

Modeling and Forecasting Influenza-like Illness (ILI) in Houston, Texas Using Three Surveillance Data Capture Mechanisms

Susannah Paul, MPH^{1*}, Osaro Mgbere, PhD, MS, MPH^{2,3}, Raouf Arafat, MD, MPH², Biru Yang, PhD, MPH², Eunice Santos, MPH²

¹ Rollins School of Public Health, Emory University, Atlanta, GA

² Houston Health Department, Houston, TX

³ Institute of Community Health, University of Houston College of Pharmacy, Texas Medical Center, Houston, TX

*Susannah Paul, MPH, Rollins School of Public Health, Emory University, Atlanta, GA. susannahpaul@outlook.com

Abstract

Objective

The objective was to forecast and validate prediction estimates of influenza activity in Houston, TX using four years of historical influenza-like illness (ILI) from three surveillance data capture mechanisms.

Background

Using novel surveillance methods and historical data to estimate future trends of influenza-like illness can lead to early detection of influenza activity increases and decreases. Anticipating surges gives public health professionals more time to prepare and increase prevention efforts.

Methods

Data was obtained from three surveillance systems, Flu Near You, ILINet, and hospital emergency center (EC) visits, with diverse data capture mechanisms. Autoregressive integrated moving average (ARIMA) models were fitted to data from each source for week 27 of 2012 through week 26 of 2016 and used to forecast influenza-like activity for the subsequent 10 weeks. Estimates were then compared to actual ILI percentages for the same period.

Results

Forecasted estimates had wide confidence intervals that crossed zero. The forecasted trend direction differed by data source, resulting in lack of consensus about future influenza activity. ILINet forecasted estimates and actual percentages had the least differences. ILINet performed best when forecasting influenza activity in Houston, TX.

Conclusion

Though the three forecasted estimates did not agree on the trend directions, and thus, were considered imprecise predictors of long-term ILI activity based on existing data, pooling predictions and careful interpretations may be helpful for short term intervention efforts. Further work is needed to improve forecast accuracy considering the promise forecasting holds for seasonal influenza prevention and control, and pandemic preparedness.

Keywords: Influenza-like illness, syndromic surveillance, forecast, Houston, Texas, Flu Near You, ILINet, Hospital Emergency Center, ARIMA, Houston, Texas

Correspondence: Susannah Paul, MPH, Rollins School of Public Health, Emory University, Atlanta, GA.
susannahpaul@outlook.com

DOI: 10.5210/ojphi.v9i2.8004

Copyright ©2017 the author(s)

This is an Open Access article. Authors own copyright of their articles appearing in the Online Journal of Public Health Informatics. Readers may copy articles without permission of the copyright owner(s), as long as the author and OJPHI are acknowledged in the copy and the copy is used for educational, not-for-profit purposes

Objective

Our objective was to forecast and validate prediction estimates of influenza-like illness (ILI) activity for MMWR weeks 27 through 37 of 2016 using historical ILI data from MMWR week 27 of 2012 through week 26 of 2016 from three surveillance sources (Flu Near You, ILINet, and the Houston Health Department's syndromic surveillance system based on hospital emergency center visits) that have diverse data capture mechanisms.

Introduction

Effective management of both seasonal and pandemic influenza requires early detection of the outbreak through timely and accurate surveillance linked with a rapid response to mitigate crowding [1]. Several diverse data sources, including historical and real-time, have been used to forecast influenza activity based on predictive models that could facilitate key preparedness actions and improve understanding of the epidemiological dynamics or evaluate potential control strategies [2-6].

Houston Health Department (HHD) conducts ILI surveillance with data from five different systems. Each system is used based on agreed upon terms for data use in public health surveillance. Three of the data sources were selected for this study; however, all five systems are routinely used for ILI surveillance. Despite the variety of syndromic surveillance systems, hospital emergency center (EC) visits are utilized for both descriptive and inferential statistics. Advanced analysis for ILI activity are performed with hospital EC data. While ILINet and Flu Near You (FNY) are used

for descriptive analysis, historically hospital EC visit data are readily available and the mechanism for reporting from healthcare to public health lends itself to the timely situational awareness for seasonal epidemic and pandemic level ILI activity. For this reason, the exploration of forecasting methods was applied and validated with a comparison of ILINet, FNY, and hospital EC visits. Ideally, the forecasting methods can become part of routine ILI surveillance at HHD.

Electronic Surveillance System for the Early Notification of Community based Epidemics (ESSENCE) includes statistical algorithms to detect aberrations in syndromic surveillance for various conditions including conditions with seasonal patterns. However, ESSENCE does not provide forecasting based on time series data. Supplemental methods such as prediction or forecasting with autoregressive integrated moving average (ARIMA) can complement the automated features in ESSENCE. Public health jurisdictions that take part in National Syndromic Surveillance Program (NSSP) have access to RStudio Pro on the BioSense Platform. The software version and methods in this study can be replicated in the BioSense Platform. Therefore, users of the NSSP BioSense Platform can complement ESSENCE automated statistical algorithms with ARIMA methods. Soon, Houston will be participating in NSSP at which time forecasting and validation can be applied as part of the Biosense platform.

ILI surveillance is a cornerstone in the early detection of seasonal influenza epidemics and other pandemics such as the H1N1. Because ILI surveillance is of paramount importance for public health authorities and healthcare, a variety of syndromic surveillance systems have been explored in the scientific literature and in practice to inform public health and clinical surveillance of ILI activity.

The HHD uses several syndromic surveillance systems to monitor ILI activity which include ILINet, hospital EC visits, and novel data sources such as FNY. The surveillance activities for influenza in Houston are adapted to meet the current public health needs. For example, since 2007 the HHD has used sentinel providers in outpatient clinics to obtain laboratory specimens for virologic surveillance of ILI and to characterize flu activity in Houston [7]. The sentinel provider network was instrumental in monitoring the novel 2009 H1N1 pandemic in Houston [7]. More recently, the ILI data has been used to explore the dynamics of patient health behavior associated with repeat episodes of ILI [8]. However, no attempt has been made thus far to evaluate the forecasting capabilities of existing syndromic surveillance systems (ILINet, hospital EC visit data, and FNY) using Houston's data. These diverse real-time data capture mechanisms may have the potential to go beyond early detection and forecast future ILI or influenza outbreaks in the community. Such estimates could help guide policy makers' decisions, and key preparedness tasks such as public health surveillance, development and use of medical countermeasures, communication strategies, deployment of Strategic National Stockpile assets in anticipation of surge demands, and hospital resource management [2].

Modern epidemiological forecasts of common illnesses, such as the flu, rely on both traditional surveillance sources as well as digital surveillance data such as social network activity and search queries [9]. Reliable forecasts could aid in the selection and implementation of interventions to reduce morbidity and mortality due to influenza illness. Previous studies have applied statistical methods on FNY and ILINet data to compare and validate their potentials to inform ILI surveillance both individually and in combination with each other [10,11]. ARIMA technique has

been used to forecast seasonal influenza surveillance, including ILI at national, state, and local levels. ILI has seasonality which can be corrected for with ARIMA. ILI, like other infectious conditions, does not assume independence and the measures of occurrence are expected to vary across time. Statistical methods used on ILI data must be robust enough to handle the violation of the independence assumption and to reduce the background noise. ARIMA results are considered reliable despite the intercorrelation of data points. ARIMA had been applied by other researchers on the FNY, ILINet, and hospital EC visits with informative and accurate results [2-6]. Results from forecasting methods applied to ILI data have been reported for FNY and ILINet at national levels [10, 11], and provided useful information for population level ILI surveillance. Prediction for ILI occurrence is being explored for the first time on data specific to Houston jurisdiction.

There is a need to understand if ARIMA with seasonal correction can be applied to various data from ILI syndromic surveillance systems in different geographic settings. ILINet and FNY Dashboard can be accessed by any health department. Since any given health department will differ in the information available for ILI surveillance, it becomes important to know which data source for ILI can provide the most accurate predictions for ILI activity. The authors fill a gap in the scientific literature on the use of these promising data sources to model and predict influenza activity. The objective of this study therefore was to forecast and validate prediction estimates of influenza activity in the City of Houston in Texas using ILI historical data from three surveillance systems, namely FNY, ILINet, and hospital EC visits.

Methods

Study Population and Setting

Houston is the fourth most populous city in the U.S. with 2.3 million residents as of January 2017 [12]. Only data collected within the jurisdiction of Houston, Texas were included in this analysis. The City of Houston jurisdiction was determined using ZIP codes.

Data Sources

The three data sources used for this study include FNY, hospital EC visits, and ILINet. Each of the three data sources represents a different stage of disease manifestation of ILI which is reflected in the magnitude, and possibly timing of ILI activity. FNY is a crowdsourced internet-based participatory syndromic surveillance program to track and monitor weekly ILI occurrence based on self-reports from volunteers [13]. The self-reports represent individuals who may or may not have accessed the healthcare system so the severity of symptoms may vary from mild to severe. Hospital EC visits from the HHD's syndromic surveillance system are a good data source for ILI surveillance and represent individuals whose symptoms are severe enough to seek medical care from an emergency center. ILINet is considered the gold standard for ILI outpatient surveillance and may be the reference point for the timing and magnitude of ILI activity for many health departments. Characteristics that are inherent in ILI data include the use of symptom data to measure the occurrence of ILI versus the use of laboratory confirmed influenza positives.

Data Collection and Measurements

Time series data were obtained from the three data sources and arranged in series of weekly occurrence of ILI percent. Like most methods for syndromic surveillance, the time series are not longitudinal and individuals are not followed across time. Data used for our analysis comprised of weekly ILI percentages from MMWR week 27 of 2012 through MMWR week 26 of 2016, which corresponds to about July 2012 through June 2016. Forecast estimates were validated using ILI percentages from MMWR weeks 27 through 37 of 2016.

All data were obtained from public health surveillance systems. Each of the data sources used captures data differently. FNY is available as both a web-based platform and a phone application. Users self-report various symptoms on a weekly basis. The number of responses with symptoms related to influenza-like illness are aggregated weekly over the total number of responses. ILINet is a collaborative effort between the Centers for Disease Control and Prevention (CDC), health departments, and health care providers. Providers report weekly on the total number of patients seen and the proportion of patients with influenza-like illness. The HHD's syndromic surveillance system receives health data via secure electronic transmission from hospital EC visits for public health surveillance. In Houston, hospital EC visits data is considered more representative compared to the other two data sources because it has better "participation" due to the real-time automated transmission of health data. Daily, the HHD receives 1,500 to 2,000 records for hospital registrations and discharges for City of Houston jurisdiction. In Houston, hospital EC visits are the preferred data source because ILI activity based on hospital EC visits consistently matches national trends.

This study received exempt status approval from the HHD Investigative Review Committee because the data were deidentified and aggregated prior to analysis.

Data Management

All analysis was completed using RStudio 3.3.1 and Microsoft Excel. There were 21 (10%) missing observations from ILINet and one missing observation (0.5%) from the hospital EC visits data. The missing data were estimated using predictive mean matching. The 'mice' package in R imputes missing values with plausible values using the predictive mean matching method. This method uses an algorithm to pull information from other values in the specified variable to predict possible values. Predictive mean matching estimates a linear regression for observed values and picks a value randomly from the posterior predictive distribution of the coefficients of the previous regression to produce a new set of coefficients. These coefficients predict values for all observations. For each missing observation, we picked a set of observations with predicted values close to predicted values of the missing observations [14].

Data were divided into segments from MMWR week 27 of the current year through MMWR week 26 of the following year. This resulted in four time periods, each containing the typical start, peak, and end of the influenza season.

ARIMA Modeling and Forecasting

We used autoregressive integrated moving average (ARIMA) models to describe autocorrelations in the data and forecast time series [15]. The ARIMA model has several advantages for forecasting compared with other methods and is especially useful in modeling the temporal dependence structure of a time series [16,17].

Stationarity

Since the ARIMA model requires stationarity, we used Augmented Dickey-Fuller (ADF) tests and Kwiatkowski-Phillips-Schmidt-Shin (KPPS) unit root tests to determine if each of the data source time series variables were stationary or would require differencing. Each of the sources were also tested using seasonal root tests to determine the appropriate number of seasonal differences required [15].

Fitting an ARIMA Model

The `auto.arima()` function in the R ‘forecast’ package which estimates parameters and model orders using maximum likelihood estimation (MLE) was used to fit the model. This method finds parameter values that maximize the probability of obtaining the observed data. After differencing d times, p and q are chosen by looking for the lowest Akaike Information Criteria (AIC) value. The autocorrelation function (ACF) plots of the residuals and portmanteau Ljung-Box tests were used to determine if there is any autocorrelation in the residuals [15].

Obtaining Point Forecasts from ARIMA Models

Once a model is fitted, the `forecast()` function can be used to predict future values. The forecasting equation (Equation 1) is a univariate linear equation where the predictors are the lags of the variable and/or lags of the forecast errors and/or a possible constant:

$$\text{Equation 1. } \hat{y}_t = \mu + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q} \text{ where,}$$

θ = moving average parameters of order q ,

ϕ = autoregressive parameters of order p ,

\hat{y}_t = prediction estimates at time t ,

y_{t-p} = lagged values of y , and

e = error term.

Results

Table 1. Peak Weeks per Year and ILI Percentages by Data Capture Mechanism

Year	Data Source	Peak Week ⁴	Peak ILI % ⁴	Mean ⁴	Standard Deviation ⁴	Mean (Oct-May)
2012-2013	<i>Flu Near You</i>	2012, 45 th	28.57	3.29	6.13	3.84 ¹
	<i>Hospital EC Visits</i>	2012, 51 st	5.25	1.94	1.20	2.41 ¹
	<i>ILINet</i>	2012, 51 st	2.62	0.72	0.79	1.01 ³
2013-2014	<i>Flu Near You</i>	2013, 51 st	10.29	2.22	2.40	2.48 ¹
	<i>Hospital EC Visits</i>	2013, 51 st	6.59	1.87	1.38	2.35 ¹
	<i>ILINet</i>	2013, 51 st	3.73	0.67	0.94	0.96 ¹
2014-2015	<i>Flu Near You</i>	2014, 37 th	9.43	2.77	2.96	3.56 ²
	<i>Hospital EC Visits</i>	2014, 51 st	3.80	1.55	0.99	2.00 ²
	<i>ILINet</i>	2014, 46 th	1.48	0.28	0.41	0.41 ²
2015-2016	<i>Flu Near You</i>	2016, 25 th	6.9	1.90	1.65	2.12 ³
	<i>Hospital EC Visits</i>	2016, 9 th	2.76	1.24	0.61	1.49 ³
	<i>ILINet</i>	2015, 51 st	0.96	0.18	0.23	0.23 ⁵
¹ Mean was calculated using weeks 40 through 22. ² Mean was calculated using weeks 40 through 21. ³ Mean was calculated using weeks 39 through 21. ⁴ The time interval was week 27 through week 26 of the following year. ⁵ Mean was calculated using weeks 38 through 20.						

Preliminary analysis depicted similarities in peak times and overall trend of increases and decreases of ILI activity (Table 1). Flu Near You did have a much larger range of ILI percentages and was heavily influenced by fluctuating and smaller sample sizes. The results also indicated different peak weeks during the 2015-2016 influenza season. Our study noted that the 2016 influenza season was later, and the overall severity of seasonal influenza was also milder compared to the previous three seasons.

FNY point estimates continually decreased from 3.26% (95% CI: -1.00 – 7.51) to 2.06% (95% CI: -2.59 – 6.72). The 95% confidence intervals for FNY point estimates crossed zero and were wide with a mean confidence range of 9.18%. The ILI percentage estimates for the hospital EC visits data continually increased from 0.72% (95% CI: 0.19 – 1.26) to 1.32% (95% CI: -0.35 – 3.00). All except the initial confidence interval for the 27th week crossed zero with a mean range of 2.43%. Lastly, ILINet ILI percentage estimates start at 0.01% (95% CI: -0.23 – 0.25), decreased to -0.09% (95% CI: -0.70 – 0.53), and increased slowly again to -0.05% (95% CI: -1.03 – 0.92). The mean range between upper and lower 95% limits was 1.18%. Like the other data sources, the 95% confidence interval also crossed zero for ILINet estimates (Table 2).

FNY had the highest average confidence interval difference (9.18%) between upper and lower 95% bounds (Figure 2-A). All the 95% confidence intervals were relatively wide and lower bounds were negative percentages, which cannot be actualized because ILI percentages can only be positive. The data sources did not agree on an increasing or decreasing predictive trend. ILINet forecasted estimates were closest to observed ILI percentages for weeks 27-37 of 2016 (Figure 2-C). The hospital EC visits data forecasted estimates resulted in a substantial difference compared to actual ILI percentages for the same period but did follow the same overall increasing trend (Figure 2-B).

Table 2. Comparison of ILI Forecast Estimates and Actual Percentages by Data Source

Data Source	Estimate Week 27 ^{1,2}	Actual Week 27 ¹	Estimate Week 37 ^{1,2}	Actual Week 37 ¹	Estimate Direction	Actual Direction
<i>Flu Near You</i>	3.26 (-1.00 – 7.51)	2.10	2.06 (-2.59 – 6.72)	2.31	Decrease	Increase
<i>Hospital EC Visits</i>	0.72 (0.19 – 1.26)	0.56	1.32 (-0.35 – 3.00)	0.84	Increase	Increase
<i>ILINet</i>	0.01 (-0.23 – 0.25)	0.02	-0.05 (-1.03 – 0.92)	0.00	Decrease ⁴	Decrease

¹ Values were from MMWR week 27 and week 37 of 2016.
² 95% confidence intervals.
³ It is not possible to have ILI percentages less than 0.
⁴ Estimates did increase slightly after the initial decreasing trend.

Figure 2-A. Flu Near You Forecasted Estimates for Weeks 210-219 Using ILI Percentages from Weeks 1-209

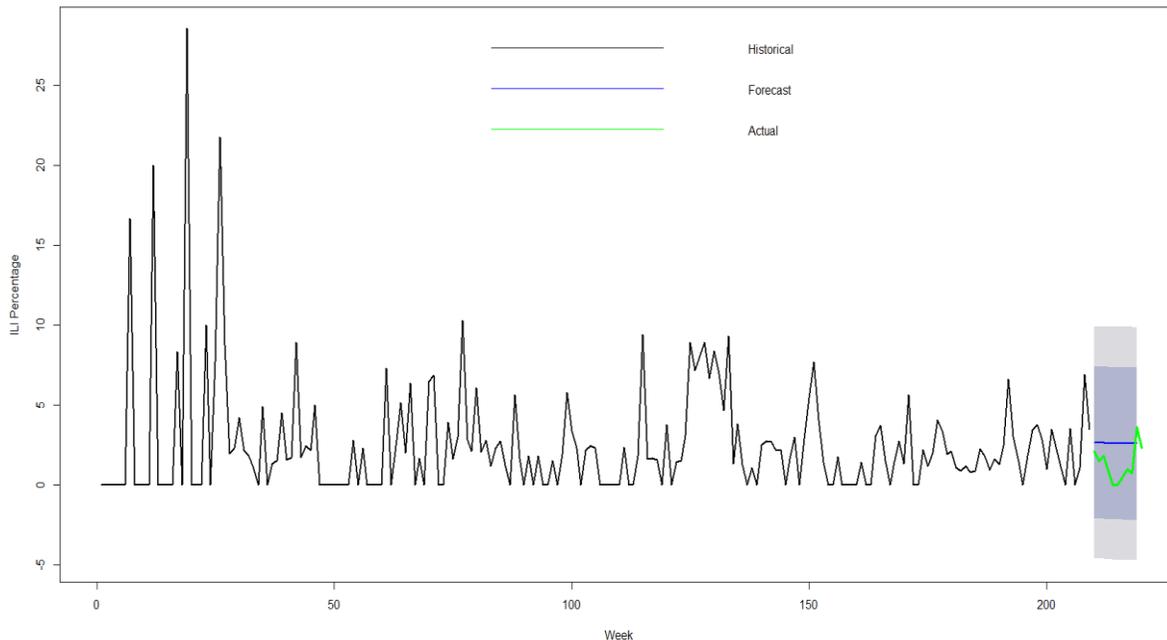


Figure 2-B Hospital EC Visits Forecasted Estimates for Weeks 210-219 Using ILI Percentages from Weeks 1-209

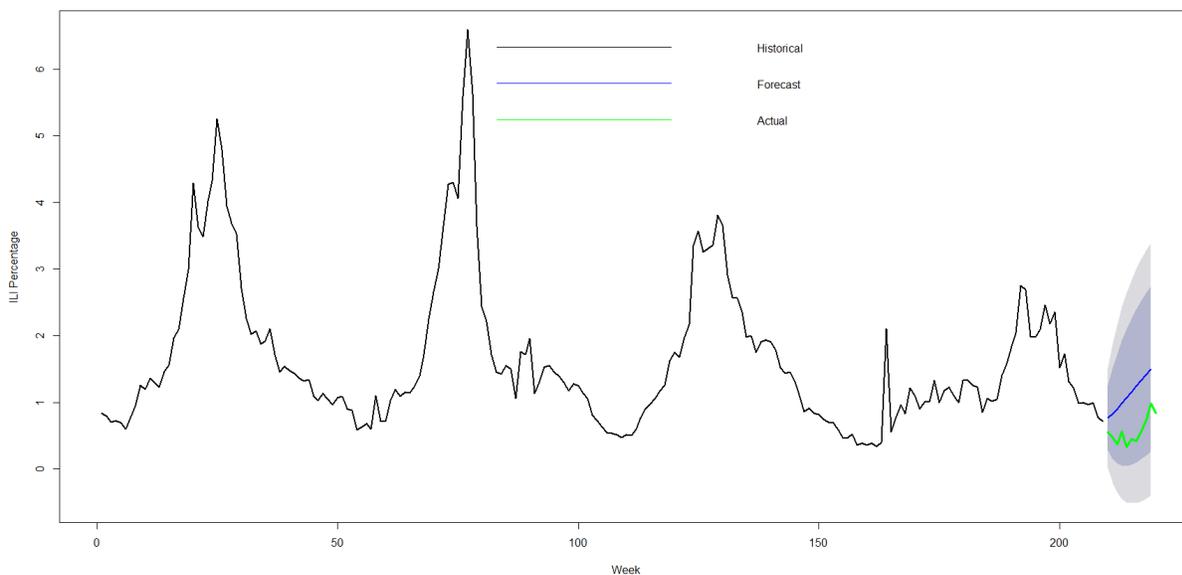
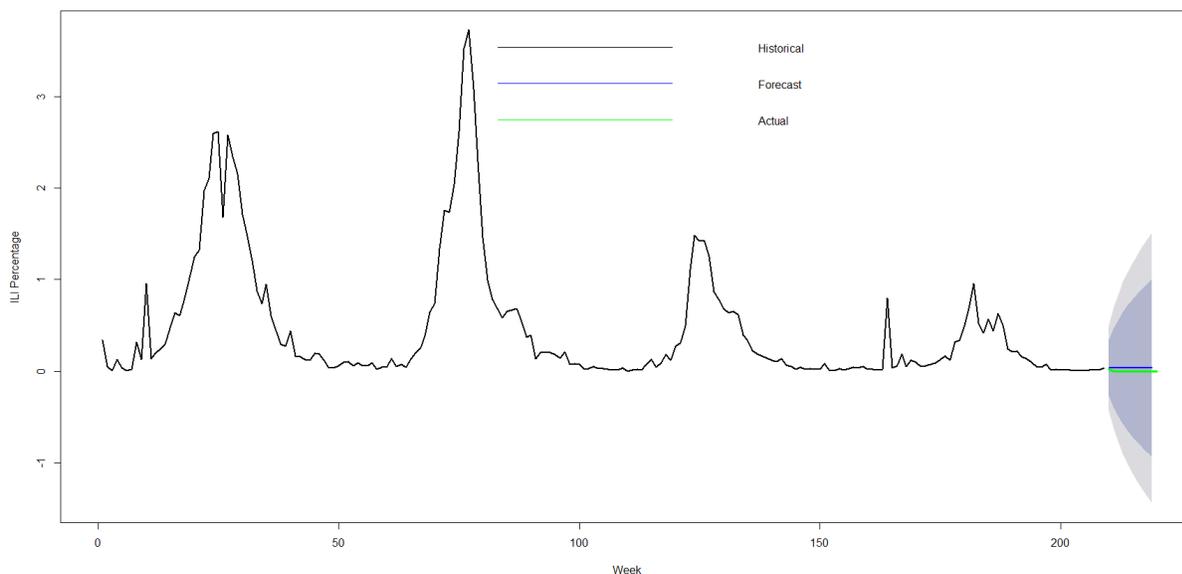


Figure 2-C ILINet Forecasted Estimates for Weeks 210-219 Using ILI Percentages from Weeks 1-209



Week refers to the range from the 26th week of 2012 through the 37th week of 2016. Each graph represents historical ILI percentages from MMWR week 27 of 2012 through MMWR week 26 of 2016 in black, forecasted estimates for MMWR weeks 27-37 of 2016 in blue, and actual ILI percentages for MMWR weeks 27-37 of 2016 in green. The dark blue shaded area represents the 80% confidence interval and the gray shaded area represents the 95% confidence interval. ILINet forecasted values (Figure 2-C, Table 2) are closer to actual ILI percentages as evidenced by smaller differences between actual and estimated values when compared to hospital EC visits data (Figure 2-B, Table 2) and FNY (Figure 2-A, Table 2) data.

Discussion

Though several surveillance systems have been designed to provide warning, few provide reliable data in near-real time, and fewer still have demonstrated the capability to provide advanced forecasting of impending influenza cases [1,18]. Our study recorded varying peak week, peak ILI %, and mean over the period from 2012-2016 with resultant effects on the ILI forecast estimates by data capture mechanisms. This is consistent with previous findings where wide variability in local and regional level data for ILI have been reported [19]. This may reflect the voluntary nature of influenza activity reporting by public health partners and health-care providers.

We noted high inconsistency and variability in the number of users who self-reported using FNY (mean (SD): 69 (41) weekly responses). This may be associated with the fact that there were fewer FNY users in the Houston area during the initial stages of the tool release. However, as awareness of participatory surveillance increased, the FNY data decreased in variability as the number of users reporting increased. Monitoring of online search trends and the use of online self-reporting illness tools such as FNY appear to have potential for supporting and enhancing traditional syndromic surveillance systems. Internet-based participatory syndromic surveillance programs

have been used to signify the start or end of transmission of seasonal pathogens [20], provide situational awareness, aid in the epidemiologic description of a disease, and to perform surveillance during a mass gathering [21].

We noted that ILI showed consistent decreases in peak ILI% and mean based on the HHD syndromic surveillance data during the study period. Although the forecasted estimates indicated a substantial difference compared to actual ILI percentages for the same period, both followed the same trend direction. EC chief complaint data have been used in the early detection of disease clusters or outbreaks [22, 23], and in the identification of the start or end of the season for pathogens such as influenza [24, 25]. For instance, the EC data had flagged the start of influenza season 2 to 3 weeks earlier than the CDC ILINet data in 2 of 3 eligible years examined [24].

None of the data sources accurately forecasted future ILI in the population due mainly to the wide variability recorded except for ILINet which exhibited the least difference with forecasted estimate close to the actual ILI report. Our results were similar to that of the CDC 2013–2014 Influenza Season Challenge where using a variety of digital data sources and methods nine of the competing teams failed to accurately forecast all influenza season milestones [26]. Our predictions were based on historical data and may not be ideal to predict unusual events that have not occurred previously.

Forecasting also must consider the noise component which is difficult to estimate or predict. While the times series in our study represent occurrence of ILI at the population level, it should be recognized that background noise is a common phenomenon that occurs in public health surveillance data. It is unclear in our study the extent to which noise affected the predicted models. In addition, forecasting time series data is difficult due to inherent uncertainties of trend and retention of historical properties. However, considering the degree of uncertainty, short-term estimates based on these data sources could be useful for preparation or to increase level of awareness in the Houston community. Similarly, the CDC noted during the influenza season challenge that it was possible to obtain reasonably accurate forecasts in the short term compared to long term [26]. Further model fitting and forecasting should consider including specific seasonal components like seasonal ARIMA models to capture specific seasonal patterns. This method can be replicated in other software that performs ARIMA. As an example, for public health surveillance, the authors described the application of forecasting methods and results from data on ILI.

Limitations

There is need to highlight some limitations that may be associated with our study outcomes. First, we were not able to include influenza laboratory results in our analysis. This may have significant impact on the accuracy of using ILI forecasted results to estimate influenza activity. Second, even though hospital EC visits data does provide a true measure of ILI cases as seen in the seasonal peaks every year which match national peaks and often coincide with the detection of influenza related outbreaks in group settings, the hospital EC visits may represent patients that access EC as their first step towards medical care or treatment, as their decisions are based in part on the severity of their symptoms. Third, since all influenza activity reporting by public health partners and health-care providers is voluntary, it is difficult to maintain a steady and consistent flow of data for ILINet and FNY during each flu season. Fourth, the meaningfulness and usefulness of predictions depend

heavily on the quality and consistency of the source of ILI data. Established surveillance data sources are more likely to produce reliable estimates due to consistency and large volume of historical data available compared to newer, untested data. Finally, while information obtained from these diverse syndromic surveillance systems is routinely used to determine influenza activity's magnitude and timing, the authors acknowledge that further work is needed before a prediction of ILI activity can be made with absolute confidence.

Strengths

Despite these limitations, we believe that our current findings highlight the inherent potentials of the three diverse data capture mechanisms for monitoring of seasonal influenza which can inform activities for prevention and control and pandemic preparedness in the Houston metropolitan area. Our study shows promise that forecasting methods can be integrated into routine public health surveillance practice. The availability of multiple years of historical ILI data helps identify overall trends and unusual activity with more reliability. The data sources provide proportion (percent) of ILI per week which offers more information for population level surveillance compared to using only counts. Also, the ARIMA forecasting with 'forecast' package within RStudio Pro makes it possible to conduct prediction of ILI occurrence within the BioSense platform. The standard features for ESSENCE do not include prediction modeling. The utility of automated real-time syndromic surveillance systems increases when forecasting component is combined with statistical detection algorithms based on change detection applied to times series. While the extent of the biases suggested by this analysis cannot be known precisely, combining these data sources may give a true picture of the influenza activity in Houston. The authors believe that there is value in the availability of participatory surveillance such as FNY not only because of the utility for tracking and monitoring ILI trends with descriptive methods but also because of the platform provides flu news, vaccine availability, and real-time maps with ILI activity [27]. These methods can be applied to ILI data sources in various geographic areas. Nonetheless, continuous improvement in the quantity and quality of data obtained from these data sources may help enhance the forecasting capabilities and early detection of potential surge in influenza activity.

Conclusions

Our study concludes that the variations in the predictive estimates were directly associated with the data sources. Forecasted estimates rely heavily on sources consistently reporting enough responses to reduce variability. Forecasts based on data from FNY and other user-reported platforms will require enough users from the population to consistently report so that historical ILI data can be a better representation of overall influenza-like illness activity. This could lead to better estimations of influenza activity. ILINet forecasted estimates and the actual ILI reports exhibited the least difference though all sources generally had wide confidence intervals. Similar surveillance systems have the potential to reasonably estimate overall trend and estimates of influenza activity. Our results demonstrate that there are significant opportunities to improve the forecasting performance and that selective superiorities among the three data sources could be leveraged upon to improve ILI surveillance in Houston. For instance, combining the data sources with enough observations to forecast future trends may provide estimates that are closer to actual influenza activity in the Houston community. These estimates could be used to empower decision

makers to lead and manage more effectively by providing timely and useful evidence for the public health surveillance to prepare and respond.

Acknowledgements

The HHD would like to acknowledge the Flu Near You team at HealthMap for their work on the dashboard and marketing material for Flu Near You, for coordinating the updates on Flu Near You for end users, which include trends of ILI across the country, and providing comments regarding the manuscript.

The HHD would like to acknowledge the many groups and individuals that conduct the operational component for seasonal influenza surveillance at HHD.

The ILINet data was obtained for Houston jurisdiction with permission from Texas DSHS.

Financial Disclosure

The HHD receives funding for this project through a joint organizational partnership between the Council of State and Territorial Epidemiologists (CSTE) and the Skoll Global Threats Fund (SGTF), provided by CSTE through sub-awards from the SGTF award #15-03126. The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the SGTF, CSTE, or the HHD.

Competing Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

1. Qualls N, Levitt A, Kanade N, et al. 2017. Community Mitigation Guidelines to Prevent Pandemic Influenza — United States, 2017. *MMWR Recomm Rep.* 66(No. RR-1), 1-34. doi:<https://doi.org/10.15585/mmwr.rr6601a1>. [PubMed](#)
2. Chretien J-P, George D, Shaman J, Chitale RA, McKenzie FE. 2014. Influenza Forecasting in Human Populations: A Scoping Review. *PLoS One.* 9(4), e94130. doi:<https://doi.org/10.1371/journal.pone.0094130>. [PubMed](#)
3. Dorjee S, Poljak Z, Revie CW, Bridgland J, McNab B, et al. 2013. A review of simulation modelling approaches used for the spread of zoonotic influenza viruses in animal and human populations. *Zoonoses Public Health.* 60, 383-411. doi:<https://doi.org/10.1111/zph.12010>. [PubMed](#)
4. Lee VJ, Lye DC, Wilder-Smith A. 2009. Combination strategies for pandemic influenza response - a systematic review of mathematical modeling studies. *BMC Med.* 7, 76. doi:<https://doi.org/10.1186/1741-7015-7-76>. [PubMed](#)

5. Prieto DM, Das TK, Savachkin AA, Uribe A, Izurieta R, et al. 2012. A systematic review to identify areas of enhancements of pandemic simulation models for operational use at provincial and local levels. *BMC Public Health*. 12, 251. doi:<https://doi.org/10.1186/1471-2458-12-251>. [PubMed](#)
6. Coburn BJ, Wagner BG, Blower S. 2009. Modeling influenza epidemics and pandemics: insights into the future of swine flu (H1N1). *BMC Med*. 7, 30. doi:<https://doi.org/10.1186/1741-7015-7-30>. [PubMed](#)
7. Khuwaja S, Mgbere O, Awosika-Olumo A, Momin F, Ngo K. 2011. Using Sentinel Surveillance System to Monitor Seasonal and Novel H1N1 Influenza Infection in Houston, Texas: Outcome Analysis of 2008–2009 Flu Season. *J Community Health*. 36(5), 857-63. doi:<https://doi.org/10.1007/s10900-011-9386-2>. [PubMed](#)
8. Mgbere O, Ngo K, Khuwaja S, Mouzoon M, Greisinger A, et al. 2017. Pandemic-related health behavior: repeat episodes of influenza-like illness related to the 2009 H1N1 influenza pandemic. *Epidemiol Infect*. 2017;145 (12):2611-2617. doi:10.1017/S0950268817001467. [PubMed](#)
9. Chakraborty P, Khadivi P, Lewis B, Mahendiran A, Chen J, et al. Forecasting a Moving Target: Ensemble Models for ILI Case Count Predictions. 2014. Proceedings of the 2014 SIAM
10. Smolinski MS, Crawley AW, Baltrusaitis K, Chunara R, Olsen JM, et al. 2015. Flu Near You: Crowdsourced Symptom Reporting Spanning 2 Influenza Seasons. *Am J Public Health*. 105(10), 2124-30. Epub Aug 2015. doi:<https://doi.org/10.2105/AJPH.2015.302696>. [PubMed](#)
11. Chunara R, Goldstein E, Patterson-Lomba O, Brownstein JS. 2015. Estimating influenza attack rates in the United States using a participatory cohort. *Sci Rep*. 5, 9540. doi:<https://doi.org/10.1038/srep09540>. [PubMed](#)
12. City of Houston. (2017). About Houston. Retrieved from <http://www.houstontx.gov/about/houston/houstonfacts.html> on June 21st, 2017.
13. Baltrusaitis K, Santillana M, Crawley AW, Chunara R, Smolinski M, Brownstein JS. (2017) Determinants of Participants' Follow-Up and Characterization of Representativeness in Flu Near You, A Participatory Disease Surveillance System. *JMIR Public Health Surveill*. Apr 7;3(2):e18. doi: 10.2196/publichealth.7304.
14. Allison P. (2015). Imputation by Predictive Mean Matching: Promise & Peril. *Statistical Horizons*. Retrieved from <http://statisticalhorizons.com/predictive-mean-matching> on April 1st, 2017.
15. Hyndman RJ, Athanasopoulos G. (2014). Forecasting: principles and practice. Otexts.com. Retrieved from <https://www.otexts.org/fpp/8> on April 1st, 2017.

16. Yurekli K, Kurunc A, Ozturk F. 2005. Application of linear stochastic models to monthly flow data of Kelkit Stream. *Ecol Modell.* 183, 67-75. <https://doi.org/10.1016/j.ecolmodel.2004.08.001>
17. Hu W, Tong S, Mengersen K, Connell D. 2007. Weather variability and the incidence of cryptosporidiosis: Comparison of time series Poisson regression and SARIMA models. *Ann Epidemiol.* 17, 679-88. [PubMed https://doi.org/10.1016/j.annepidem.2007.03.020](https://doi.org/10.1016/j.annepidem.2007.03.020)
18. Dugas AF, Hsieh YH, Levin SR, Pines JM, Mareiniss DP, et al. 2012. Google Flu Trends: correlation with emergency department influenza rates and crowding metrics. *Clin Infect Dis.* 54, 463-69. [PubMed https://doi.org/10.1093/cid/cir883](https://doi.org/10.1093/cid/cir883)
19. Centers for Disease Control and Prevention. Overview of Influenza Surveillance in the United States - Outpatient Influenza-like Illness Surveillance Network (ILINet) Retrieved from <https://www.cdc.gov/flu/weekly/overview.htm> on June 22, 2017.
20. Dalton C, Durrheim D, Fejsa J, Francis L, Carlson S, et al. 2009. Flutracking: A weekly Australian community online survey of influenza-like illness in 2006, 2007 and 2008. *Commun Dis Intell.* 33(3), 316-22. [PubMed](https://doi.org/10.1017/S095026880999149X)
21. Sugiura H, Ohkusa Y, Akahane M, Sugahara T, Okabe N, et al. 2010. Construction of syndromic surveillance using a web-based daily questionnaire for health and its application at the G8 Hokkaido Toyako Summit meeting. *Epidemiol Infect.* 138(10), 1493-502. [PubMed https://doi.org/10.1017/S095026880999149X](https://doi.org/10.1017/S095026880999149X)
22. Heffernan R, Mostashari F, Das D, Karpati A, Kulldorff M, et al. 2004. Syndromic surveillance in public health practice, New York City. *Emerg Infect Dis.* 10(5), 858-64. [PubMed https://doi.org/10.3201/eid1005.030646](https://doi.org/10.3201/eid1005.030646)
23. Zheng W, Aitken R, Muscatello DJ, Churches T. 2007. Potential for early warning of viral influenza activity in the community by monitoring clinical diagnoses of influenza in hospital emergency departments. *BMC Public Health.* 7, 250. [PubMed https://doi.org/10.1186/1471-2458-7-250](https://doi.org/10.1186/1471-2458-7-250)
24. May LS, Griffin BA, Bauers NM, Jain A, Mitchum M, et al. 2010. Emergency department chief complaint and diagnosis data to detect influenza-like illness with an electronic medical record. *West J Emerg Med.* 11(1), 1-9. [PubMed](https://doi.org/10.1186/1471-2458-9-483)
25. Griffin BA, Jain AK, Davies-Cole J, Glymph C, Lum G, et al. 2009. Early detection of influenza outbreaks using the DC Department of Health's syndromic surveillance system. *BMC Public Health.* 9, 483. [PubMed https://doi.org/10.1186/1471-2458-9-483](https://doi.org/10.1186/1471-2458-9-483)
26. Biggerstaff M, Alper D, Dredze M, et al. 2016. Results from the Centers for Disease Control and Prevention's predict the 2013–2014 Influenza Season Challenge. *BMC Infect Dis.* 16, 357. Published online Jul 2016. doi:<https://doi.org/10.1186/s12879-016-1669-x>. [PubMed](https://doi.org/10.1186/s12879-016-1669-x)
27. Baltrusaitis K, Santillana M, Crawley AW, Chunara R, Smolinski M, et al. Determinants of Participants' Follow-Up and Characterization of Representativeness in Flu Near You, A

Participatory Disease Surveillance System. JMIR Public Health Surveill 2017;3(2):e18.
URL: <https://publichealth.jmir.org/2017/2/e18> DOI: 10.2196/publichealth.7304 PMID:
28389417. PMCID: 5400887