



ORIGINAL ARTICLE

Modeling for Estimating Influenza Patients from ILI Surveillance Data in Korea

Joo-Sun Lee^a, Sun-Hee Park^a, Jin-Woong Moon^a, Jacob Lee^b,
Yong Gyu Park^c, Yong Kyun Roh^{d,*}

^aDivision of Infectious Disease Surveillance, Korea Centers for Disease Control and Prevention, Osong, Korea.

^bDepartment of Internal Medicine, College of Medicine, Hallym University, Chuncheon, Korea.

^cDepartment of Medical Life Science, College of Medicine, The Catholic University of Korea, Seoul, Korea.

^dDepartment of Family Medicine, College of Medicine, Hallym University, Chuncheon, Korea.

Received: May 30, 2011
Revised: July 12, 2011
Accepted: July 18, 2011

KEYWORDS:

influenza patients,
influenza surveillance
system,
Korea,
prediction model

Abstract

Objective: Prediction of influenza incidence among outpatients from an influenza surveillance system is important for public influenza strategy.

Methods: We developed two influenza prediction models through influenza surveillance data of the Korea Centers for Disease Control and Prevention (each year, each province and metropolitan city; total reported patients with influenza-like illness stratified by age) for 6 years from 2005 to 2010 and disease-specific data (influenza code J09-J11, monthly number of influenza patients, total number of outpatients and hospital visits) from the Health Insurance Review and Assessment service.

Results: Incidence of influenza in each area, year, and month was estimated from our prediction models, which were validated by simulation processes. For example, in November 2009, Seoul and Joenbuk, the final number of influenza patients calculated by prediction models A and B underestimated actual reported cases by 64 and 833 patients, respectively, in Seoul and 6 and 9 patients, respectively, in Joenbuk. R-square demonstrated that prediction model A was more suitable than model B for estimating the number of influenza patients.

Conclusion: Our prediction models from the influenza surveillance system could estimate the nationwide incidence of influenza. This prediction will provide important basic data for national quarantine activities and distributing medical resources in future pandemics.

*Corresponding author.

E-mail: rohyk@hallym.ac.kr, rohyk@korea.com

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

Influenza remains a global concern; estimates show that annual epidemics may cause <5 million severe cases and 500,000 deaths worldwide [1].

We established the Korean Influenza Surveillance Scheme (KISS) to monitor outbreaks of influenza-like illness (ILI) and detect new influenza virus strains in 2000 [2]. Surveillance of ILI in KISS is based on reports made by private sentinel physicians including pediatricians, internists, and general practitioners and physicians in county public health centers. Every Tuesday these physicians report the number of patients with ILI and the total number of patients who visited during the previous week via an internet reporting system to the Korea Centers for Disease Control and Prevention (KCDC). ILI is defined as fever (>38°C) with cough or sore throat. KCDC designated one clinic per 50,000 persons as sentinel sites participating in KISS; 820 sentinel clinics participated in reporting ILI in 2011 [3].

Data collected from ILI surveillance are used to calculate national outbreaks each year and advise vaccination strategy—notably in 2009 influenza

A (H1N1) pandemic [4–6]. Despite these efforts, KISS has some limitation in that it cannot accurately estimate the nationwide number of ILI patients with data collected only from sentinel physicians, because it is not a population-based surveillance.

The objective of the present study was to model the number of nationwide influenza patients based on ILI surveillance data and to analyze the usefulness of our model to prevent influenza outbreak.

2. Materials and Methods

We collected ILI surveillance data of KCDC (each year, each province and metropolitan city, total number of patients reporting ILI stratified by age) over 6 years from 2005 to 2010, as well as disease-specific data (influenza code J09-J11, monthly number of influenza patients, total number of outpatients and hospital visits) from the Health Insurance Review and Assessment (HIRA) service.

We performed statistical analyses using SAS software version 9.1 (SAS Institute). First, we estimated monthly the number of influenza patients and hospital

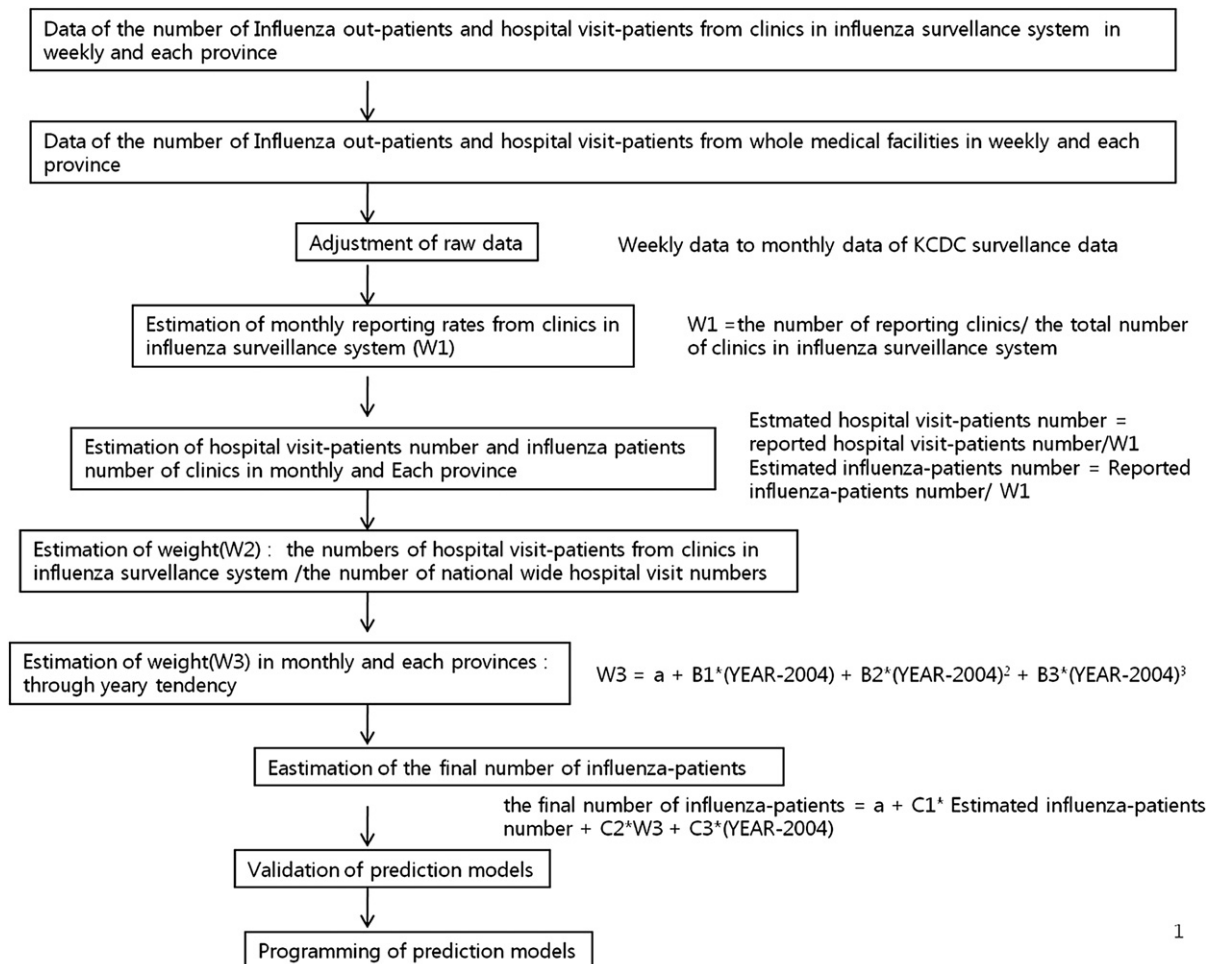


Figure 1. Development process of prediction model A.

visits in each province with estimated monthly reporting rates (W1). Second, we estimated the weight of scale of sentinel clinics (W2) compared with national clinics using numbers of hospital visit patients from clinics in the influenza surveillance system and that of national hospital visits. To estimate W2 in advance, we calculated fixed W2 with 3rd-order polynomial regression, because we did not know the number of national hospital visits (model A) [7–9]. Using fixed W2 after 2012, it will be estimated <1 in most regions. On the other hand, we also used W2 as the mean estimated number of hospital visits and influenza patients over 7 years by region (model B) [10,11]. Third, we estimated W3 as the weight in each province through yearly tendency. Finally, we obtained the final number of influenza patients by the following equations (Figure 1):

Model A

$$a + C1 * \text{Estimated number of influenza patients} + C2 * W3 + C3 (\text{year}-2004)$$

Model B

$$a + C1 * \text{Estimated number of influenza patients} + C2 * (\text{year}-2004)$$

3. Results

We developed two prediction models and explored final number of influenza patients by region, monthly using these models.

For example, in Seoul in November 2009, we obtained a reporting rate of 0.50 (324 of 648 total sentinel sites). The estimated hospital visit number was

calculated to be 213,982 from the reported total hospital visit number divided by the abovementioned reporting rate (0.50). Also, the estimated influenza patient number was calculated to be 7983 divided by above-obtained reporting rate (0.50). Next, we calculated W3 (13.57) as the value of 3rd-order polynomial regression. In the last step, we obtained 320,503 as the final number of influenza patients using model A.

In model B, the 0.50 calculated reporting rate and 15,966 estimated hospital visit number were the same as for model A. Next, we calculated W3 (17.33) as mean estimated hospital visits and estimated the number of influenza patients over 7 years, and finally obtained 319,734 using model B.

The final number of influenza patients calculated by prediction models A and B was 320,503 and 319,734, respectively, whereas according to HIRA it was 320,567 during pandemic influenza season (Figure 2). Therefore prediction models A and B underestimated the incidence of influenza by 64 and 833 influenza patients according to HIRA.

To compare differences between the numbers of predicted and real influenza patients, we applied standardized residual techniques in which no difference was demonstrated if the value range was from -2 to $+2$. The standardized residual was 1.00 in validation of estimation value, demonstrating no difference.

As a second example, in Joenbuk in November 2009 the number of ILI patients predicted by models A and B was 68,610 and 68,607 whereas that by HIRA was 68,616 (Figure 3). Again, the standardized residual was 1.00 as was demonstrated for the above example using data from Seoul.

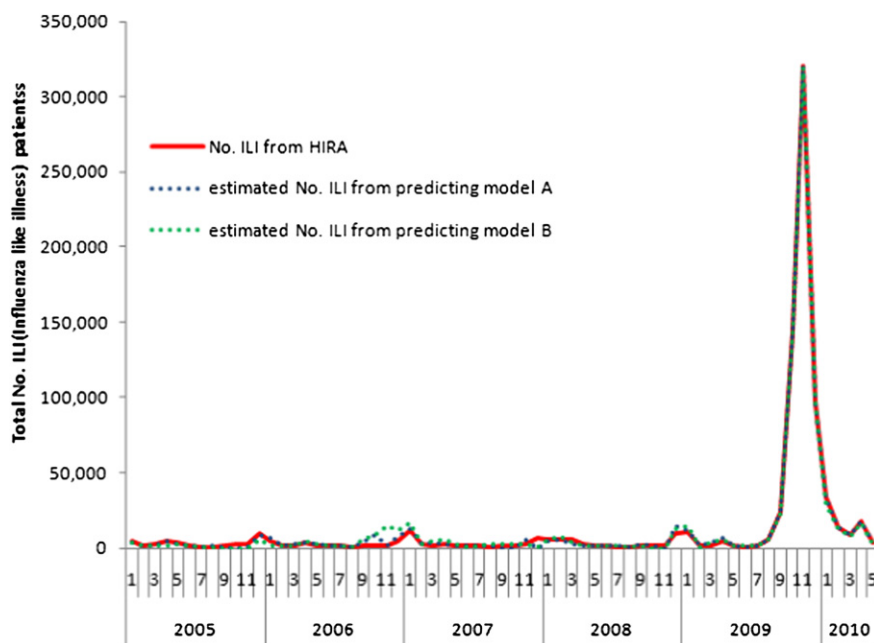


Figure 2. Estimated number of ILI from prediction model in Seoul.

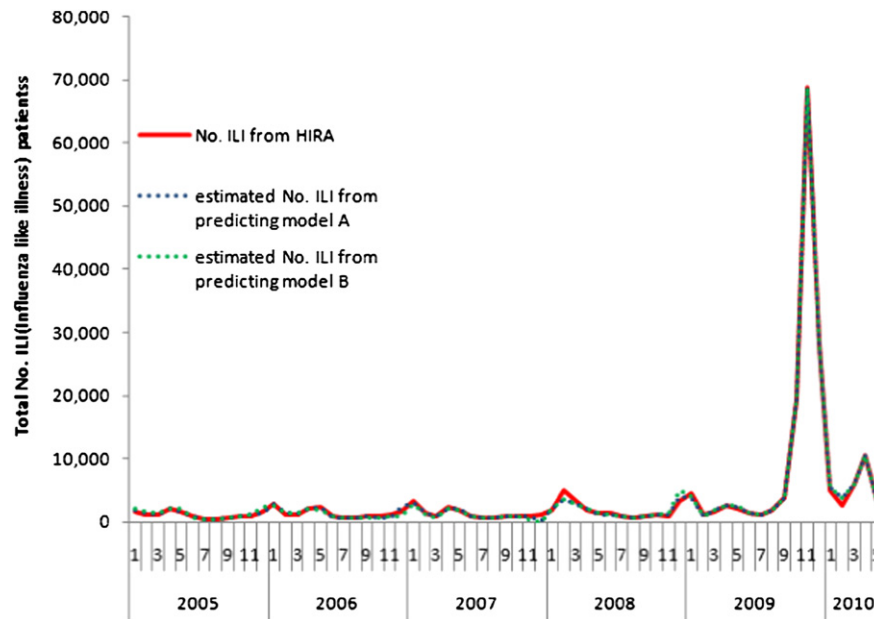


Figure 3. Estimated number of ILI from prediction model in Joenbuk.

Finally, we attempted to validate our model using root mean square error (RMSE) and R-square [12]. The R-square value of model A was higher than that of model B in Seoul (model A: 0.9998; B: 0.9967) and Joenbuk (model A: 0.9999; model B: 0.9998).

4. Discussion

We developed prediction models that were able to estimate nationwide numbers of influenza patients using data of ILI patients and hospital visit patients collected from sentinel sites. Because we know the number of reporting sentinel sites, the total number of sentinel sites, and number of patients and reported hospital visits by region monthly, we can calculate the total number of influenza patients by applying these data to models A and B.

In the above examples, compared versus real influenza patients, that predicted by models from ILI surveillance data exhibited no difference according to standardized residual 1.00. However, validation of our models using data on Seoul and Joenbuk suggested that model A was better as an influenza-estimating model. As mentioned above, as W_3 will be calculated <1 after 2012 because 3rd-order polynomial regression equation is applied in model A; we constantly have to revise this model.

Although we collect and analyze ILI data weekly, we will be able to estimate the number of influenza-patients only on a monthly basis because this model uses ILI surveillance data that are changed monthly. Therefore our models will need further refinement to obtain the number of nationwide influenza patients weekly.

Although our models have certain limitations, the number of nationwide influenza patients they predict

will provide important basic data for national quarantine activities and distributing medical resources, including influenza vaccine, admission beds, and so on, in future pandemics [13–15].

Although we developed only equations to estimate numbers of nationwide influenza patients in this study, it would be more useful to build a program to handle data automatically, simply, and in real time.

Acknowledgements

This study was supported by an intramural grant (No. 4838-304-260-00) of Korea Centers Disease Control and Prevention.

References

1. World Health Organizations (WHO). Influenza (seasonal). Available from: <http://www.who.int/demicentre/factsheets/fs11/en/>; 2009 [accessed January 2009].
2. Lee JS, Shin KC, Na BK, et al. Influenza surveillance in Korea: establishment and first results of an epidemiological and virological surveillance scheme. *Epidemiol Infect* 2007 Oct;135(7): 1117–23.
3. Kim JH, Yoo HS, Lee JS, et al. The spread of pandemic H1N1 2009 by age and region and the comparison among monitoring tools. *J Korean Med Sci* 2010 Jul;25(7):1109–12.
4. Flemming DM, Elliot AJ. Lessons from 40 years' surveillance of influenza in England and Wales. *Epidemiol Infect* 2008 Jul; 136(7):866–75.
5. CDC. Interim pre-pandemic planning guidance: community strategy for pandemic influenza mitigation in the United States. [monograph on the Internet]. Atlanta: Centers for Disease Control and Prevention. Available at: <http://www.flu.gov/professional/community/mitigation.pdf>; 2007 [Date accessed: 11 December 2010].

6. The ANZIC influenza investigators. Critical care services and 2009 H1N1 influenza in Australia and New Zealand. *N Engl J Med* 2009;361(20):1925–34.
7. Montgomery DC, Peck EA. Introduction to linear regression analysis. New York: John Wiley & Sons; 1982. p. 181–211.
8. Draper NR, Smith H. Applied regression analysis. New York: John Wiley & Sons; 1998. p. 251–271.
9. Belsley DA, Kuh E, Welsch RE. Regression diagnostics. New York: John Wiley & Sons; 1980. p. 152–173.
10. Box GP, Jenkins GM. Time series analysis. San Francisco: Holden-Day; 1970. p. 46–166.
11. Montgomery DC, Johnson LA. Forecasting and time series analysis. New York: McGraw-Hill Company; 1976. p. 75–95.
12. Cook RD, Weisberg S. Residuals and influence in regression. New York: Chapman and Hall; 1982. p. 101–112.
13. Balcan D, Colizza V, Singer AC, et al. Modeling the critical care demand and antibiotics resources needed during the Fall 2009 wave of influenza A(H1N1) pandemic. *PLoS Curr Influenza* 2009;7:1133–45.
14. Colizza V, Barrat A, Barthelemy M, et al. Modeling the worldwide spread of pandemic influenza: baseline case and containment interventions. *Plos medicine* 2007;4:95–110.
15. Ajelli M, Merler S, Pugliese A, et al. Model predictions and evaluation of possible control strategies for the 2009 A/H1N1v influenza pandemic in Italy. *Epidemiol Infect* 2011;139: 68–79.