



OPEN

Suggestion of a simpler and faster influenza-like illness surveillance system using 2014–2018 claims data in Korea

HeeKyoung Choi^{1,2,4}, Won Suk Choi^{3,4} & Euna Han¹✉

Influenza is an important public health concern. We propose a new real-time influenza-like illness (ILI) surveillance system that utilizes a nationwide prospective drug utilization monitoring in Korea. We defined ILI-related claims as outpatient claims that contain both antipyretic and antitussive agents and calculated the weekly rate of ILI-related claims, which was compared to weekly ILI rates from clinical sentinel surveillance data during 2014–2018. We performed a cross-correlation analysis using Pearson's correlation, time-series analysis to explore actual correlations after removing any dubious correlations due to underlying non-stationarity in both data sets. We used the moving epidemic method (MEM) to estimate an absolute threshold to designate potential influenza epidemics for the weeks with incidence rates above the threshold. We observed a strong correlation between the two surveillance systems each season. The absolute thresholds for the 4-years were 84.64 and 86.19 cases per 1000claims for claims data and 12.27 and 16.82 per 1000 patients for sentinel data. The epidemic patterns were more similar in the 2016–2017 and 2017–2018 seasons than the 2014–2015 and 2015–2016 seasons. ILI claims data can be loaded to a drug utilization review system in Korea to make an influenza surveillance system.

Influenza, which can cause epidemics and pandemics through antigenic shift, and localized outbreaks through antigenic drift¹, is one of the most important infectious diseases in public health². Therefore, we need a national surveillance system to monitor and respond to an influenza epidemic. In Korea, such a system is currently based on a clinical sentinel surveillance system, a laboratory sentinel surveillance system, and a hospitalization and mortality monitoring system, all of which use information collected from a restricted number of selected outpatient clinics^{3,4}. This traditional influenza surveillance system is considered the gold standard for relevant studies and public interventions^{5–7}. The clinical sentinel surveillance system reports the proportion of influenza-like illness (ILI) visits among weekly outpatients through the voluntary participation of 200 local clinics nationwide. This system has been well used in Korea for a long time⁸. However, due to the small number of participating clinics in each region, calculating the ILI rate by region or age is limited. In addition, as pediatric clinics account for half of the 200 surveillance clinics⁸, it is possible that the entire epidemiology cannot be sufficiently reflected. As the surveillance system relies on voluntary reporting, there is a concern of under-reporting. Also, the degree of under-reporting may differ depending on the characteristics of the clinic because case reporting can be affected by the patient volume and the physician's compliance⁹.

Patient visit, data submission, analysis of weekly ILI rates and public announcement take one to two weeks⁵, and such a time lag from actual peak to alarm is inevitable. The external validity of these reports has not been determined. Such caveats are widely recognized globally, and an alternative system using prescription drug sales^{10,11}, emergency department use¹², or school absenteeism data¹³ has been proposed. More recently, research has attempted to predict the incidence of ILI in real-time using novel data from Google or other social media^{14,15}, Wikipedia logs^{16,17}, Twitter¹⁸, or a combination of multiple data sources¹⁹. These methods are limited in their generalization to other regions²⁰, and may still have instabilities, such as overestimation²¹. Koreans need a timely and valid surveillance system to complement the sentinel surveillance system⁵.

¹College of Pharmacy, Yonsei Institute of Pharmaceutical Research, Yonsei University, 162-1 Songdo-dong, Yeonsu-gu, Incheon, Seoul, Republic of Korea. ²Division of Infectious Diseases, Department of Internal Medicine, National Health Insurance Service Ilsan Hospital, Ilsan, Republic of Korea. ³Division of Infectious Diseases, Department of Internal Medicine, Ansan Hospital, Korea University College of Medicine, Ansan, Republic of Korea. ⁴These authors contributed equally: HeeKyoung Choi and Won Suk Choi. ✉email: eunahan@yonsei.ac.kr

We sought to develop a simpler and faster surveillance system using real-time prospective drug utilization system in Korea. There is only one public health insurer in South Korea, and all Koreans and legally-residing foreigners are mandatory beneficiaries and all medical institutions (including pharmacies) are compulsory providers. Therefore, the single-payer National Health Insurance (NHI) system of Korea allows national and regional level research of healthcare service use for the entire population or representative samples in Korea. Given the comprehensive coverage of insurance under the NHI, the access barrier to healthcare service use is relatively low and thus sample selection is less of a concern. The Health Insurance Review and Assessment Service (HIRA) in Korea operates a drug utilization review (DUR) system that monitors the prescribing and dispensing of drugs nation-wide and in real-time. Therefore, by monitoring the rate of drug prescriptions related to ILI, we can capture the ILI status by country-wide and sub-regional in real time. To prove this, we defined ILI-related prescription claims, calculated the rate of the ILI-related claims among the total claims, and assessed its association with the existing clinical sentinel surveillance system.

Results

We collected a total of 208 weeks of ILI data from Korea Centers for Disease Control and Prevention (KCDC) and outpatient claims data from the NHI system from 2014 to 2018. The epidemic peak in the ILI data was observed in the 8th week (45.5 ILI visits per 1000 patients) in the 2014–2015 season, the 7th week (53.8 ILI visits per 1000 patients) in the 2015–2016 season, the 52nd week (86.2 ILI visits per 1000 patients) in the 2016–2017 season, and the 1st week (72.1 ILI visits per 1000 patients) in the 2017–2018 season. There was a total of 109,214,840 NHI outpatient claims during the 2014–2018 period. The proportions of ILI-related claims to the total outpatient claims were 6.2%, 5.9%, 6.4%, 6.1%, and 6.4% in 2014, 2015, 2016, 2017, and 2018, respectively. Those claims were arranged weekly as the 36th week of the previous year to the 35th week of the years 2014–2015, 2015–2016, 2016–2017, and 2017–2018. Approximately 21–22 million claims were included in each season. Similar to the sentinel surveillance data, the epidemic peak in the ILI data was observed in the 8th week in the 2014–2015 season, the 7th week in the 2015–2016 season, the 52nd week in the 2016–2017 season, and the 1st week in the 2017–2018 season (Fig. 1).

Figure 2 shows the relationship between the ILI-related claims rate and the reported ILI rate from the KCDC data. The two surveillance systems show similar trends. A strong correlation was observed over each season (2014–2015 season, $\rho = 0.7001$, $P < 0.001$; 2015–2016 season, $\rho = 0.7774$, $P < 0.001$; 2016–2017 season, $\rho = 0.8074$, $P < 0.001$; and 2017–2018 season, $\rho = 0.8939$, $P < 0.001$).

When comparing the ILI-related claims rate and the KCDC surveillance data rate by age group, the correlation coefficient was the highest in the 50–64-year-old age group for all three seasons ($\rho = 0.7735$ for the 2014–2016 season, $\rho = 0.7455$ for the 2014–2015 season, and $\rho = 0.8283$ for the 2015–2016 season). For the 50–64-year-old age group, the incidence rates peaked at week 8 in the 2014–2015 season and week 7 in the 2015–2016 season (42.4 cases per 1000 patients in the 2014–2015 season and 33.1 cases per 1000 patients in the 2015–2016 season for KCDC data, and 106.9 claims per 1000 claims in the 2014–2015 season and 118 claims per 1000 claims in the 2015–2016 season for NHI claims data). The correlations were statistically significant for all other age groups, with the correlation coefficients ranging from 0.5529 to 0.7732 across seasons (Table 1).

Figure 3 shows the ILI trends according to age group during the 2014–2016 seasons. In most weeks and age groups, the ILI-related claims rates were higher than the KCDC reported ILI rates. In individuals under six years of age, there was a more pronounced fluctuation in the KCDC rate than the ILI-related claims rate. Conversely, in those aged 65 years and older, the ILI-related claims rate showed significant fluctuation compared to the KCDC rates.

We examined the partial autocorrelation function for sentinel and claims surveillance data, respectively, which showed a statistically significant autocorrelation at the 5% level for a lag of one week for the sentinel data and a lag of two weeks for the claims data (results not shown). The augmented Dickey-Fuller test for the null hypothesis of the unit root process (i.e., non-stationarity) was not rejected at the 5% level for either data, i.e., there was no statistically significant persistent time series, and differencing was not applied. We applied the autoregressive function for respective lags for each dataset and examined the cross-correlation coefficients between the two datasets. The residuals for the correlations were approximately normally distributed (results not shown). Figure 4 shows the cross-correlations between the residuals of the first-order (sentinel data) or the second-order (claims data) autoregressive function of each data, representing a gradual decrease in the correlation coefficients for lags greater than zero. These results indicate that the claims data neither lead nor lag the sentinel data.

Figure 5 and Supplementary Fig. 1 show the MEM analysis results. The absolute thresholds for the four-year surveillance period (2014–2018) were 84.64 and 86.19 claims per 1,000 claims for the claims data and 12.27 and 16.82 per 1000 patients for the sentinel data (Supplementary Fig. 1). Both the claims and sentinel data surpassed the respective epidemic threshold in each of the four seasons. The epidemic was relatively longer in the sentinel data than the claims data, and the epidemic peaked in the claims data one to two weeks later than in the sentinel data. The epidemic pattern showed greater similarity in terms of the peak during the epidemic period in the 2016–2017 and 2017–2018 seasons than the 2014–2015 and 2015–2016 seasons (Fig. 5).

Discussion

The influenza surveillance system of Korea is operated by the KCDC and consists of the three systems; clinical sentinel surveillance, laboratory sentinel surveillance, and an influenza hospitalization and mortality surveillance system^{3,4}. Among them, the clinical sentinel surveillance system began operation on a pilot basis in 1997 with more than 70 private medical institutions. In 2000, the system was expanded to the Korea Influenza Surveillance Scheme, which consists of a clinical and a laboratory monitoring system involving public health centers and private medical institutions. In 2008, the public health center's zero reporting rate was high, even during

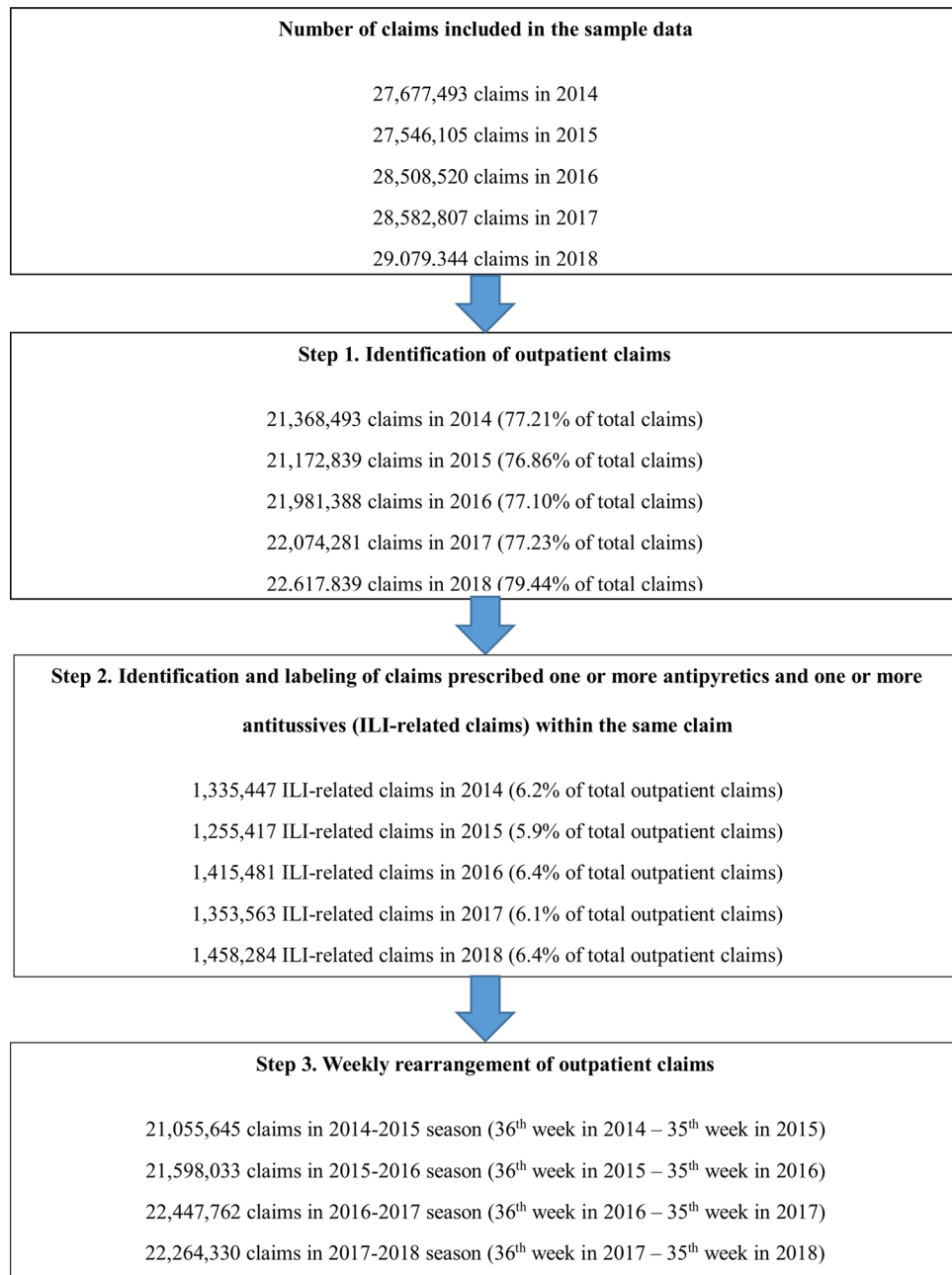


Figure 1. Defining ILI-related claims.

the influenza epidemic, so the public health center was replaced with a private medical institution to secure more reliable data. In 2009, the number of monitoring agencies was expanded, but there was still a limit to the number of zero reporting sites even during an epidemic period. During expert meetings held in 2013, 200 clinical surveillance institutions were designated as active participants, 36 of which were invited to participate in a laboratory sentinel surveillance system. Currently, the clinical sentinel surveillance involves the selection of 200 outpatient clinics, designated by a medical association, including 100 pediatric clinics and 100 internal or family medicine clinics with specialties in internal medicine, pediatrics, and family medicine.⁸ Site selection is based on geographical distribution and population characteristics (Supplementary Table 3)²². The number of ILI patients and the total number of outpatients should be reported weekly from April to November of the current year and daily from December to April of the following year. After collecting each clinic's data, KCDC releases the proportion of ILI visits per 1000 patients per week each week.

However, one disadvantage to such a surveillance system is that a one to two-week reporting lag is inevitable. Considering the potential mismatch of influenza vaccines and circulating virus strains and the limited capacity to prevent an influenza epidemic in advance²³, it is essential to control an epidemic promptly⁵. Such a sentinel alert system is also time and labor-intensive: a network of 400 outpatient providers across Korea submit ILI

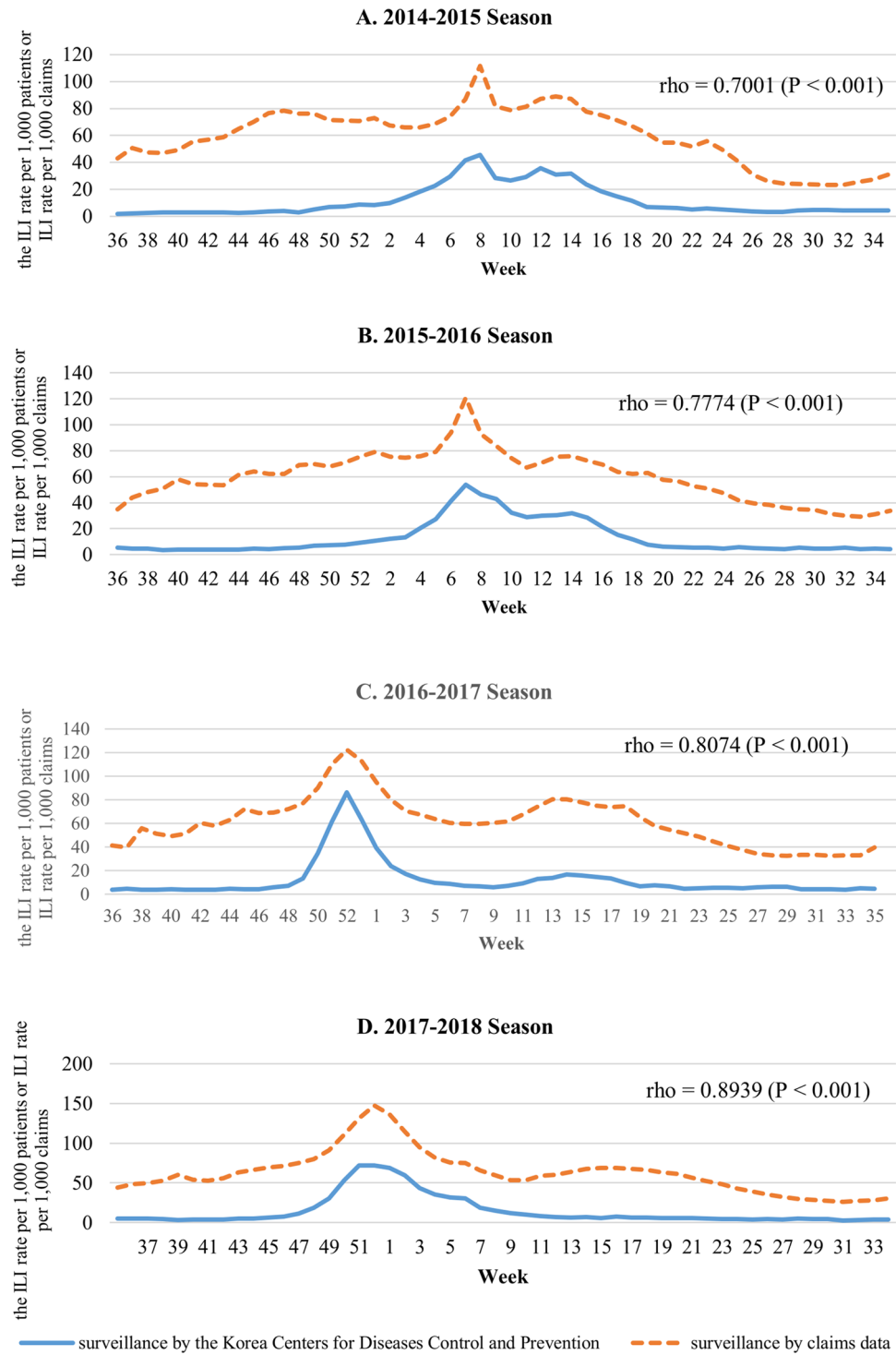


Figure 2. (A) Comparison of the ILI-related claims rate and clinical sentinel surveillance report, 2014–2015 season. (B) Comparison of the ILI-related claims rate and clinical sentinel surveillance report, 2015–2016 season. (C) Comparison of the ILI-related claims rate and clinical sentinel surveillance report, 2016–2017 season. (D) Comparison of the ILI-related claims rate and clinical sentinel surveillance report, 2017–2018 season.

counts, which are then summarized by the KCDC⁸. Also, reports may not always reflect the trends across the country^{5,24}, as participation is voluntary⁸ and results are from conveniently selected sample clinics⁸. This can lead to under-reporting, especially early in the season, when both doctors and the general population are unaware of

	ILI claims	Total claims	Number of observations (week)	Rho	P-value*
2014–2016 season					
0–6 years of age	244,804	4,857,228	104	0.6350	<0.001
7–18 years of age	334,639	3,287,888	104	0.6865	<0.001
19–49 years of age	918,294	12,121,075	104	0.7078	<0.001
50–64 years of age	582,828	10,748,959	104	0.7735	<0.001
65 +	493,298	11,638,528	104	0.7208	<0.001
All	2,573,863	42,653,678	104	0.7328	<0.001
2014–2015 season					
0–6 years of age	116,658	2,387,055	52	0.5529	<0.001
7–18 years of age	168,546	1,638,138	52	0.6755	<0.001
19–49 years of age	449,593	6,027,427	52	0.6537	<0.001
50–64 years of age	289,251	5,272,972	52	0.7455	<0.001
65 +	250,284	5,730,053	52	0.7234	<0.001
All	1,274,332	21,055,645	52	0.7001	<0.001
2015–2016 season					
0–6 years of age	128,146	2,470,173	52	0.7057	<0.001
7–18 years of age	166,093	1,649,750	52	0.7376	<0.001
19–49 years of age	468,701	6,093,648	52	0.7647	<0.001
50–64 years of age	293,577	5,475,987	52	0.8283	<0.001
65 +	243,014	5,908,475	52	0.7732	<0.001
All	1,299,531	21,598,033	52	0.7774	<0.001

Table 1. Comparison of the weekly ILI-related claims rate and clinical sentinel surveillance report, 2014–2016. Since HIRA's age group has been provided differently from KCDC since 2017, only the 2014–2016 data set was used for analysis by age group. *P < 0.05 is considered statistically significant (indicated by bold text).

an influenza outbreak. Several weeks may pass between the actual onset of a seasonal epidemic and the official alert. As reporting can be affected by the patient volume and the physician's compliance⁹, the amount of under-reporting varies from clinic to clinic.

Therefore, new surveillance systems attempt to overcome these shortcomings and detect early influenza signals^{14,15,19,25–30}. Google Flu Trends was a web service that predicted an influenza epidemic based on 45 queries¹⁴. Although initial reports had a 97% accuracy compared to the United States Center for Disease Control data³¹, subsequent reports revealed Google Flu Trends to be inaccurate²¹. The errors in Google Flu Trends may have been due to media-stoked panic and the fact that people making flu-related searches may not have symptoms, but were researching other disease symptoms similar to influenza²¹. Other social media platforms, such as Twitter, have made efforts to remove unrelated messages to improve programme accuracy¹⁸. Regardless, searches on Google, Wikipedia, or social media are likely to vary from country to country³⁰. The coverage even within a country varies, as rural areas with older populations are likely to have low access to social media¹⁸ and are thus systematically underrepresented. External validation of algorithms using this data is questionable.

ILI does not have a diagnostic code that is adequately responsive. By contrast, drug prescription data are reliable because they reflect the doctor's judgment regarding the patient's condition. Therefore, claims with antiviral agents for influenza can be used for surveillance, which is likely to detect confirmed influenza cases. Claims data, including prescription records, are valid as the patient's subjective symptoms are confirmed by the prescribing physician. A surveillance system based on claims data can predict influenza epidemics more efficiently and reliably than internet search engines or a self-report system, without any additional data collection. The national health insurance system in South Korea is particularly useful for complementary influenza surveillance given that the claims data represents healthcare service use for the entire population, and most of the services are reimbursed.

There may be a time lag between the actual practice and the reimbursement claim. Therefore, we propose applying this study's concepts to the nationwide Drug Utilization Review (DUR) system in South Korea. DUR is operated by the HIRA, which has all of the reimbursement claims data for review and assessment. The DUR system detects potential side effects and unsafe use at the time of prescribing and dispensing the medications, providing patient records of drug use and real-time alerts³². In this system, a physician's prescription is cross-checked by the HIRA using a specialized database that can warn doctors via their computer screen in real-time (Supplementary Fig. 2). The doctor may then change the prescription or note the reason they are keeping it as is. The final decision will be sent to HIRA. Pharmacists will follow the same process. A warning message on the DUR system prompts the pharmacist to follow-up with the doctor before dispensing the medication. Final dispensing details are sent to the HIRA. This process is completed within 0.5 s. All providers in Korea have access to the DUR system and send prescription and dispensing details to the HIRA. Therefore, ILI surveillance can be performed in real-time by adding an indicator for ILI-related claims in the DUR system.

The proposed alert system is similar to a real-time alert system for influenza epidemics in other countries, including Japan and Taiwan^{5,24}. Taiwan's system proactively sends an alert to subscribers via a mobile application

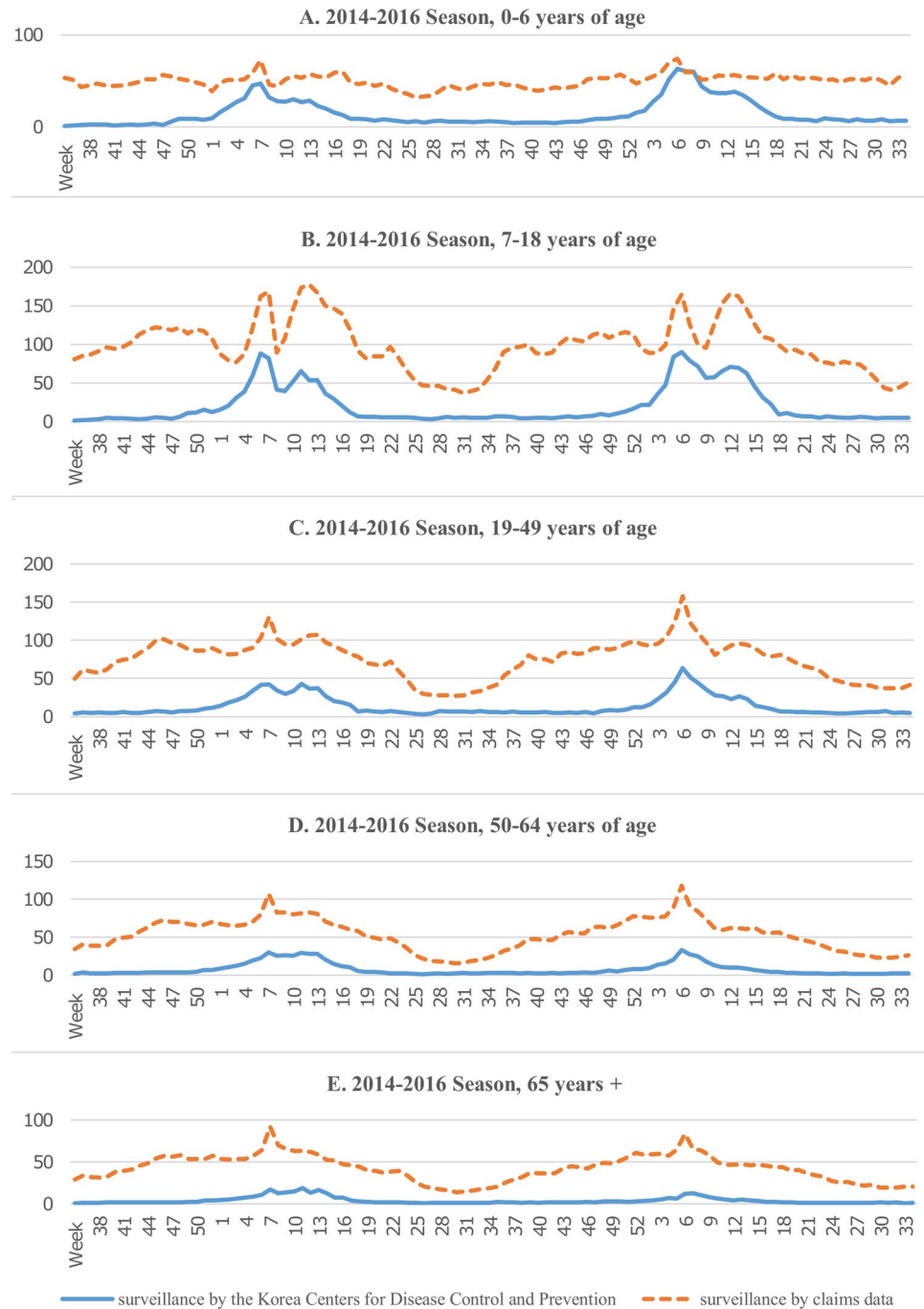


Figure 3. Comparison of the ILI-related claims rate and clinical sentinel surveillance report by age group, 2014–2016.

or computerized physician order entry when one of the three surveillance systems, based on the sentinel reports, insurance claims, or the electronic medical records of selected hospitals, presents incidence values over the epidemic threshold⁵. However, planting the surveillance system in the prospective DUR is a more novel and efficient manner of capturing and controlling the emerging surveillance alert given that it is based on the compulsory participation of all pharmacies and clinicians in a nation. This ensures that all healthcare providers are subscribers and the government is already part of the system.

There are several limitations in this study. First, we did not compare ILI claims rates with laboratory-confirmed influenza cases. The ILI rates do not represent true cases of influenza. ILI has a low sensitivity (30–70%) for predicting laboratory confirmed influenza^{33,34}. Patients who were diagnosed with laboratory confirmed influenza are prescribed antiviral agents, but not everyone is prescribed antiviral agents. Moreover, some antiviral agents are not covered by insurance. Therefore, all confirmed influenza cases are not detected by claims data alone, and laboratory data is required to ensure that influenza-specific policies are initiated. We also acknowledge

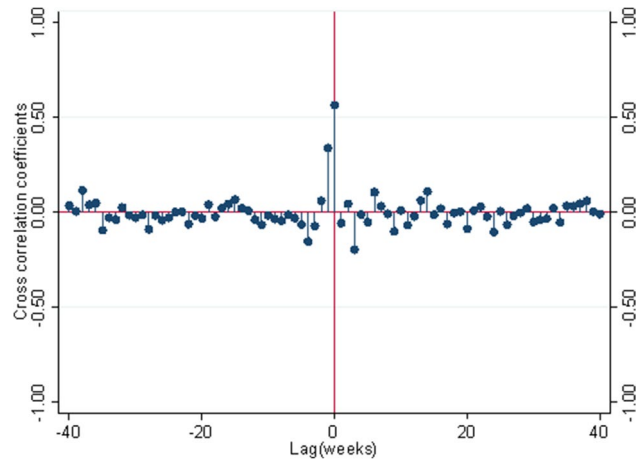


Figure 4. Cross-correlation between the KCDC sentinel ILI rates and National Health Insurance claims ILI rates.

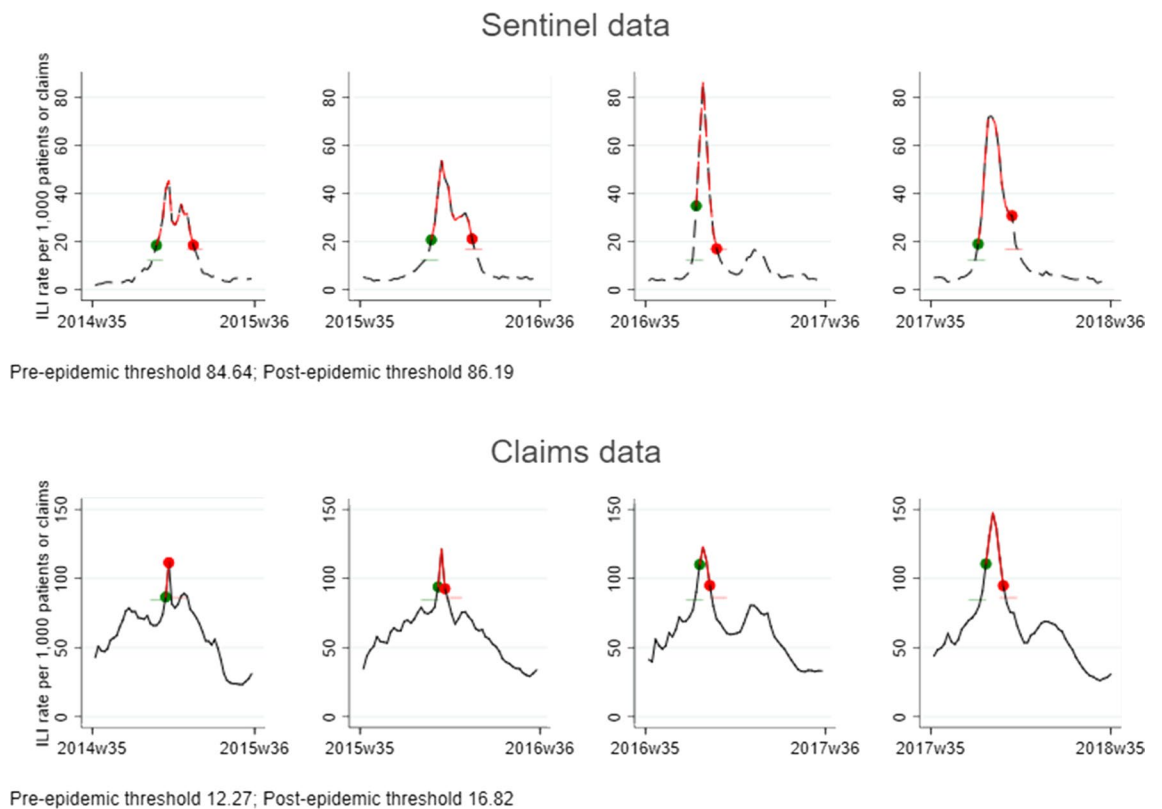


Figure 5. Epidemic thresholds for the surveillance of ILI rates from the KCDC sentinel data and National Health Insurance claims data.

that monitoring ILI-related prescription claims does not provide information about the virus or disease severity and cannot replace the entire influenza surveillance system. The system suggested in this study cannot replace all of the traditional influenza surveillance systems, and the need for laboratory or hospitalization and mortality surveillance still exists. Regardless, this may complement the traditional clinical sentinel surveillance to provide quicker and easier data collection and analysis⁵. Second, national health insurance in Korea is a single-payer program, and has an electronic prescription monitoring system that streamline real-time data collection and reporting. Therefore, implementation of this scheme is limited in other countries with other health insurance systems. Third, physicians may avoid prescribing antitussive agents despite the patient's symptoms. This trend is expected to occur in young children, because it is recommended to use antitussive agents more carefully in younger children due to concerns about side effects³⁵. In conclusion, the weekly fraction of outpatient claims having both antipyretic and antitussive agents among the total claims were similar to the existing sentinel ILI

surveillance system. This suggests that it is possible to integrate a new, real-time influenza surveillance system to the existing system for efficient and timely surveillance.

Methods

Data source. We used the Health Insurance Review and Assessment Service-National Patient Samples (HIRA-NPS), 2014–2018, to identify ILI-related insurance claims. The HIRA-NPS is a stratified random sample of 3% of the Korean population that includes approximately 1.4 million individuals. Because the HIRA-NPS extracts data from the National Health Insurance System (NHIS), it only includes data on claims that are reimbursed by the NHIS³⁶.

National weekly ILI rates were pulled from public data provided by the Korea Centers for Diseases Control and Prevention (KCDC)³⁷. The selected sites for the clinical sentinel surveillance system report the number of ILI patients and the number of outpatients. Based on this report, KCDC releases the weekly incidence of ILI per 1000 patients.

To reinforce personal information protection, HIRA has revised the principle of providing data for research from 2017. They offered only the age group instead of the exact age, and the age groups of KCDC and HIRA were different. Therefore, only data prior to 2017 was used for comparison by age group.

Operational definition of ILI. ILI is defined as an acute respiratory illness with a measured temperature of 38.0 °C or higher and a cough with onset within the last 10 days³⁸. For sentinel surveillance, ILI is defined as the sudden onset of fever (> 38.0 °C) with cough or sore throat³⁹. The ILI rate, calculated using the total number of weekly outpatient patients as the denominator and the number of weekly patients with ILI as the numerator, is reported as the number of ILI patients per 1000 outpatients per week.

We defined an influenza-like illness (ILI)-related claim as an outpatient claim that contains both antipyretic and antitussive agents (Supplementary Table 1, 2). The ILI-related claims rate was defined as the proportion of ILI-related claims of the entire outpatient claims for a given period. We used the following process to define an ILI-related claim (Fig. 1): (1) we pulled outpatient claims; (2) we identified those outpatient claims with both antipyretic and antitussive agents (ILI-related claims, hereafter); (3) we reorganized claims weekly, like the clinical sentinel surveillance data issued by the KCDC; and (4) the ILI-related claims rate was calculated as the weekly number of ILI-related claims per 1000 outpatient claims per week.

Statistical analysis. We first performed a cross-correlation analysis using Pearson's correlation to compare ILI-related claims rates and weekly ILI rates from the KCDC's clinical sentinel surveillance data. We then used a time-series analysis to examine the autocorrelation in both surveillance data to explore actual correlations after removing any dubious correlations due to underlying non-stationarity in both data sets. We performed the partial autocorrelation function for all lags and the augmented Dickey-Fuller test for the null hypothesis of the unit root process (i.e., non-stationarity)^{40,41}. We then generated cross-correlation coefficients for residuals after controlling for any autocorrelation properties. Lastly, we used the moving epidemic method (MEM) to estimate an absolute threshold from historical influenza incidence data and to designate potential influenza epidemics for the weeks with incidence rates above the threshold^{42,43}. We estimated the absolute threshold for the sentinel data and insurance claims data, respectively, using the MEM, and compared the epidemic period and whether the peak in the epidemic period is visually overlapped for each data set. Data were analyzed weekly over 12 months for the entire study duration and separately by each surveillance period starting from the 36th week of a calendar year to the 35th week of the following calendar year.

Data analysis was performed using SAS software, version 9.4 (Cary, NC), Stata software version 14 (Stata-Corp, College Station, TX, USA, <https://www.stata.com/stata14/>), and R 3.4.2 (CRAN) using the MEM library version 2.11⁴⁴.

Ethical statement. This study was approved by the Institutional Review Board (IRB) of Yonsei University (approval number: 201905-h-1535-01) and Korea University Ansan Hospital (approval number: 2019AS0107).

Received: 11 December 2020; Accepted: 6 May 2021

Published online: 27 May 2021

References

1. Webster, R. G., Wright, S. M., Castrucci, M. R., Bean, W. J. & Kawaoka, Y. Influenza—A model of an emerging virus disease. *Intervirology* **35**, 16–25. <https://doi.org/10.1159/000150292> (1993).
2. Cassini, A. *et al.* Impact of infectious diseases on population health using incidence-based disability-adjusted life years (DALYs): results from the Burden of Communicable Diseases in Europe study, European Union and European Economic Area countries, 2009 to 2013. *Euro Surveill.* **23**. <https://doi.org/10.2807/1560-7917.ES.2018.23.16.17-00454> (2018).
3. *2018–2019 Season Influenza Management Guideline.* (Korea Centers for Disease Control and Prevention (KCDC), 2018).
4. Choi, W. S. The national influenza surveillance system of Korea. *Infect. Chemother.* **51**, 98–106. <https://doi.org/10.3947/ic.2019.51.2.98> (2019).
5. Yang, C. Y., Chen, R. J., Chou, W. L., Lee, Y. J. & Lo, Y. S. An integrated influenza surveillance framework based on national influenza-like illness incidence and multiple hospital electronic medical records for early prediction of influenza epidemics: Design and evaluation. *J. Med. Internet. Res.* **21**, e12341. <https://doi.org/10.2196/12341> (2019).
6. Marsden-Haug, N. *et al.* Code-based syndromic surveillance for influenza like illness by International Classification of Diseases, ninth revision. *Emerg. Infect. Dis.* **13**, 207 (2007).

7. Brammer, L., Budd, A. & Cox, N. Seasonal and pandemic influenza surveillance considerations for constructing multicomponent systems. *Influenza Other Respir. Viruses* **3**, 51–58. <https://doi.org/10.1111/j.1750-2659.2009.00077.x> (2009).
8. *Infectious Diseases Surveillance Yearbook, 2017* (Korea Centers for Diseases Control and Prevention, 2017).
9. Temte, J. L. & Beasley, J. W. Rate of case reporting, physician compliance, and practice volume in a practice-based research network study. *J. Fam. Pract.* **50**, 977 (2001).
10. Buehler, J. W., Sonricker, A., Paladini, M., Soper, P. & Mostashari, F. Syndromic surveillance practice in the United States: Findings from a survey of state, territorial, and selected local health departments. *Adv. Dis. Surveill.* **6**, 1–20 (2008).
11. Patwardhan, A. & Bilkovski, R. Comparison: Flu prescription sales data from a retail pharmacy in the US with Google Flu trends and US ILINet (CDC) data as flu activity indicator. *PLoS ONE* **7**, e43611. <https://doi.org/10.1371/journal.pone.0043611> (2012).
12. Olson, D. R. *et al.* Monitoring the impact of influenza by age: Emergency department fever and respiratory complaint surveillance in New York City. *PLoS Med.* **4**, e247. <https://doi.org/10.1371/journal.pmed.0040247> (2007).
13. Egger, J. R. *et al.* Usefulness of school absenteeism data for predicting influenza outbreaks, United States. *Emerg. Infect. Dis.* **18**, 1375 (2012).
14. Ginsberg, J. *et al.* Detecting influenza epidemics using search engine query data. *Nature* **457**, 1012–1014. <https://doi.org/10.1038/nature07634> (2009).
15. Yuan, Q. *et al.* Monitoring influenza epidemics in china with search query from baidu. *PLoS ONE* **8**, e64323. <https://doi.org/10.1371/journal.pone.0064323> (2013).
16. Hickmann, K. S. *et al.* Forecasting the 2013–2014 influenza season using Wikipedia. *PLoS Comput. Biol.* **11**, e1004239. <https://doi.org/10.1371/journal.pcbi.1004239> (2015).
17. Sharpe, J. D., Hopkins, R. S., Cook, R. L. & Striley, C. W. Evaluating Google, Twitter, and Wikipedia as tools for influenza surveillance using Bayesian change point analysis: A comparative analysis. *JMIR Public Health Surveill.* **2**, e161. <https://doi.org/10.2196/publichealth.5901> (2016).
18. Broniatowski, D. A., Paul, M. J. & Dredze, M. National and local influenza surveillance through Twitter: An analysis of the 2012–2013 influenza epidemic. *PLoS ONE* **8**, e83672. <https://doi.org/10.1371/journal.pone.0083672> (2013).
19. Santillana, M. *et al.* Combining search, social media, and traditional data sources to improve influenza surveillance. *PLoS Comput. Biol.* **11**, e1004513. <https://doi.org/10.1371/journal.pcbi.1004513> (2015).
20. Olson, D. R., Konty, K. J., Paladini, M., Viboud, C. & Simonsen, L. Reassessing Google Flu Trends data for detection of seasonal and pandemic influenza: A comparative epidemiological study at three geographic scales. *PLoS Comput. Biol.* **9**, e1003256. <https://doi.org/10.1371/journal.pcbi.1003256> (2013).
21. Lazer, D., Kennedy, R., King, G. & Vespignani, A. Big data. The parable of Google Flu: Traps in big data analysis. *Science* **343**, 1203–1205. <https://doi.org/10.1126/science.1248506> (2014).
22. Lee, J. S. *et al.* Influenza surveillance in Korea: establishment and first results of an epidemiological and virological surveillance scheme. *Epidemiol. Infect.* **135**, 1117–1123. <https://doi.org/10.1017/S0950268807007820> (2007).
23. Tricco, A. C. *et al.* Comparing influenza vaccine efficacy against mismatched and matched strains: A systematic review and meta-analysis. *BMC Med.* **11**, 153. <https://doi.org/10.1186/1741-7015-11-153> (2013).
24. Sugawara, T. *et al.* Real-time prescription surveillance and its application to monitoring seasonal influenza activity in Japan. *J Med Internet Res* **14**, e14. <https://doi.org/10.2196/jmir.1881> (2012).
25. Espino, J. U., Hogan, W. R. & Wagner, M. M. Telephone triage: A timely data source for surveillance of influenza-like diseases. *AMIA Annu. Symp. Proc.* 215–219 (2003).
26. Johnson, H. A. *et al.* Analysis of web access logs for surveillance of influenza. *Stud. Health Technol. Inform.* **107**, 1202–1206 (2004).
27. Eysenbach, G. Infodemiology: Tracking flu-related searches on the web for syndromic surveillance. *AMIA Annu. Symp. Proc.* 244–248 (2006).
28. Martin, L. J., Lee, B. E. & Yasui, Y. Google Flu trends in Canada: A comparison of digital disease surveillance data with physician consultations and respiratory virus surveillance data, 2010–2014. *Epidemiol. Infect.* **144**, 325–332. <https://doi.org/10.1017/S0950268815001478> (2016).
29. Santillana, M. *et al.* Cloud-based electronic health records for real-time, region-specific influenza surveillance. *Sci. Rep.* **6**, 25732. <https://doi.org/10.1038/srep25732> (2016).
30. Cho, S. *et al.* Correlation between national influenza surveillance data and google trends in South Korea. *PLoS ONE* **8**, e81422. <https://doi.org/10.1371/journal.pone.0081422> (2013).
31. Ortiz, J. R. *et al.* Monitoring influenza activity in the United States: A comparison of traditional surveillance systems with Google Flu Trends. *PLoS ONE* **6**, e18687. <https://doi.org/10.1371/journal.pone.0018687> (2011).
32. Introduction to HIRA Drug Utilization Review (DUR) System and Its Function to Manage Appropriate Drug Use. <https://www.hira.or.kr/bbsDummy.do?pgmid=HIRAJ03000001000&brdScnBltno=4&brdBltno=46903#none>.
33. Dugas, A. F. *et al.* Clinical diagnosis of influenza in the ED. *Am. J. Emerg. Med.* **33**, 770–775. <https://doi.org/10.1016/j.ajem.2015.03.008> (2015).
34. Ong, A. K. *et al.* Improving the clinical diagnosis of influenza—A comparative analysis of new influenza A (H1N1) cases. *PLoS ONE* **4**, e8453. <https://doi.org/10.1371/journal.pone.0008453> (2009).
35. Siwek, J. & Lin, K. W. Choosing wisely: More good clinical recommendations to improve health care quality and reduce harm. *Am. Fam. Phys.* **88**, 164–168 (2013).
36. Kim, L., Kim, J. A. & Kim, S. A guide for the utilization of health insurance review and assessment service national patient samples. *Epidemiol. Health* **36**, e2014008. <https://doi.org/10.4178/epih/e2014008> (2014).
37. Infectious Disease Portal. <http://www.cdc.go.kr/npt/biz/npp/iss/influenzaStatisticsMain.do>.
38. Fitzner, J. *et al.* Revision of clinical case definitions: Influenza-like illness and severe acute respiratory infection. *Bull World Health Organ* **96**, 122–128. <https://doi.org/10.2471/BLT.17.194514> (2018).
39. In Hyekyung, L. D., Gu, K.M., Hyuk, C., Joo-Yeon, L., Kisoon, K. *Public Health Weekly Report*. Vol. 10 185–193 (Korea Centers for Diseases Control and Prevention, 2017).
40. Dickey, D. A. & Fuller, W. A. Distribution of the estimators for autoregressive time-series with a unit root. *J. Am. Stat. Assoc.* **74**, 427–431. <https://doi.org/10.2307/2286348> (1979).
41. Stata, A. *Stata Base Reference Manual Release 14*. (2015).
42. Vega, T. *et al.* Influenza surveillance in Europe: Establishing epidemic thresholds by the moving epidemic method. *Influenza Other Respir. Viruses* **7**, 546–558. <https://doi.org/10.1111/j.1750-2659.2012.00422.x> (2013).
43. Vega, T. *et al.* Influenza surveillance in Europe: comparing intensity levels calculated using the moving epidemic method. *Influenza Other Respir. Viruses* **9**, 234–246. <https://doi.org/10.1111/irv.12330> (2015).
44. Lozano, J. [lozalojo/mem: Second Release of the MEM R Library. Zenodo](https://zenodo.org/record/165983). <https://zenodo.org/record/165983>. Accessed 1 Feb 2017.

Acknowledgements

This data is based on the Health Insurance Review and Assessment Service (HIRA)'s national patient sample data (HIRA-NPS-2014, HIRA-NPS-2015, HIRA-NPS-2016, HIRA-NPS-2017, and HIRA-NPS-2018), and the research results are not related to the HIRA or Ministry of Health and Welfare.

Author contributions

H.K.C. conducted the formulation of the research idea, developed the analysis plan, conceptualized the paper, analyzed the data, and wrote the manuscript. W.S.C. participated in the formulation of the research idea, study design, analysis, and interpretation of data. E.H. conducted the formulation of the research idea, design of the study, study planning, and analysis, interpreted the data, and critically revised the manuscript.

Funding

Funding was provided by National Research Foundation of Korea (Grant no. 2019R1A2C1003259) and National Evidence-based Healthcare Collaborating Agency (Grant no. HC20C0010).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-90511-0>.

Correspondence and requests for materials should be addressed to E.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021