RESEARCH

Estimation of the Time-dependent Reproduction Number of Influenza A(H1N1)pdm2009 in South Korea

Yunjeong Lee^{1†}, Dong Han Lee^{2†}, Hee-Dae Kwon³, Changsoo Kim⁴ and Jeehyun Lee^{5*}

*Correspondence:

ezhyun@yonsei.ac.kr

⁵Department of Mathematics,
Yonsei University, 50, Yonsei-ro,
03722, Seoul, South Korea

Full list of author information is
available at the end of the article

[†]Equal contributor

Abstract

Background: The reproduction number is one of the most crucial parameters in determining disease dynamics, providing a summary measure of the transmission potential. However, estimating this value is particularly challenging owing to the characteristics of epidemic data, including non-reproducibility and incompleteness.

Methods: In this study, we propose mathematical models with different population structures; each of these models can produce data on the number of cases of the influenza A(H1N1)pdm09 epidemic in South Korea. These structured models incorporating the heterogeneity of age and region are used to estimate the time-dependent effective reproduction numbers. Subsequently, the age- and region-specific reproduction numbers are also computed to analyze the differences illustrated in the incidence data.

Results: The basic SIR fails to provide a reasonable estimation of the reproduction numbers. The estimated values demonstrate a large variation and remains outside of the feasible range for the influenza, regardless of the time period for data. Real-time estimation using age- and region-structured models demonstrated that the effective reproduction number rose sharply during mid-October when the number of patients increased dramatically. The reproduction number fell below unity at the end of October and stayed lower than unity indicating that the epidemic starts decreasing, which is consistent with the incidence data.

Conclusions: Numerical results reveal that the introduction of heterogeneity into the population to represent the general characteristics of dynamics is essential for the robust estimation of parameters.

Keywords: Influenza; Reproduction number; Antiviral agents

Background

The reproduction number is defined as the average number of secondary cases generated by a typical primary case. It is a measure of the transmission potential associated with the contact rate, duration of infectivity, and probability of transmission per contact. The maximum reproduction number is attained when an infectious person is introduced into a totally susceptible population and is called the basic reproduction number, R_0 . Various approaches such as the exponential growth rate of infections during the early epidemic stage, model-based schemes, and maximum-likelihood estimations have been used to analyze this number [1–4].

When an infection spreads throughout a population, the time-dependent effective reproduction number, R_t , is often more useful for assessing the transmission potential throughout a pandemic, especially during the period with the highest level of activity. Real-time estimation continues to track the number of secondary infections caused by a single infective, providing a quantitative measure of the time evolution of the epidemic force. Cruz-Pacheco et al. demonstrated the manner in which sanitary measures reduce the prevalence of an infected population [1]. Estimates of the reproduction number were shown to decrease from 1.4–1.5 initially to 1.1–1.2 later in the summer, which was most likely because of the vacation period and the seasonality of influenza transmission [5]. In addition to capturing temporal dynamics, it is important to consider heterogeneous patterns of the transmission. It is well known that school-age children are disproportionately responsible for influenza transmissions. Estimates of the age-specific reproduction number help with our understanding of the role of each group in the transmission dynamics and with devising effective targeting mitigation strategies.

If all incident cases could be traced back to their index cases, estimating the reproduction number would simply be a matter of counting the number of secondary cases. However, with most epidemics, only the epidemic curve is observed, and there is no available information regarding who infected whom. To appropriately estimate the reproduction number from the influenza outbreak data, it is essential that the selected model capture the underlying dynamics embedded in the data. The objective of this study is to estimate the reproduction numbers based on the incidence data.

¹ When the World Health Organization announced the emergence of influenza¹
²A(H1N1)pdm09 (pH1N1) in 2009 [6], the first probable patient in South Korea²
³was identified on April 28. A total of 763,759 confirmed cases, of which 270 were³
⁴fatal, were reported by the end of August 2010 [7]. During the initial epidemic⁴
⁵phase, the main control measure was containment through quarantine and isola-⁵
⁶tion. Surveillance programs in schools and medical facilities were implemented, and⁶
⁷all confirmed cases were investigated. However, when community outbreaks were⁷
⁸detected in June, the intervention policy switched from containment to mitigation,⁸
⁹including vaccination and antiviral prescription. Vaccination was started on Octo-⁹
¹⁰ber 27, 2009, and 12.7 million people were vaccinated by the end of August 2010.¹⁰
¹¹Before August 20, antiviral agents were prescribed to patients with acute febrile¹¹
¹²respiratory illness (AFRI) and who had a history of travel abroad or contact with¹²
¹³a confirmed patient. However, when the number of community-acquired cases in-¹³
¹⁴creased, antiviral agents were prescribed to patients with AFRI symptoms. ¹⁴

¹⁵ According to the database, 3,087,788 courses of antiviral agents were prescribed¹⁵
¹⁶from August 21, 2009 to April 30, 2010. The daily number of incident patients¹⁶
¹⁷was estimated based on the amount of prescribed antiviral agents (Refer to [8] for¹⁷
¹⁸details). Figure 1 shows the temporal incidence distribution of pH1N1 in South¹⁸
¹⁹Korea. The amount of antiviral agents prescribed and the number of incident pa-¹⁹
²⁰tients soared from mid-October, reached its peak at the end of October, and started²⁰
²¹declining in mid-November. ²¹

²² Demographic and regional characteristics are illustrated in Table 1 and Figure²²
²³2. The incidence rate is higher in children and students than in other age group²³
²⁴individuals (Table 1). The rate is higher in urban areas than in rural areas and is²⁴
²⁵the highest in the national capital and the south-eastern region (Figure 2). ²⁵

²⁶ In this paper, two different structured models are proposed to estimate reproduc-²⁶
²⁷tion numbers on the basis of the epidemic curve. We begin by introducing a basic²⁷
²⁸SIR model to describe a single outbreak and build age- and region-structured mod-²⁸
²⁹els by incorporating population heterogeneity. Numerical simulations are conducted²⁹
³⁰to analyze the impact of terminal time and the effect of heterogeneous structures³⁰
³¹on the estimation of parameters. Finally, the proposed models are applied to the³¹
³²2009 incidence data of novel pH1N1 in South Korea to compute the age- and region-³²
³³specific reproduction numbers. ³³

Methods

Basic SIR model

We consider the standard SIR (Susceptible-Infected-Recovered) model to represent single-outbreak influenza dynamics. The model classifies individuals into three key compartments: susceptible, infected, and recovered. The nonlinear system of differential equations describing the dynamics is given by the following equation.

$$S'(t) = -\beta S(t)I(t)$$

$$I'(t) = \beta S(t)I(t) - \gamma I(t)$$

$$R'(t) = \gamma I(t).$$

The state variables $S(t)$, $I(t)$, and $R(t)$ denote the number of individuals who are susceptible, infected, and recovered, respectively, at time t . The number of contact events sufficient for transmitting an infection during unit time per individual is βN based on the mass action incidence. Infective individuals leave the compartment at the recovery rate γ , thereby acquiring immunity to the disease. We can drop the equation for $R(t)$ as it has no effect on the dynamics of others and hence is determined once $S(t)$ and $I(t)$ are known. Based on the SIR model, we have the time-dependent net reproduction number $R_t = \beta S(t)/\gamma$, which quantifies the level of transmission at time t . Note that R_t is the per-infective rate at which new infections occur within the average duration of infection at time t .

Age-structured model

An age-structured model is employed to estimate the reproduction number of pH1N1 because the transmission rate is higher in preschool and schoolchildren than in other age group individuals, in general. We consider a subgroup SIR model where the population is divided into n_a age groups with different transmission dynamics. We denote the number of susceptible and infected individuals within the i^{th} age group by S_i and I_i , respectively. Let β_{ij} refer to the transmission from the j^{th} age group to the i^{th} age group, and $\beta = [\beta_{ij}]$ denote the transmission matrix, also

¹known as Who-Acquires-Infection-From-Whom matrix. Putting these elements to-¹
²gether, we have the following system of differential equations. ²

$$\begin{aligned} S'_i(t) &= - \sum_{j=1}^{n_a} \beta_{ij} S_i(t) I_j(t) \\ I'_i(t) &= \sum_{j=1}^{n_a} \beta_{ij} S_i(t) I_j(t) - \gamma_i I_i(t). \end{aligned} \quad (1)$$

⁸ In a general structured model of the form (1) with n_a distinct classes, n_a^2 trans-⁸
⁹mission terms are required. However, one transmission term is available at most for ⁹
¹⁰each class. The typical way to address this lack of specificity is to constrain the ¹⁰
¹¹structure of the transmission matrix and/or to use prior knowledge of social mixing ¹¹
¹²behavior. For an age-structured model, we assume that the transmission rates are ¹²
¹³proportional to the rates of social contact, which can be estimated from contact ¹³
¹⁴patterns. A large multi-country population-based survey conducted in Europe as a ¹⁴
¹⁵part of the POLYMOD [9] enables us to implement this approach. The transmis- ¹⁵
¹⁶sion is modeled as the product of the contact rate in the survey and an age-specific ¹⁶
¹⁷proportionality factor to account for characteristics related to susceptibility and ¹⁷
¹⁸infectiousness, which are not captured by contact rates. This leads to ¹⁸

$$\beta_{ij} = \begin{cases} q_i c_{ij} & i = j \\ \sigma c_{ij} & i \neq j \end{cases}$$

²³where c_{ij} is the contact rate and q_i and σ are proportionality factors. ²³

²⁴ Based on the age-structured SIR model (1), the reproduction number can be ²⁴
²⁵calculated by following Driessche and Watmough [10]. It is the spectral radius of ²⁵
²⁶the next generation matrix M where ²⁶

$$M_{ij} = \frac{\beta_{ij} S_i(t)}{\gamma_i}.$$

³⁰ The details are given in Appendix. ³⁰

³¹Region-structured model ³¹

³² The second mechanism incorporates a heterogeneous population based on regions to ³²
³³account for the wave of the pH1N1 pandemic. We denote the number of susceptible ³³

¹and infected individuals within the i^{th} region by S_i and I_i , respectively. Let β_{ij} ¹
²refer to the transmission from the j^{th} subgroup to the i^{th} subgroup and $\beta = [\beta_{ij}]$ ²
³denote the transmission matrix. In the same manner as the age-structured model,³
⁴we have the following system of differential equations.⁴

$$\begin{aligned} S'_i(t) &= - \sum_{j=1}^{n_r} \beta_{ij} S_i(t) I_j(t) \\ I'_i(t) &= \sum_{j=1}^{n_r} \beta_{ij} S_i(t) I_j(t) - \gamma_i I_i(t). \end{aligned} \quad (2)$$

¹⁰We assume that transmission rates between distinct regions in the region-¹⁰
¹¹structured model can be expressed as the frequency of transportations multiplied¹¹
¹²by a region-specific proportionality factor. The transportation information was ex-¹²
¹³tracted from the highway portal site and Kakao map for number of buses and¹³
¹⁴highway traffic, respectively [11, 12]. Let the number of buses and highway traffic¹⁴
¹⁵from region j to region i be denoted by $w_{i,j}$ and $W_{i,j}$, respectively. Note that w ¹⁵
¹⁶is symmetric because the bus route is circular, although W is not necessarily. The¹⁶
¹⁷transmission rate can be written as¹⁷

$$\beta_{ij} = \begin{cases} q_i & i = j \\ q_i \sigma_l w_{ij} + q_i \sigma_g W_{ij} & i \neq j \end{cases}$$

²³where q_i is the proportionality factor, and σ_l are σ_g can be chosen such that they²³
²⁴balance the weight between different types of transportations.²⁴

²⁵The same argument as that presented in the age-structured model gives the ex-²⁵
²⁶pression of the effective reproduction number R_t , which is the spectral radius of the²⁶
²⁷following next generation matrix²⁷

$$\left[\frac{\beta_{ij} S_i(t)}{\gamma_i} \right].$$

³⁰Study subjects and parameter estimation³⁰

³¹Study subjects were patients who were prescribed antiviral agents from the national³¹
³²stockpile from August 21, 2009 to April 30, 2010. Because of mandatory antiviral³²
³³agent management program during study period, all patients who were prescribed³³

antiviral agent were included in this study. The data employed to estimate the parameters are the daily number of incident patients in figure 1. It was estimated based on the aggregation of prescribed antiviral agents from deidentified database. This study was approved by the Institutional Review Board (IRB) of Yonsei University Health System. Since this study used retrospective data and the study subjects were anonymized, the IRB waived the requirement for written consent from the patients.

Our goal is to estimate the optimal parameters that provide the states that are best fit to the given data. This section briefly reviews the parameter estimation technique of the least squares method. In general, parameter estimation is conducted by minimizing the cost function, which measures the difference between the model prediction and observation. Let θ be the parameter set and y_j be the measurement data at time t_j . We recast the mathematical model as

$$\mathbf{x}'(t) = \mathbf{g}(t, \mathbf{x}(t), \theta),$$

and assume a statistical model for measurement of the form

$$y_j = f(t_j; \theta) + \varepsilon_j, \quad j = 1, \dots, N$$

where $f(t_j; \theta)$ is the model prediction at time t_j with parameter θ and the measurement error $\varepsilon_j \sim \mathcal{N}(0, a^2)$. The least squares estimator can be obtained by minimizing the following cost function over the given parameter space Ω_θ [13]:

$$\sum_{j=1}^N [y_j - f(t_j; \theta)]^T [y_j - f(t_j; \theta)]$$

The parameter sets to be estimated for the basic SIR model, age-structured model and region-structured model, are

$$\theta_{\text{basic}} = \{\beta, \gamma\},$$

$$\theta_{\text{age}} = \{q_i, \sigma, \gamma_i \text{ for } i = 1, 2, \dots, n_a\},$$

$$\theta_{\text{region}} = \{q_i, \sigma_l, \sigma_g, \gamma_i \text{ for } i = 1, 2, \dots, n_r\}$$

where n_a and n_r denote the number of age groups and regions, respectively.

¹Results

²Time-dependent reproduction number ³

⁴We illustrate the proposed methodology and investigate its performance by apply-
⁵ing it to 2009 incidence data of pH1N1 in South Korea. This section presents the
⁶results of the estimation obtained by applying the least squares method to basic
⁷SIR, age- and region-structured models. In each experiment, data with different
⁸time periods by varying the terminal time are tested to determine the earliest stages
⁹of the epidemic sufficient to provide a reasonable estimation. Figure 3 displays the
¹⁰predicted incidence based on the basic SIR model compared with the observed
¹¹data. Predictions using data only during the initial growth phase cannot effectively
¹²exhibit the dynamics and substantially overestimate the spread of the infection.
¹³The results of simulation improved after the peak of the epidemic, and the wave is
¹⁴roughly generated at a later stage. However, the simple SIR model does not provide
¹⁵a reasonable estimation of the reproduction numbers. The estimated values demon-
¹⁶strate a large variation and remains outside of the feasible range for the influenza,
¹⁷regardless of the time period for data. The plausible reason for this involves the
¹⁸model assumptions that are too simple to capture the underlying mechanisms.

¹⁹For the age-structured SIR model (1), the total population is split into 14 sub-
²⁰groups of 5-year age bands and one with 70 years and older (i.e., 0-4, 5-9, ..., 65-69,
²¹70⁺). This incorporates a heterogeneous population into the model in order to re-
²²flect different transmission rates in each age group. In figure 4, the outbreaks are
²³simulated using data during various time periods in the same manner as mentioned
²⁴above. Results of both the models show similar trends as long as the terminal time
²⁵is earlier than mid-November when the gentle growth begins during the decline
²⁶stage. Additionally, as the growth begins to decline, the age-structured SIR model
²⁷fits the incidence data better than the basic SIR model. The reproduction number
²⁸starts increasing in early October, peaks at 2.5 on October 17, and then decreases
²⁹to unity at the end of October. Real-time estimation demonstrated that the effec-
³⁰tive reproduction number rose sharply during mid-October when the number of
³¹patients increased dramatically. The reproduction number fell below unity at the
³²end of October and stayed lower than unity indicating that the epidemic starts
³³decreasing, which is consistent with the incidence data. Incorporation of the age

structure allows for robust estimation of reproduction numbers, as observed on the bottom of figure 4.

The general characteristics of regional difference led us to consider a second type of heterogeneity in the model. The nation is split into 252 in the region-structured model (2), where the transmission rates are implemented based on transportation patterns. Figure 5 compares the predicted cases based on the region-structured SIR model with the observed data over the course of the epidemic. As it was discussed in the previous experiment, it is not earlier than the epidemic peak for estimation to start adjusting to outbreak data. Since this outbreak, the incidence data is well described in the form of the characteristic exponential rise, turnover, and decline pattern predicted by the process model. The estimated time-dependent effective reproduction numbers are illustrated in figure 5, which demonstrates a patterns similar to that obtained using the age-structured SIR model in figure 4.

Age-specific and region-specific reproduction numbers

It is widely known that the transmission is considerably different among various age groups. We also observe from the pH1N1 epidemic data that the incidence rate is higher in children and students than in other age groups (Table 1). Estimates of the age-specific reproduction number help in clarifying the role of each age group in the transmission dynamics and in suggesting guidelines for effective targeting intervention strategies. The estimated age-specific reproduction numbers are displayed in figure 6. The result is closely related to the cumulative incidence for each age group because it is often the contact rate within the same age group is higher than with other groups.

The incidence rate is higher in urban areas than in rural areas, and the highest in the national capital and the south-eastern region, as shown in figure 2. We estimated the region-specific reproduction number and observed that it is more than two in some areas and less than one in the others (Figure 7). This is consistent with regions having larger cumulative incidence with a similar argument regarding contact patterns to age-specific cases.

Discussion

An estimation of reproduction numbers is crucial because it provides a measure of the transmission potential when an infection is spreading throughout a population.

¹The reproductive numbers in the early phase of the pandemic have been estimated¹
²in several countries with different settings, yielding median 1.46 and range 1.0–3.6²
³[14]. Many of these studies focused on cases confirmed in the early stage of the³
⁴pandemic. Because laboratory tests focused on severe cases and there are possible⁴
⁵changes in laboratory testing and notification rates, the number of confirmed cases⁵
⁶does not necessarily represent the underlying epidemic. It also does not reflect the⁶
⁷dynamics during the period of the highest level of activity, which is the winter in⁷
⁸temperate climates. Some studies used the number of cases from sentinel surveillance⁸
⁹that is much less than the actual number of influenza patients. It is necessary to⁹
¹⁰estimate the reproductive number using the number of all the patients throughout¹⁰
¹¹a pandemic, including the period with the highest level of activity. In this study,¹¹
¹²the reproductive number of was estimated based on the national data of incidence¹²
¹³deduced from antiviral agent prescription in South Korea during the pandemic. ¹³

¹⁴ We discussed parameter estimation methodologies based on deterministic SIR¹⁴
¹⁵models incorporating population heterogeneity. Age-structured and region-structured¹⁵
¹⁶models were employed to describe the underlying epidemic process. The proposed¹⁶
¹⁷mechanisms were applied to influenza A(H1N1)pdm09 in South Korea to compute¹⁷
¹⁸the time-dependent effective reproduction numbers. We found that the introduc-¹⁸
¹⁹tion of heterogeneity into the population and sufficient data to represent general¹⁹
²⁰characteristics of dynamics are essential to the robust estimation of parameters.²⁰
²¹Real-time estimation showed that the reproduction number started increasing in²¹
²²early October, peaked at 2.5 on October 17, and then decreased to unity at the²²
²³end of October. The effective number rose sharply during the mid-October when²³
²⁴the number of patients increased dramatically. The reproduction number fell below²⁴
²⁵unity at the end of October and remained lower than unity, indicating that the²⁵
²⁶epidemic starts decreasing, which is consistent with the incidence data. ²⁶

²⁷ Subsequently, age-specific and region-specific basic reproduction numbers were²⁷
²⁸estimated to account for the differences of incidence. We observe from the pH1N1²⁸
²⁹epidemic data that the incidence rate is higher in children and students than in other²⁹
³⁰age groups. The estimated age-specific reproduction numbers agree with the cumu-³⁰
³¹lative incidence for each age group because the mixing is assortative. The incidence³¹
³²rate is higher in urban areas than in rural areas, highest in the national capital and³²
³³in the south-eastern region. We estimated the region-specific reproduction number³³

¹whose trend is similar to the number of cases in each region. Estimates of the age-¹
²specific and region-specific reproduction number help to predict the transmission²
³dynamics, and to suggest guidelines for effective targeting intervention strategies. ³
⁴This study has both limitations and strengths. The first limitation is that the⁴
⁵number of cases is estimated from the amount of prescribed antiviral agents as-⁵
⁶suming the time lag between symptom onset and antiviral agent prescription, the⁶
⁷proportion of prescription and the proportion of pH1N1 confirmation among AFRI⁷
⁸patients. Another limitation is that vaccination is not considered in the model.⁸
⁹However, the effect of vaccination on the transmission of pH1N1 may have been in-⁹
¹⁰significant because the vaccination for general group was initiated in January 2010.¹⁰
¹¹On the contrary, the present research has its strengths compared to previous studies.¹¹
¹²The reproduction number was estimated based on the national level antiviral agent¹²
¹³prescription data in South Korea throughout the pandemic including the period¹³
¹⁴of the highest level of activity. The real-time estimation incorporating population¹⁴
¹⁵structures can be used to predict the disease dynamics, thereby providing guidelines¹⁵
¹⁶for the optimal implementation of preventive measures, such as school closing and¹⁶
¹⁷distribution of antiviral agents. ¹⁷

¹⁸Conclusions ¹⁹

²⁰In this study, we proposed age- and region-structured SIR models to estimate the²⁰
²¹time-dependent reproduction number of Influenza A(H1N1)pdm2009 in South Ko-²¹
²²rea. The basic SIR fails to provide a reasonable estimation of the effective reproduc-²²
²³tion numbers, possibly due to the model assumptions that are too simple to capture²³
²⁴the underlying mechanisms. Numerical results reveal that the introduction of het-²⁴
²⁵erogeneity into the population to represent the general characteristics of dynamics²⁵
²⁶is essential for the robust estimation of parameters. ²⁶

²⁷Appendix ²⁸

²⁹The reproduction number for the age-structured SIR model (1), can be calculated²⁹
³⁰following the approach of Driessche and Watmough [10]. Let $\mathcal{F}_i$ be the new infections³⁰
³¹and $\mathcal{V}_i$ be the transitions of i^{th} compartment, then ³¹

$$\mathcal{F}_i = S_i \left(q_i c_{ii} I_i + \sum_{j \neq i} \sigma c_{ij} I_j \right) \quad (1) \quad \supseteq 32$$

¹and 1

² 2

³ $\mathcal{V}_i = \gamma I_i.$ (2)³

⁴ 4

⁵for $i = 1, \dots, n_a.$ 5

⁶ Subsequently, the derivatives of $\mathcal{F} = [\mathcal{F}_i]$ and $\mathcal{V} = [\mathcal{V}_i]$ are 6

$$\begin{aligned} & \begin{matrix} 7 \\ 8 \\ 9 \\ 10 \\ 11 \end{matrix} \quad F = \begin{bmatrix} S_1 q_1 c_{11} & S_1 \sigma c_{12} & \cdots & S_1 \sigma c_{1n_a} \\ S_2 \sigma c_{21} & S_1 q_2 c_{22} & \cdots & S_2 \sigma c_{2n_a} \\ \vdots & \vdots & \ddots & \vdots \\ S_{n_a} \sigma c_{n_a 1} & S_{n_a} \sigma c_{n_a 2} & \cdots & S_{n_a} q_{n_a} c_{n_a n_a} \end{bmatrix}, \quad V = \begin{bmatrix} \gamma & 0 & \cdots & 0 \\ 0 & \gamma & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \gamma \end{bmatrix}, \quad \begin{matrix} 7 \\ 8 \\ 9 \\ 10 \\ 11 \end{matrix} \end{aligned}$$

¹²respectively, and the next generation matrix is 12

$$\begin{aligned} & \begin{matrix} 13 \\ 14 \\ 15 \\ 16 \\ 17 \end{matrix} \quad FV^{-1} = \frac{1}{\gamma} \begin{bmatrix} S_1 q_1 c_{11} & S_1 \sigma c_{12} & \cdots & S_1 \sigma c_{1n_a} \\ S_2 \sigma c_{21} & S_1 q_2 c_{22} & \cdots & S_2 \sigma c_{2n_a} \\ \vdots & \vdots & \ddots & \vdots \\ S_{n_a} \sigma c_{n_a 1} & S_{n_a} \sigma c_{n_a 2} & \cdots & S_{n_a} q_{n_a} c_{n_a n_a} \end{bmatrix}. \quad \begin{matrix} 13 \\ 14 \\ 15 \\ 16 \\ 17 \end{matrix} \end{aligned} \quad (3)^{15}$$

¹⁸Thus, the reproduction number is the spectral radius of FV^{-1} and the age-specific¹⁸

¹⁹reproduction number is the column sum of FV^{-1} corresponding to the age of in-¹⁹

²⁰terest. 20

²¹ 21

²²Abbreviations 22

²³ R_0 : Basic reproduction number; R_t : Time-dependent effective reproduction number; pH1N1: Influenza 23

A(H1N1)pdm09; AFRI: Acute febrile respiratory illness; SIR: Susceptible-Infected-Recovered; IRB: Institutional 24

²⁴Review Board; LSM: Least squares method 24

²⁵Declarations 25

²⁶ Ethics approval and consent to participate 26

²⁷ This study was approved by the Institutional Review Board (IRB) of Yonsei University Health System. 27

²⁸ Consent for publication 28

²⁹ Not applicable. 29

³⁰ Availability of data and materials 30

³¹ The datasets used and/or analysed during the current study are available from the corresponding author on 31
³² reasonable request. 32

³³ Competing interests 33

The authors declare that they have no competing interests.

- 1 Funding 1
- 2 The work of Jeehyun Lee was supported by NRF-2015R1A5A1009350 and NRF-2016R1A2B4014178. The funders 2
- 3 had a major role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. 3
- 4 Authors' contributions 4
- 5 Conceived and designed the experiments: DHL JL. Analyzed the data: DHL CK. Performed the experiments: YL 5
- 6 HDK. Wrote the paper: YL DHL JL. Interpreted the data: YL DHL HDK. Contributed to draft and revised the 6
- 7 manuscript for content: All. Read and approved the manuscript: All. 7
- 8 Acknowledgements 8
- 9 Not applicable. 9
- 10 Authors' information 10
- 11 Not applicable. 11
- 12 Author details 12
- 13 Department of Computational Science and Engineering, Yonsei University, 50, Yonsei-ro, 03722, Seoul, South 13
- 14 Korea. ²Korea Centers for Disease Control and Prevention, 187, Osongsaengmyeong 2-ro, 28159, Cheongju-si, 14
- 15 South Korea. ³Department of Mathematics, Inha University, 100, Inha-ro, 22212, Incheon, South Korea. 15
- 16 Department of Preventive Medicine and Public Health, Severance Hospital, Yonsei University College of Medicine, 16
- 17 50-1, Yonsei-ro, 03722, Seoul, South Korea. ⁵Department of Mathematics, Yonsei University, 50, Yonsei-ro, 03722, 17
- 18 Seoul, South Korea. 18
- 19 References 19
- 20 1. Cruz-Pacheco G, Duran L, Esteva L, Minzoni AA, López-Cervantes M, Panayotaras P, et al. Modelling of the 20
- 21 influenza A(H1N1) outbreak in Mexico City, April-May 2009, with control sanitary measures. Euro Surveill. 21
- 22 2009;14(26). 22
- 23 2. Fraser C, Donnelly CA, Cauchemez S, Hanage WP, Van Kerkhove MD, Hollingsworth TD, et al. Pandemic 23
- 24 potential of a strain of influenza A (H1N1): early findings. Science. 2009;324(5934):1557–61. 24
- 25 3. Nishiura H, Chowell G, Safan M, Castillo-Chavez C. Pros and cons of estimating the reproduction number from 25
- 26 early epidemic growth rate of influenza A (H1N1) 2009. Theor Biol Med Model. 2010;7(1). 26
- 27 4. Yang Y, Sugimoto JD, Halloran ME, Basta NE, Chao DL, Matrajt L, et al. The transmissibility and control of 27
- 28 pandemic influenza A (H1N1) virus. Science. 2009;326(5953):729–33. 28
- 29 5. Cowling BJ, Lau MSY, Ho LM, Chuang SK, Tsang T, Liu SH, et al. The effective reproduction number of 29
- 30 pandemic influenza: prospective estimation. Epidemiology. 2010;21(6):842–6. 30
- 31 6. World Health Organization. Influenza-like illness in the United States and Mexico, 24 April 2009; 2009. 31
- 32 Available from: https://www.who.int/csr/don/2009_04_24/en/. Accessed 4 Sep 2019. 32
- 33 7. Korea Centers for Disease Control and Prevention. Analysis of reported pandemic influenza (A/H1N1 2009) 33
- 34 virus infections in Korea - From April, 2009 through August, 2010. Public Health Weekly Report 2010a. 34
- 35 2010;3:637–42. 35
- 36 8. Lee DH. Estimation of the reproduction number of pandemic influenza A(H1N1) 2009 from national data of 36
- 37 antiviral agent prescription. Yonsei University, Seoul, South Korea; 2011. 37
- 38 9. Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, et al. Social contacts and mixing patterns 38
- 39 relevant to the spread of infectious diseases. PLoS Med. 2008 03;5(3). 39
- 40 10. van den Driessche P, Watmough J. Reproduction numbers and sub-threshold endemic equilibria for 40
- 41 compartmental models of disease transmission. Math Biosci. 2002;180:29–48. 41
- 42 11. Kakao map. Kakao Corporation, Jeju-si, South Korea; 2017. Available from: <https://map.kakao.com/>. 42
- 43 Updated 25 February 2019; Accessed 15 March 2017. 43
- 44 12. Highway public data portal. Korea Expressway Corporation, Gimcheon-si, South Korea; c2015. Available from: 44
- 45 <http://data.ex.co.kr/portal/fdwn/view?type=TCS&num=39&requestfrom=dataset#>. Accessed 15 March 45
- 46 2017. 46

13. Banks HT, Hu S, Thompson WC. Modeling and inverse problems in the presence of uncertainty. New York: Chapman and Hall/CRC; 2014.
14. Biggerstaff M, Cauchemez S, Reed C, Gambhir M, Finelli L. Estimates of the reproduction number for seasonal, pandemic, and zoonotic influenza: a systematic review of the literature. BMC Infect Dis. 2014 Sep;14:480.

Figure Legends

Figure 1 Daily antiviral agent prescription and incident patients from September 1, 2009 to March 30, 2010 [8]. The adjusted antiviral agent prescription (dotted line) is obtained by tuning the discrepancy in the amount of prescribed antiviral agents between weekdays and weekends using cubic splines. The estimated number of incident patients (solid line) is approximated by reflecting the time lag between symptom onset and antiviral agent prescription, proportion of antiviral agent prescription, number of pH1N1 confirmations among AFRI patients, and number of asymptomatic cases.

Figure 2 Cumulative incidence by region.

Figure 3 Comparison of pH1N1 incidence data with predictions of least squares method (LSM) using the basic SIR model (top and bottom left): The red dots show the number of new cases per day and the blue line presents the predicted number of cases. Terminal time of data used for estimation is displayed by a black dotted vertical bar. In each figure, the end of time period was set at October 24, 2009, November 14, 2009, and March 30, 2010. Estimation of the time-dependent effective reproduction number using the basic SIR model (bottom right): The black dots show the estimated value using data during the time period at different terminal times.

Figure 4 Comparison of pH1N1 incidence data with predictions of least squares method (LSM) using age-structured SIR model (top and bottom left): The red dots show the number of new cases per day, and the blue line presents the predicted number of cases. Terminal time of data used for estimation is displayed by a black dotted vertical bar. In each figure, the end of time period was set at October 24, 2009, November 14, 2009, and March 30, 2010. Estimation of the time-dependent effective reproduction number using the age-structured SIR model (bottom right): The black dots show the estimated values at different terminal times.

Figure 5 Comparison of pH1N1 incidence data with the predictions of least squares method (LSM) obtained using the region-structured SIR model (top and bottom left): The red dots show the number of new cases per day, and the blue line presents the predicted number of cases. The terminal time of data used for estimation is displayed by a black dotted vertical bar. In each figure, the end of time period was set at October 24, 2009, November 14, 2009, and March 30, 2010. The estimation of the time-dependent effective reproduction number using the region-structured SIR model (bottom right): The black dots show the estimated value at different terminal times.

Tables

Figure 6 Age-specific reproduction number (left) and cumulative incidence for each age group (right).

Figure 7 Region-specific reproduction number (left) and cumulative incidence for each region (right).

Table 1 Cumulative incidence by age.

Age	Population size	Total number of cases	Incidence rate(%)
0-4	2273634	917793	40
5-9	2671496	1084094	41
10-14	3284134	920937	28
15-19	3368818	781238	23
20-24	3153001	305403	10
25-29	4001090	341940	9
30-34	3856438	290354	8
35-39	4466962	303670	7
40-44	4244225	220743	5
45-49	4321262	164735	4
50-54	3761143	149067	4
55-59	2611705	97571	4
60-64	2117041	73540	3
65-69	1870279	66033	4
70+	3306607	100728	3