RESEARCH

Estimation of the Reproduction Number of Influenza A(H1N1)pdm2009 in South Korea Using Heterogeneous Models

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.

Results: 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 reproduction numbers at various terminal times. Subsequently, the age- and region-specific reproduction numbers are also computed to analyze the differences illustrated in 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

¹reproduction number, R_0 . Various approaches such as the exponential growth rate¹
²of infections during the early epidemic stage, model-based schemes, and maximum-²
³likelihood estimations have been used to analyze this number [1–4].³

⁴ When an infection spreads throughout a population, the time-dependent effective⁴
⁵reproduction number, R_t , is often more useful for assessing the transmission po-⁵
⁶tential throughout a pandemic, especially during the period with the highest level⁶
⁷of activity. Real-time estimation continues to track the number of secondary in-⁷
⁸fections caused by a single infective, providing a quantitative measure of the time⁸
⁹evolution of the epidemic force. Cruz-Pacheco et al. demonstrated the manner in⁹
¹⁰which sanitary measures reduce the prevalence of an infected population [1]. Esti-¹⁰
¹¹mates of the reproduction number were shown to decrease from 1.4-1.5 initially to¹¹
¹²1.1-1.2 later in the summer, which was most likely because of the vacation period¹²
¹³and the seasonality of influenza transmission [5]. In addition to capturing temporal¹³
¹⁴dynamics, it is important to consider heterogeneous patterns of the transmission.¹⁴
¹⁵It is well known that school-age children are disproportionately responsible for in-¹⁵
¹⁶fluenza transmissions. Estimates of the age-specific reproduction number help with¹⁶
¹⁷our understanding of the role of each group in the transmission dynamics and with¹⁷
¹⁸devising effective targeting mitigation strategies.¹⁸

¹⁹ If all incident cases could be traced back to their index cases, estimating the¹⁹
²⁰reproduction number would simply be a matter of counting the number of secondary²⁰
²¹cases. However, with most epidemics, only the epidemic curve is observed, and²¹
²²there is no available information regarding who infected whom. To appropriately²²
²³estimate the reproduction number from the influenza outbreak data, it is essential²³
²⁴that the selected model capture the underlying dynamics embedded in the data.²⁴
²⁵The objective of this study is to estimate the reproduction numbers based on the²⁵
²⁶incidence data.²⁶

²⁷ When the World Health Organization announced the emergence of influenza²⁷
²⁸A(H1N1)pdm09 (pH1N1) in 2009 [6], the first probable patient in South Korea²⁸
²⁹was identified on April 28. A total of 763,759 confirmed cases, of which 270 were²⁹
³⁰fatal, were reported by the end of August 2010 [7]. During the initial epidemic³⁰
³¹phase, the main control measure was containment through quarantine and isola-³¹
³²tion. Surveillance programs in schools and medical facilities were implemented, and³²
³³all confirmed cases were investigated. However, when community outbreaks were³³

¹detected in June, the intervention policy switched from containment to mitigation,¹
²including vaccination and antiviral prescription. Vaccination was started on Octo-²
³ber 27, 2009, and 12.7 million people were vaccinated by the end of August 2010.³
⁴Before August 20, antiviral agents were prescribed to patients with acute febrile⁴
⁵respiratory illness (AFRI) and who had a history of travel abroad or contact with⁵
⁶a confirmed patient. However, when the number of community-acquired cases in-⁶
⁷creased, antiviral agents were prescribed to patients with AFRI symptoms.⁷

⁸ According to the database, 3,087,788 courses of antiviral agents were prescribed⁸
⁹from August 21, 2009 to April 30, 2010. The daily number of incident patients⁹
¹⁰was estimated based on the amount of prescribed antiviral agents (Refer to [8] for¹⁰
¹¹details). Figure 1 shows the temporal incidence distribution of pH1N1 in South¹¹
¹²Korea. The amount of antiviral agents prescribed and the number of incident pa-¹²
¹³tients soared from mid-October, reached its peak at the end of October, and started¹³
¹⁴declining in mid-November.¹⁴

¹⁵ Demographic and regional characteristics are illustrated in Table 1 and Figure¹⁵
¹⁶2. The incidence rate is higher in children and students than in other age group¹⁶
¹⁷individuals (Table 1). The rate is higher in urban areas than in rural areas and is¹⁷
¹⁸the highest in the national capital and the south-eastern region (Figure 2).¹⁸

¹⁹ In this paper, two different structured models are proposed to estimate reproduc-¹⁹
²⁰tion numbers on the basis of the epidemic curve. We begin by introducing a basic²⁰
²¹SIR model to describe a single outbreak and build age- and region-structured mod-²¹
²²els by incorporating population heterogeneity. Numerical simulations are conducted²²
²³to analyze the impact of terminal time and the effect of heterogeneous structures²³
²⁴on the estimation of parameters. Finally, the proposed models are applied to the²⁴
²⁵2009 incidence data of novel pH1N1 in South Korea to compute the age- and region-²⁵
²⁶specific reproduction numbers.²⁶

²⁸**Methods**²⁸

²⁹**Basic SIR model**²⁹

³⁰ We consider the standard SIR model to represent single-outbreak influenza dy-³⁰
³¹namics. The model classifies individuals into three key compartments: susceptible,³¹
³²infected, and recovered. The nonlinear system of differential equations describing³²
³³

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

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

In a general structured model of the form (1) with n_a distinct classes, n_a^2 transmission terms are required. However, one transmission term is available at most for

each class. The typical way to address this lack of specificity is to constrain the¹
 structure of the transmission matrix and/or to use prior knowledge of social mixing²
 behavior. For an age-structured model, we assume that the transmission rates are³
 proportional to the rates of social contact, which can be estimated from contact⁴
 patterns. A large multi-country population-based survey conducted in Europe as a⁵
 part of the POLYMOD [9] enables us to implement this approach. The transmis-⁶
 sion is modeled as the product of the contact rate in the survey and an age-specific⁷
 proportionality factor to account for characteristics related to susceptibility and⁸
 infectiousness, which are not captured by contact rates. This leads to⁹

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

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

Based on the age-structured SIR model (1), the reproduction number can be¹⁵
 calculated by following Driessche and Watmough [10]. It is the spectral radius of¹⁶
 the next generation matrix M where¹⁷

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

The details are given in Appendix.

Region-structured model

The second mechanism incorporates a heterogeneous population based on regions to²⁴
 account for the wave of the pH1N1 pandemic. We denote the number of susceptible²⁵
 and infected individuals within the i^{th} region by S_i and I_i , respectively. Let β_{ij} ²⁶
 refer to the transmission from the j^{th} subgroup to the i^{th} subgroup and $\beta = [\beta_{ij}]$ ²⁷
 denote the transmission matrix. In the same manner as the age-structured model,²⁸
 we have the following system of differential equations.

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

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

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

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

The same argument as that presented in the age-structured model gives the expression of the effective reproduction number R_t , which is the spectral radius of the following next generation matrix

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

Study subjects and parameter estimation

Study subjects were patients who were prescribed antiviral agents from the national stockpile from August 21, 2009 to April 30, 2010. Because of mandatory antiviral agent management program during study period, all patients who were prescribed 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 time points t_j ($j = 1, \dots, N$) are uniformly distributed with daily time step. The data vector y_j ($j = 1, \dots, N$) denotes the number of cases at time t_j . It is a scalar for basic SIR model, but it is a vector structured by age and region for age-stratified and region-stratified models, respectively. For example, the y_j is a column vector of length 15 for the age-structured model. We recast the mathematical model as

$$x'(t) = g(t, x(t), \theta),$$

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.

Results

Time-dependent reproduction number

We illustrate the proposed methodology and investigate its performance by applying it to 2009 incidence data of pH1N1 in South Korea. This section presents the results of the estimation obtained by applying the least squares method to basic SIR, age- and region-structured models. In each experiment, data with different

time periods by varying the terminal time are tested to determine the earliest stage¹ of the epidemic sufficient to provide a reasonable estimation. Figure 3 displays the² predicted incidence based on the basic SIR model compared with the observed³ data. Predictions using data only during the initial growth phase cannot effectively⁴ exhibit the dynamics and substantially overestimate the spread of the infection.⁵ The results of simulation improved after the peak of the epidemic, and the wave is⁶ roughly generated at a later stage. However, the simple SIR model does not provide⁷ a reasonable estimation of the reproduction numbers. The estimated values demon-⁸ strate a large variation and remains outside of the feasible range for the influenza,⁹ regardless of the time period for data. The plausible reason for this involves the¹⁰ model assumptions that are too simple to capture the underlying mechanisms.¹¹

For the age-structured SIR model (1), the total population is split into 14 sub-¹² groups of 5-year age bands and one with 70 years and older (i.e., 0-4, 5-9, ..., 65-69,¹³ 70⁺). This incorporates a heterogeneous population into the model in order to re-¹⁴ flect different transmission rates in each age group. In figure 4, the outbreaks are¹⁵ simulated using data during various time periods in the same manner as mentioned¹⁶ above. Results of both the models show similar trends as long as the terminal time¹⁷ is earlier than mid-November when the gentle growth begins during the decline¹⁸ stage. Additionally, as the growth begins to decline, the age-structured SIR model¹⁹ fits the incidence data better than the basic SIR model. The reproduction number²⁰ starts increasing in early October, peaks at 2.5 on October 17, and then decreases²¹ to unity at the end of October. Real-time estimation demonstrated that the effec-²² tive reproduction number rose sharply during mid-October when the number of²³ patients increased dramatically. The reproduction number fell below unity at the²⁴ end of October and stayed lower than unity indicating that the epidemic starts²⁵ decreasing, which is consistent with the incidence data. In the age-stratified model,²⁶ heterogeneity was incorporated by WAIFW matrix where the transmission was as-²⁷ sumed to be proportional to the contacts. The effective contacts were measured by²⁸ POLYMOD contact survey, which showed a clear evidence for an age-dependency²⁹ in contact patterns. Taking heterogeneous mixing into the model enabled better³⁰ description of the dynamics, because the trend in behavior was consistent with the³¹ demographic characteristics of cases (as shown in Table 1). Estimated parameters³² are possible indicators to determine the feasibility of models. Incorporation of the³³

age structure allows for robust estimation of reproduction numbers as observed on the bottom of figure 4, while the basic SIR model provides estimated values beyond the reasonable range with severe fluctuation.

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.

¹Discussion

²An estimation of reproduction numbers is crucial because it provides a measure of²
³the transmission potential when an infection is spreading throughout a population.³
⁴The reproductive numbers in the early phase of the pandemic have been estimated⁴
⁵in several countries with different settings, yielding median 1.46 and range 1.0–3.6⁵
⁶[14]. Many of these studies focused on cases confirmed in the early stage of the⁶
⁷pandemic. Because laboratory tests focused on severe cases and there are possible⁷
⁸changes in laboratory testing and notification rates, the number of confirmed cases⁸
⁹does not necessarily represent the underlying epidemic. It also does not reflect the⁹
¹⁰dynamics during the period of the highest level of activity, which is the winter in¹⁰
¹¹temperate climates. Some studies used the number of cases from sentinel surveillance¹¹
¹²that is much less than the actual number of influenza patients. It is necessary to¹²
¹³estimate the reproductive number using the number of all the patients throughout¹³
¹⁴a pandemic, including the period with the highest level of activity. In this study,¹⁴
¹⁵the reproductive number of was estimated based on the national data of incidence¹⁵
¹⁶deduced from antiviral agent prescription in South Korea during the pandemic.¹⁶

¹⁷We discussed parameter estimation methodologies based on deterministic SIR¹⁷
¹⁸models incorporating population heterogeneity. Age-structured and region-structured¹⁸
¹⁹models were employed to describe the underlying epidemic process. The proposed¹⁹
²⁰mechanisms were applied to influenza A(H1N1)pdm09 in South Korea to compute²⁰
²¹the time-dependent effective reproduction numbers. We found that the introduc-²¹
²²tion of heterogeneity into the population and sufficient data to represent general²²
²³characteristics of dynamics are essential to the robust estimation of parameters.²³
²⁴Real-time estimation showed that the reproduction number started increasing in²⁴
²⁵early October, peaked at 2.5 on October 17, and then decreased to unity at the²⁵
²⁶end of October. The effective number rose sharply during the mid-October when²⁶
²⁷the number of patients increased dramatically. The reproduction number fell below²⁷
²⁸unity at the end of October and remained lower than unity, indicating that the²⁸
²⁹epidemic starts decreasing, which is consistent with the incidence data.²⁹

³⁰Subsequently, age-specific and region-specific basic reproduction numbers were³⁰
³¹estimated to account for the differences of incidence. We observe from the pH1N1³¹
³²epidemic data that the incidence rate is higher in children and students than in other³²
³³age groups. The estimated age-specific reproduction numbers agree with the cumu-³³

relative incidence for each age group because the mixing is assortative. The incidence rate is higher in urban areas than in rural areas, highest in the national capital and in the south-eastern region. We estimated the region-specific reproduction number whose trend is similar to the number of cases in each region. Estimates of the age-specific and region-specific reproduction number help to predict the transmission dynamics, and to suggest guidelines for effective targeting intervention strategies.

This study has both limitations and strengths. The first limitation is that the number of cases is estimated from the amount of prescribed antiviral agents assuming the time lag between symptom onset and antiviral agent prescription, the proportion of prescription and the proportion of pH1N1 confirmation among AFRIPatients. The effective contacts were employed from POLYMOD contact survey, which possibly yields discrepancy in mixing pattern of Korea. The structured models may not be enough to account for some underlying mechanism of dynamics due to those factors. However, they provide feasible estimates of reproduction number despite the limitations, which was not possible using simple SIR model. We will be able to improve the outcome as we gather more information, because additional knowledge is required to achieve a better result. Another limitation is that vaccination is not considered in the model. However, the effect of vaccination on the transmission of pH1N1 may have been insignificant because the vaccination for general group was initiated in January 2010. On the contrary, the present research has its strengths compared to previous studies. The reproduction number was estimated based on the national level antiviral agent prescription data in South Korea throughout the pandemic including the period of the highest level of activity. The real-time estimation incorporating population structures can be used to predict the disease dynamics, thereby providing guidelines for the optimal implementation of preventive measures, such as school closing and distribution of antiviral agents.

Appendix

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

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

and

2

$$\mathcal{V}_i = \gamma I_i. \quad (2)$$

4

for $i = 1, \dots, n_a$.

5

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

6

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

respectively, and the next generation matrix is

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

Thus, the reproduction number is the spectral radius of FV^{-1} and the age-specific reproduction number is the column sum of FV^{-1} corresponding to the age of interest.

Ethics approval and consent to participate

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

Consent for publication

Not applicable.

Availability of data and materials

The dataset used and/or analysed during the current study were employed from the literature [8].

Competing interests

The authors declare that they have no competing interests.

Funding

Not applicable.

Authors' contributions

DH and J designed the work, and C contributed to collect data. Y carried out data processing and numerical implementation, and J, HD, and Y interpreted the results. J wrote the manuscript and DH, C, HD and Y assisted with drafting the work. All authors read and approved the final manuscript.

Acknowledgements

The work of Jeehyun Lee was supported by NRF-2015R1A5A1009350 and HG20C0003030020.

¹Authors' information

²Not applicable.

³Author details

¹Department of Computational Science and Engineering, Yonsei University, 50, Yonsei-ro, 03722, Seoul, South

⁴Korea. ²Korea Centers for Disease Control and Prevention, 187, Osongsaengmyeong 2-ro, 28159, Cheongju-si,

⁵South Korea. ³Department of Mathematics, Inha University, 100, Inha-ro, 22212, Incheon, South Korea.

⁴Department of Preventive Medicine and Public Health, Severance Hospital, Yonsei University College of Medicine,

650-1, Yonsei-ro, 03722, Seoul, South Korea. ⁵Department of Mathematics, Yonsei University, 50, Yonsei-ro, 03722,

Seoul, South Korea.

References

- ¹ Cruz-Pacheco G, Duran L, Esteva L, Minzoni AA, López-Cervantes M, Panayotaros P, et al. Modelling of the
influenza A(H1N1) outbreak in Mexico City, April-May 2009, with control sanitary measures. *Euro Surveill.*
2009;14(26).
- ² Fraser C, Donnelly CA, Cauchemez S, Hanage WP, Van Kerkhove MD, Hollingsworth TD, et al. Pandemic
potential of a strain of influenza A (H1N1): early findings. *Science.* 2009;324(5934):1557–61.
- ³ Nishiura H, Chowell G, Safan M, Castillo-Chavez C. Pros and cons of estimating the reproduction number from
early epidemic growth rate of influenza A (H1N1) 2009. *Theor Biol Med Model.* 2010;7(1).
- ⁴ Yang Y, Sugimoto JD, Halloran ME, Basta NE, Chao DL, Matrajt L, et al. The transmissibility and control of
pandemic influenza A (H1N1) virus. *Science.* 2009;326(5953):729–33.
- ⁵ Cowling BJ, Lau MSY, Ho LM, Chuang SK, Tsang T, Liu SH, et al. The effective reproduction number of
pandemic influenza: prospective estimation. *Epidemiology.* 2010;21(6):842–6.
- ⁶ World Health Organization. Influenza-like illness in the United States and Mexico, 24 April 2009; 2009.
Available from: https://www.who.int/csr/don/2009_04_24/en/. Accessed 4 Sep 2019.
- ⁷ Korea Centers for Disease Control and Prevention. Analysis of reported pandemic influenza (A/H1N1 2009)
virus infections in Korea - From April, 2009 through August, 2010. *Public Health Weekly Report* 2010a.
2010;3:637–42.
- ⁸ Lee DH. Estimation of the reproduction number of pandemic influenza A(H1N1) 2009 from national data of
antiviral agent prescription. Yonsei University, Seoul, South Korea; 2011.
- ⁹ Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, et al. Social contacts and mixing patterns
relevant to the spread of infectious diseases. *PLoS Med.* 2008 03;5(3).
- ¹⁰ van den Driessche P, Watmough J. Reproduction numbers and sub-threshold endemic equilibria for
compartmental models of disease transmission. *Math Biosci.* 2002;180:29–48.
- ¹¹ Kakao map. Kakao Corporation, Jeju-si, South Korea; 2017. Available from: <https://map.kakao.com/>.
Updated 25 February 2019; Accessed 15 March 2017.
- ¹² Highway public data portal. Korea Expressway Corporation, Gimcheon-si, South Korea; c2015. Available from:
<http://data.ex.co.kr/portal/fdwn/view?type=TCS&num=39&requestfrom=dataset#>. Accessed 15 March
2017.
- 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.

Figures

Figure 1 Daily antiviral agent prescription (dotted line) and incident patients (solid line) from
September 1, 2009 to March 30, 2010. The incidence data used for model calibration was
imported from literature and refer to [8] for details.

Tables

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. The bottom right shows the reproduction number at different time point t using data until that specific time t . First, the parameters were estimated fitting the number of cases up to the time t . Then, the reproduction number at time t was calculated with the estimated parameters and $S(t)$. Therefore, not only the reproduction number but also parameters are updated as the more observations are available.

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. The reproduction number shown on the bottom right was obtained similarly to the basic SIR model. The only difference was to obtain the reproduction number at time t by calculating the spectral radius of the next generation matrix with estimated parameters and the number of susceptible in the i^{th} age-group at time t , $S_i(t)$.

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 reproduction number shown on the bottom right was obtained similarly to the age-structured model, except that $S_i(t)$ denoted the number of susceptible in the i^{th} region at time t .

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).

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

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