathematical Frameworks and Applications," Springer, New York, 1993.

[4] S. A. Call, M. A. Vollenweider, C. A. Hornung, D. L. Simel and W. P. McKinney, *Does this patient have influenza?* JAMA, **293** (2005), 987–997.

[5] G. Chowell, M. A. Miller and C. Viboud, *Seasonal influenza in the United States, France, and Australia: Transmission and prospects for control*, Epidemiol. Infect., **136** (2008), 852–864.

[6] B. J. Cowling, L. L. Lau, P. Wu, H. W. Wong, V. J. Fang, S. Riley and H. Nishiura, *Entry screening to delay local transmission of 2009 pandemic influenza A (H1N1)*, BMC Infect. Dis., **10** (2010), 82.

[7] C. Fraser, C. A. Donnelly, S. Cauchemez, W. P. Hanage, M. D. van Kerkhove, T. D. Hollingsworth, J. Griffin, R. F. Baggaley, H. E. Jenkins, E. J. Lyons, T. Jombart, W. R. Hinsley, N. C. Grassly, F. Balloux, A. C. Ghani, N.M. Ferguson, A. Rambaut, O.G. Pybus, H. Lopes-Gatell, C.M. Alpuche-Aranda, I.B. Chapela, E.P. Zavala, D. M. Guevara, F. Checchi, E. Garcia, S. Hugonnet, C. Roth and WHO Rapid Pandemic Assessment Collaboration, *Pandemic potential of a strain of influenza A (H1N1): Early findings*, Science, **324** (2009), 1557–1561.

[8] T. Garske, J. Legrand, C.A. Donnelly, H. Ward, S. Cauchemez, C. Fraser, N. M. Ferguson and A. C. Ghani, *Assessing the severity of the novel influenza A/H1N1 pandemic*, BMJ, **339** (2009), b2840.

[9] A. L. Greer, A. Tuite and D. N. Fisman, *Age, influenza pandemics and disease dynamics*, Epidemiol. Infect., (2010), in press.

[10] Y. H. Hsieh and C. W. Chen, *Turning points, reproduction number, and impact of climatological events for multi-wave dengue outbreaks*, Trop. Med. Int. Health, **14** (2009), 628–638.

[11] Y. H. Hsieh and S. Ma, *Intervention measures, turning point, and reproduction number for dengue, Singapore, 2005*, Am. J. Trop. Med. Hyg., **80** (2009), 66–71.

[12] A. C. Hurt, R. Alexander, J. Hibbert, N. Deed and I. G. Barr, *Performance of six influenza rapid tests in detecting human influenza in clinical specimens*, J. Clin. Virol., **39** (2007), 132–135.

[13] M. Lipsitch, S. Riley, S. Cauchemez, A. C. Ghani and N. M. Ferguson, *Managing and reducing uncertainty in an emerging influenza pandemic*, N. Engl. J. Med., **361** (2009), 112–115.

[14] Mie Infectious Disease Surveillance Center, "Situation of Influenza Epidemic from 2007-08," Mie Prefecture, Mie, Japan, 2008 (in Japanese; available from: http://www.kenkou.pref.mie.jp/).

[15] S. M. Moghadas, N. J. Pizzi, J. Wu and P. Yan, *Managing public health crises: The role of models in pandemic preparedness*, Influenza Other Respi. Viruses, **3** (2009), 75–79.

[16] N. J. Nagelkerke, R. N. Chunge and S. N. Kinoti, *Estimation of parasitic infection dynamics when detectability is imperfect*, Stat. Med., **9** (1990), 1211–1219.

[17] Narita Airport Quarantine Station, "Report on Responses to the Pandemic Influenza at Narita Airport Quarantine Station," Narita Airport Quarantine Station, Chiba, Japan, 2009 (in Japanese).

[18] H. Nishiura, *Case fatality ratio of pandemic influenza*, Lancet Infect. Dis., **10** (2010), 442–443.

[19] H. Nishiura, *The virulence of pandemic influenza A (H1N1) 2009: An epidemiological perspective on the case-fatality ratio*, Expert Rev. Respir. Med., **4** (2010), 329–338.

[20] H. Nishiura, *Time variatins in the generation time of an infectious disease: Implications for sampling to appropriately quantify transmission potential*, Math. Biosci. Eng., **7** (2010), 851–869.

[21] H. Nishiura, C. Castillo-Chavez, M. Safan and G. Chowell, *Transmission potential of the new influenza A(H1N1) virus and its age-specificity in Japan*, Euro. Surveill., **14** (2009), pii:19227.

[22] H. Nishiura, G. Chowell, M. Safan and C. Castillo-Chavez, *Pros and cons of estimating the reproduction number from early epidemic growth rate of influenza A (H1N1) 2009*, Theor. Biol. Med. Model., **7** (2010), 1.

[23] H. Nishiura, D. Klinkenberg, M. Roberts and J. A. Heesterbeek, *Early epidemiological assessment of the virulence of emerging infectious diseases: A case study of an influenza pandemic*, PLoS One, **4** (2009), e6852.

[24] H. Nishiura and M. G. Roberts, *Estimation of the reproduction number for 2009 pandemic influenza A(H1N1) in the presence of imported cases*, Euro. Surveill., **15** (2010), pii:19622.

[25] H. Nishiura, N. Wilson and M. G. Baker, *Quarantine for pandemic influenza control at the borders of small island nations*, BMC Infect. Dis., **9** (2009), 27.

[26] R. J. Pitman, B. S. Cooper, C. L. Trotter, N. J. Gay and W. J. Edmunds, *Entry screening for severe acute respiratory syndrome (SARS) or influenza: Policy evaluation*, BMJ, **331** (2005), 1242–1243.

[27] B. Pourbohloul, A. Ahued, B. Davoudi, R. Meza, L. A. Meyers, D. M. Skowronski, I. Villasenor, F. Galvan, P. Cravioto, D. J. Earn, J. Dushoff, D. Fisman, W. J. Edmunds, N. Hupert, S. V. Scarpino, J. Trujillo, M. Lutzow, J. Morales, A. Contreras, C. Chavez, D. M. Patrick and R. C. Brunham, *Initial human transmission dynamics of the pandemic (H1N1) 2009 virus in North America*, Influenza Other Respi. Viruses, **3** (2009), 215–222.

[28] A. M. Presanis, D. De Angelis, New York City Swine Flu Investigation Team, A. Hagy, C. Reed, S. Riley, B. S. Cooper, L. Finelli, P. Biedrzycki and M. Lipsitch, *The severity of pandemic H1N1 influenza in the United States, from April to July 2009: A Bayesian analysis*, PLoS Med., **6** (2009), e1000207.

[29] H. Rashid, S. Shafi, H. El Bashir, E. Haworth, Z. A. Memish, K. A. Ali and R. Booy, *Influenza and the Hajj: defining influenza-like illness clinically*, Int. J. Infect. Dis., **12** (2008), 102–103.

[30] W. J. Rogan and B. Gladen, *Estimating prevalence from the results of a screening test*, Am. J. Epidemiol., **107** (1978), 71–76.

[31] H. Sato, H. Nakada, R. Yamaguchi, S. Imoto, S. Miyano and M. Kami, *When should we intervene to control the 2009 influenza A(H1N1) pandemic?* Euro. Surveill., **15** (2010), pii:19455.

[32] L. Simonsen, M. J. Clarke, D.F. Stroup, G. D. Williamson, N. H. Arden and N. J. Cox, *A method for timely assessment of influenza-associated mortality in the United States*, Epidemiology, **8** (1997), 390–395.

[33] A. R. Tuite, A. L. Greer, M. Whelan, A. L. Winter, B. Lee, P. Yan, J. Wu, S. Moghadas, D. Buckeridge, B. Pourbohloul and D. N. Fisman, *Estimated epidemiologic parameters and morbidity associated with pandemic H1N1 influenza*, CMAJ, **182** (2010), 131–136.

[34] United Nations System Influenza Coordination (UNSIC) Asia-Pacific Regional Hub, "Simulation Exercises on Influenza Pandemic Responses in the Asia-Pacific Region," UNSIC Asia-Pacific Regional Hub, Bangkok, 2008.

[35] A. M. v d Hoeven, M. Scholing, P. C. Wever, R. Fijnheer, M. Hermans and P. M. Schneeberger, *Lack of discriminating signs and symptoms in clinical diagnosis of influenza of patients admitted to the hospital*, Infection, **35** (2007), 65–68.

[36] J. van den Broek and H. Nishiura, *Using epidemic prevalence data to jointly estimate reproduction and removal*, Ann. Appl. Stat., **3** (2009), 1505–1520.

[37] M. D. van Kerkhove, T. Asikainen, N. G. Becker, S. Bjorge, J. C. Desenclos, T. dos Santos, C. Fraser, G. M. Leung, M. Lipsitch, I. M. Longini, E. S. McBryde, C. E. Roth, D. K. Shay, D. J. Smith, J. Wallinga, P. J. White, N. M. Ferguson, S. Riley, WHO Informal Network for Mathematical Modelling for Pandemic Influenza H1N1 2009 (Working Group on Data Needs), *Studies needed to address public health challenges of the 2009 H1N1 influenza pandemic: Insights from modeling*, PLoS Med., **7** (2010), e1000275.

[38] J. T. Wittes, T. Colton and V. W. Sidel, *Capture-recapture methods for assessing the completeness of case ascertainment when using multiple information sources*, J. Chronic. Dis., **27** (1974), 25–36.

[39] T. Yanagawa and B. Gladen, *Estimating disease rates from a diagnostic test*, Am. J. Epidemiol., **119** (1984), 1015–1023.

[40] Y. Yang, J. D. Sugimoto, M. E. Halloran, N. E. Basta, D. L. Chao, L. Matrajt, G. Potter, E. Kenah and I. M. Longini, *The transmissibility and control of pandemic influenza A (H1N1) virus*, Science, **326** (2009), 729–733.