

This is a repository copy of *Results from the centers for disease control and prevention's predict the 2013-2014 Influenza Season Challenge*.

White Rose Research Online URL for this paper:
<https://eprints.whiterose.ac.uk/157285/>

Version: Published Version

Article:

Fox, S. (2016) Results from the centers for disease control and prevention's predict the 2013-2014 Influenza Season Challenge. *BMC Infectious Diseases*. 357 (2016). ISSN 1471-2334

<https://doi.org/10.1186/s12879-016-1669-x>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

RESEARCH ARTICLE

Open Access



Results from the centers for disease control and prevention's predict the 2013–2014 Influenza Season Challenge

Matthew Biggerstaff^{1*}, David Alper², Mark Dredze³, Spencer Fox⁴, Isaac Chun-Hai Fung⁵, Kyle S. Hickmann^{6,7}, Bryan Lewis⁸, Roni Rosenfeld⁹, Jeffrey Shaman¹⁰, Ming-Hsiang Tsou¹¹, Paola Velardi¹², Alessandro Vespignani¹³, Lyn Finelli¹ and for the Influenza Forecasting Contest Working Group

Abstract

Background: Early insights into the timing of the start, peak, and intensity of the influenza season could be useful in planning influenza prevention and control activities. To encourage development and innovation in influenza forecasting, the Centers for Disease Control and Prevention (CDC) organized a challenge to predict the 2013–14 United States influenza season.

Methods: Challenge contestants were asked to forecast the start, peak, and intensity of the 2013–2014 influenza season at the national level and at any or all Health and Human Services (HHS) region level(s). The challenge ran from December 1, 2013–March 27, 2014; contestants were required to submit 9 biweekly forecasts at the national level to be eligible. The selection of the winner was based on expert evaluation of the methodology used to make the prediction and the accuracy of the prediction as judged against the U.S. Outpatient Influenza-like Illness Surveillance Network (ILINet).

Results: Nine teams submitted 13 forecasts for all required milestones. The first forecast was due on December 2, 2013; 3/13 forecasts received correctly predicted the start of the influenza season within one week, 1/13 predicted the peak within 1 week, 3/13 predicted the peak ILINet percentage within 1 %, and 4/13 predicted the season duration within 1 week. For the prediction due on December 19, 2013, the number of forecasts that correctly forecasted the peak week increased to 2/13, the peak percentage to 6/13, and the duration of the season to 6/13. As the season progressed, the forecasts became more stable and were closer to the season milestones.

Conclusion: Forecasting has become technically feasible, but further efforts are needed to improve forecast accuracy so that policy makers can reliably use these predictions. CDC and challenge contestants plan to build upon the methods developed during this contest to improve the accuracy of influenza forecasts.

Keywords: Influenza, Forecasting, Prediction, Modeling

Background

Each year annual seasonal epidemics of influenza occur in the United States; however, these seasonal epidemics vary in their timing and intensity [1–3]. Preparing for and responding appropriately to influenza epidemics and pandemics is a critical function of public health. Traditionally, planning and response have relied on

surveillance data to provide situational awareness [3–5] and information on historic experiences to inform qualitative judgments about what may happen next. A promising new approach has emerged that could help provide a timelier and systematic foundation for public health decision-making: infectious disease forecasting. Infectious disease forecasting combines traditional and internet-derived data on influenza activity with novel mathematical strategies to forecast the progression of an epidemic over a season. These forecasts can provide information for early public health action, such as

* Correspondence: mbiggerstaff@cdc.gov

¹Epidemiology and Prevention Branch, Influenza Division, Centers for Disease Control and Prevention, Atlanta, Georgia, USA

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

targeting resources for influenza prevention and control and communicating prevention messages to the public. Initial work describing influenza forecasting methods was promising [6, 7].

To better understand influenza forecasts and improve their usefulness to public health decision making, the Centers for Disease Control and Prevention (CDC) organized a flu forecasting challenge with the primary objectives of 1) examining the accuracy of influenza forecasts, 2) increasing interest in influenza forecasting, and 3) improving the utility of influenza forecasts. CDC also wanted to encourage researchers to utilize novel sources of digital surveillance data (e.g., Twitter data, internet search query data, internet-based surveys) to make their forecasts and connect forecasts to public health decision making. The development of better forecasting models would help CDC improve its ability to monitor influenza in the United States and ultimately improve the prevention and control of influenza.

On November 25, 2013, CDC announced the Predict the Influenza Season Challenge [8]. Challenge contestants were asked to forecast the start, peak, and intensity of the 2013–2014 influenza season at the national level and at any or all Health and Human Services (HHS) region level(s) in the United States. Contestants were free to use any mathematical or statistical model that utilized digital surveillance data. In this report, we present the aggregated results and the lessons learned from the challenge.

Methods

The requirements and criteria for the selection of the winner have been described more fully in the Federal Register Notice announcing the challenge on November 23, 2013 [8]. Briefly, the challenge period ran from December 1, 2013–March 27, 2014, and contestants were required to submit a total of 9 biweekly forecasts over the challenge period that contained forecasts for the start, peak, length, and intensity of the 2013–2014 influenza season at the national level to be eligible for judging. Teams had to use a form of digital surveillance data (e.g., Twitter data, mining internet search term data, Internet-based surveys) as part of their forecasts but could use other data sources, including those of traditional influenza surveillance systems (e.g. the U.S. Outpatient Influenza-like Illness Surveillance Network [ILINet]). A team's submission also had to include a narrative describing the methodology of the forecasting model. The forecasting methodology could be changed during the course of the contest, but teams had to submit an updated narrative describing the changes.

All milestones were compared to ILINet. The current ILINet system began during the 1997–1998 influenza season, and ILINet has since demonstrated the ability to provide accurate information on the timing and impact

of influenza activity each season [9]. ILINet consists of more than 2,900 outpatient healthcare providers around the country who report data to CDC weekly on the number of patients with influenza-like illness (ILI) and the total number of patients seen in their practices [5]. ILINet data are based on a reporting week that starts on Sunday and ends on Saturday of each week; data are reported through the FluView surveillance report the following Friday [4]. Therefore, the most current ILINet data can lag the calendar date by 1–2 weeks.

We defined the start of the season as the first surveillance week in ILINet where the weighted number of visits for ILI divided by the total number of patient visits (the ILINet percentage) was above the national baseline value of 2.0 % and remained there for at least two additional weeks. We defined the peak week of the season as the surveillance week that the ILINet percentage was the highest. Two values were used to measure the intensity of the influenza season. The first was season duration, which was defined as the number of weeks that the ILINet percentage remained above baseline. The second was the highest numeric value that the ILINet percentage reached in the United States [4]. Weeks for the contest were defined by Morbidity and Mortality Weekly Report (MMWR) surveillance weeks; the MMWR calendar is available at <http://www.cdc.gov/nndss/script/downloads.aspx>. Contestants were also eligible to submit milestone forecasts for any or all of the 10 HHS regions to add to their final scores.

Forecasts were considered accurate if they were within 1 week or one percent of the actual value calculated from ILINet. The selection of the winner for this challenge was based on an evaluation of the methodology used to make the forecast and the accuracy of the forecast. Contestant submissions were judged by a panel of reviewers that included two CDC staff outside the Influenza Division and one faculty member from a noncompeting university. Judges scored submissions on a scale of 0 to 100 points using the following criteria: the strength of the methodology (25 points), which assessed how clearly the results and uncertainty in the forecasts were presented and how the data sources and forecast methodology were described; the accuracy, timeliness, and reliability of the forecasts for the start, peak week, and intensity of the influenza season (65 points); and the scope of the geography (US plus one or more HHS Regions) that the source data represented (10 points). Up to 50 bonus points were awarded to any contestant that submitted forecasts for the 10 HHS regions; the number of bonus points was based on the number of regions with a forecast and the strength of the methodology and the accuracy, timeliness, and reliability of the forecasts. The winner was awarded a cash prize of \$75,000. The results presented in this analysis are based on the forecasts from teams that met the eligibility criteria.

Results

Sixteen individuals or teams initially registered for the challenge, 15 entered at least one forecast, 11 submitted nine biweekly forecasts, and 9 submitted forecasts for all required milestones and are included in this report. The majority of teams used Twitter ($n = 6$ teams) and/or Google Flu Trends data ($n = 5$ teams) as a data source to inform their forecasting models. Teams also utilized digital data sources such as Wikipedia search inquiries and HealthMap data; 3/9 (33 %) teams utilized more than one digital data source (Table 1). Five out of 9 (56 %) teams employed statistical methods like time series analysis and generalized linear models, and 4/9 (44 %) employed mechanistic models that incorporated compartmental modeling (e.g., Susceptible-Exposed-Infected-Recovered [SEIR] models) (Table 1). Eight out of 9 teams made forecasts for at least one HHS region during the challenge period (Table 1). Two teams provided multiple forecasts using distinct methods or data sources as part of their submission. A total of 13 forecasts were evaluated over the contest period.

Based on values from ILINet, the 2013–14 influenza season in the United States began on MMWR week 48 (late November), peaked on week 52 (late December) at 4.6 %, and lasted for 14 consecutive weeks (Table 2). Virologic data indicated that pH1N1 viruses predominated nationally and in all 10 regions for the majority of the influenza season but that influenza B viruses became the predominant virus nationally in week 13 (late March) [3]. The median submitted milestone forecasts are

shown for the United States in Table 3 and for the 10 HHS regions in Additional file 1: Tables S1–S10. The 13 biweekly forecasts of the milestones for the United States submitted by the 9 teams (the grey lines) are presented in Figs. 1, 2, 3 and 4 along with the milestones as calculated from ILINet (the black line). The first forecast was due on December 2, 2013, when ILINet data for surveillance week 46 were available. The median value of the 13 forecasts received for the start of the influenza season was week 50 (corresponding to the calendar week beginning December 8, 2013), for the week that the ILINet percentage would peak was week 5 (the week beginning January 26, 2014), for the ILINet peak was 3.45 %, and for the median forecasted season duration was 13 weeks (Table 3). When compared to the ILINet results, 3/13 (23 %) individual forecasts correctly forecasted the start of the influenza season within one week, 1/13 (8 %) correctly forecasted the peak within 1 week, 3/13 (23 %) correctly forecasted the peak ILINet percentage within 1 %, and 4/13 (31 %) correctly forecasted the duration within 1 week (Table 4); national-level accuracy results for the 13 predictions received by each milestone are available in Additional file 1: Tables S11–S14.

For the forecast submitted on December 19, when ILINet data from surveillance week 49 were available, the contestants adjusted their models based on updated ILINet and digital data and the median forecast for the peak week shifted from week 5 to week 2 (the week beginning January 5, 2014). The median forecast for the

Table 1 Characteristics of nine teams that competed in the Predict the 2013–14 Influenza Season Challenge

Team	Digital Data source	Model type	Regional forecast ^a	Brief description ^d
A	Wikipedia	mechanistic ^b	Yes	Susceptible-Exposed-Infected-Recovered (SEIR) model using data assimilation to probabilistically fit models to ILINet data
B	Twitter	mechanistic	Yes	SEIR model initialized with current Twitter and ILINet data
C	Google Flu Trends; Twitter	statistical ^c	Yes	Utilized method of analogues, Kalman filtering, Poisson regression, and an ensemble method averaging the results of the three models to forecast ILINet
D	Google Flu Trends	statistical	Yes	Utilized empirical Bayes model and a spatio-temporal likelihood function
E	Google Flu Trends; Twitter	statistical	Yes	Utilized multiplicative time series model
F	Google Flu Trends	mechanistic	Yes	Susceptible-infected-recovered-susceptible (SIRS) model initialized with Google Flu Trends data and data assimilation methods
G	Twitter	statistical	No	Extrapolation of filtered Twitter data
H	Google Flu Trends; HealthMap; Twitter	mechanistic	Yes	Statistical models used to make short term forecasts and agent based models combined with mean field models with non-linear optimization techniques used to output long term forecasts.
I	Twitter	statistical	Yes	Utilized time series model and method of analogues

^aYes denotes forecast for ≥ 1 region (for all weeks)

^bIncludes models that incorporate compartmental modeling like Susceptible-Exposed-Infected-Recovered [SEIR] models

^cIncludes models like time series analysis and generalized linear models

^dAdditional information on methodology and results for select teams available in references [34–38]

Table 2 Forecasting targets for the 2013–2014 influenza season as calculated from the U.S. Outpatient Influenza-like Illness Surveillance Network (ILINet), United States

	Start week ^a	Peak week ^a	Peak percentage	Duration of influenza season
United States	48	52	4.6	14

Legend: The start of the season was defined as the first surveillance week in ILINet where the number of visits for ILI divided by the total number of patient visits (the ILINet percentage) was above the national baseline value of 2.0 % and remained there for at least two additional weeks. The peak week of the season was defined as the surveillance week that the ILINet percentage was the highest during the 2013–14 influenza season. The ILINet percent peak was defined as the highest numeric value that the ILINet percentage reached in the United States during the 2013–14 influenza season. The duration was defined as the number of weeks that the ILINet percentage remained above the national baseline.

^aWeeks are given in Morbidity and Mortality Weekly Report surveillance weeks. For calendar start and end dates of each week, please see <http://www.cdc.gov/nndss/script/downloads.aspx>

peak ILINet percentage increased to 4.48 %, and the median forecast for season duration increased to 14 weeks. The number of submissions that correctly forecasted the peak week increased to 2/13 (15 %), the peak percentage to 6/13 (46 %), and the duration of the season to 6/13 (46 %). As the season progressed, the median forecasts for peak week, peak percentage, and peak duration for the United States became more stable and a majority of the 13 predictions converged on the season milestones calculated from ILINet (Tables 2, 3 and 4; Figs. 1, 2, 3, 4).

After a review of the forecasts and the results from the ILINet system, the judges found that Jeffrey Shaman's team from Columbia University was the overall winner of CDC's Predict the Influenza Season Challenge [10].

Discussion

This challenge represents the first nationally coordinated effort to forecast an influenza season in the United States. CDC organized the challenge to support the continued technical innovation required to forecast the influenza season. The results of this challenge indicate

that while forecasting has become technically feasible and reasonably accurate forecasts can be made in the short term, further work refining forecasting methodology and identifying the best uses of forecasts for public health decision making are needed. CDC continues to work with the researchers who participated in the challenge to refine and adapt methodology and determine the best uses of forecasts to inform public health decision making.

This forecasting challenge was useful to CDC in many ways. First, it promoted the development of forecasting models using data readily available through existing surveillance and digital data (e.g. Google Flu Trends) platforms. Second, it encouraged the connection between forecasters, subject matter experts, and public health decision-makers. During the challenge, forecasters needed support understanding the nuances of CDC's surveillance data while public health decision makers needed support understanding the different digital data sources and forecasting techniques. This challenge also incorporated forecasting milestones that would be most

Table 3 Median forecasted start week, peak week, peak percentage, and duration of the 2013–14 influenza season, by forecast date, United States ($n = 13$)

Date of forecast (Week of ILINet data availability ^{a,b})	Median forecasted start week ^b	Median forecasted peak week ^b	Median forecasted peak percentage	Median forecasted duration of influenza season
12/2/2013 (WK. 46)	50	5	3.5	13
12/19/2013 (WK. 49)	49	3	4.5	14
1/2/2014 (WK. 51)	48	3	4.5	15
1/16/2014 (WK.1)	48	2	4.9	14
1/30/2014 (WK. 3)	48	52	4.6	14
2/13/2014 (WK. 5)	48	52	4.6	13
2/27/2014 (WK. 7)	48	52	4.6	13
3/13/2014 (WK. 9)	48	52	4.6	14
3/27/2014 (WK. 11)	48	52	4.6	14

Legend: Forecasts presented here are from the 9 teams that successfully completed the CDC Predict the 2013–2014 Influenza Season Challenge

^aILINet data are based on a reporting week that starts on Sunday and ends on Saturday of each week, and data are reported out through the FluView surveillance report the following Friday. Therefore, the most current ILINet data can lag the calendar date by 1–2 weeks

^bWeeks are given in Morbidity and Mortality Weekly Report surveillance weeks. For calendar start and end dates of each week, please see <http://www.cdc.gov/nndss/script/downloads.aspx>

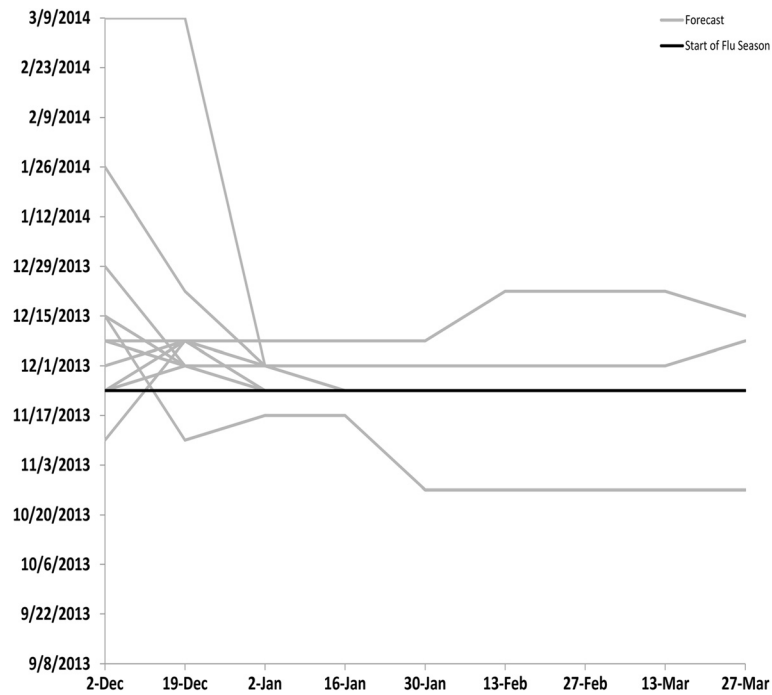


Fig. 1 Forecasted start week of the 2013–2014 influenza season, by forecast date, United States ($n = 13$). Forecasts presented here are from teams that successfully completed the CDC Predict the 2013–2014 Influenza Season Challenge. The start of the season was defined as the first surveillance week in ILINet where the number of visits for ILI divided by the total number of patient visits (the ILINet percentage) was above the national baseline value of 2.0 % and remained there for at least two additional weeks

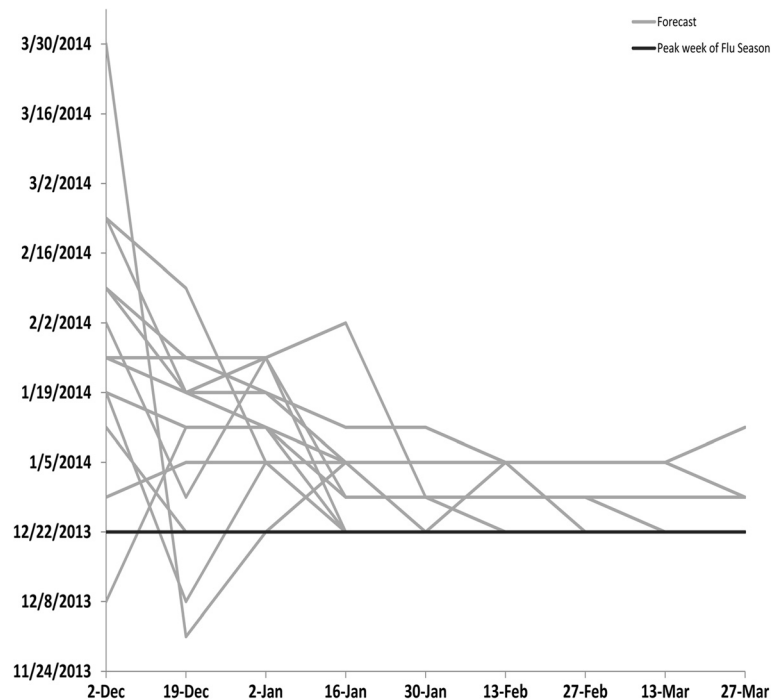


Fig. 2 Forecasted peak week of the 2013–2014 influenza season, by forecast date, United States ($n = 13$). Forecasts presented here are from teams that successfully completed the CDC Predict the 2013–2014 Influenza Season Challenge. The peak week of the season was defined as the surveillance week that the ILINet percentage was the highest during the 2013–14 influenza season

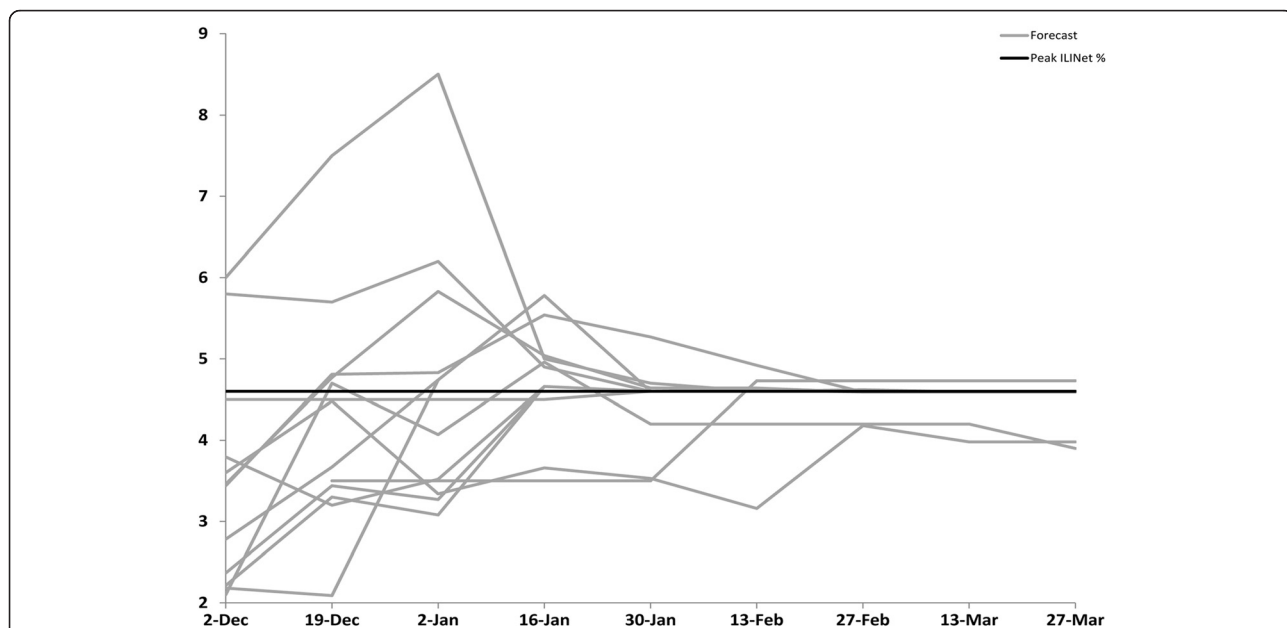


Fig. 3 Forecasted peak ILINet percent of the 2013–2014 influenza season, by forecast date, United States ($n = 13$). Forecasts presented here are from teams that successfully completed the CDC Predict the 2013–2014 Influenza Season Challenge. The ILINet percent peak was defined as the highest numeric value that the ILINet percentage reached in the United States during the 2013–14 influenza season

useful to public health decision makers in planning for influenza prevention and control activities, including the start week of the influenza season. This milestone was rarely included in previous efforts to forecast influenza but could inform public health action, including the development of tailored recommendations to the public

about vaccination timing and to physicians about influenza antiviral treatment [6, 7].

The results from the challenge indicated that no team was entirely accurate at forecasting all of the influenza season milestones. Public health actions informed by forecasts that later prove to be inaccurate can have

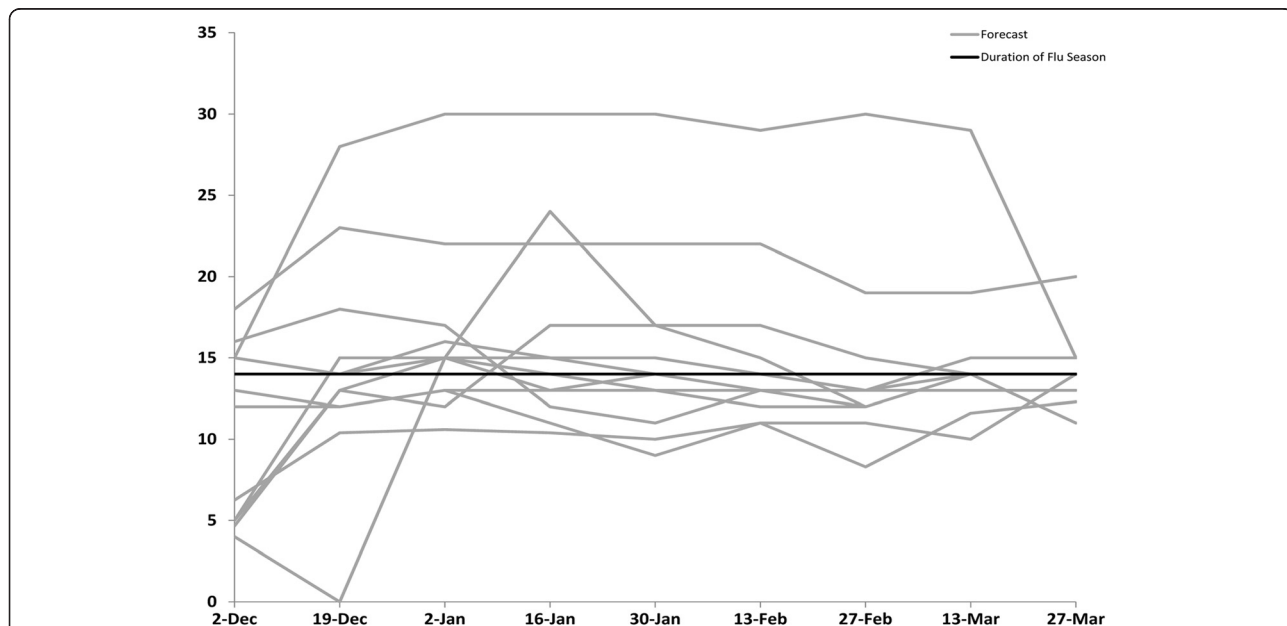


Fig. 4 Forecasted duration of the 2013–2014 influenza season, by forecast date, United States ($n = 13$). Forecasts presented here are from teams that successfully completed the CDC Predict the 2013–2014 Influenza Season Challenge. The duration was defined as the number of weeks that the ILINet percentage remained above the national baseline

Table 4 Forecasts within 1 week or percent of the start week, peak week, peak percentage, and duration of the 2013–14 influenza season, by forecast date, United States (n = 13)

Date of forecast (Week of ILINet data availability ^{a,b})	Start week	Peak week	Peak percentage	Duration of influenza season
12/2/2013 (WK. 46)	3 (23 %)	1 (8 %)	3 (23 %)	4 (31 %)
12/19/2013 (WK. 49)	6 (46 %)	2 (15 %)	6 (46 %)	6 (46 %)
1/2/2014 (WK. 51)	12 (92 %) ^c	2 (15 %)	5 (38 %)	7 (54 %)
1/16/2014 (WK.1)	12 (92 %)	6 (46 %)	10 (77 %)	6 (46 %)
1/30/2014 (WK. 3)	11 (85 %)	11 (85 %)	11 (85 %)	6 (46 %)
2/13/2014 (WK. 5)	11 (85 %)	10 (77 %)	12 (92 %)	6 (46 %)
2/27/2014 (WK. 7)	11 (85 %)	11 (85 %)	13 (100 %)	5 (38 %)
3/13/2014 (WK. 9)	11 (85 %)	11 (85 %)	13 (100 %)	9 (69 %)
3/27/2014 (WK. 11)	10 (77 %)	12 (92 %)	13 (100 %)	10 (77 %)

Legend: Forecasts presented here are from the 9 teams that successfully completed the CDC Predict the 2013–2014 Influenza Season Challenge. The start of the season was defined as the first surveillance week in ILINet where the number of visits for ILI divided by the total number of patient visits (the ILINet percentage) was above the national baseline value of 2.0 % and remained there for at least two additional weeks. The peak week of the season was defined as the surveillance week that the ILINet percentage was the highest during the 2013–14 influenza season. The ILINet percent peak was defined as the highest numeric value that the ILINet percentage reached in the United States during the 2013–14 influenza season. The duration was defined as the number of weeks that the ILINet percentage remained above the national baseline

^aILINet data are based on a reporting week that starts on Sunday and ends on Saturday of each week, and data are reported out through the FluView surveillance report the following Friday. Therefore, the most current ILINet data can lag the calendar date by 1–2 weeks

^bWeeks are given in Morbidity and Mortality Weekly Report surveillance weeks. For calendar start and end dates of each week, please

see <http://wwwn.cdc.gov/nndss/script/downloads.aspx>

^cLast forecast received before milestone observed in ILINet

negative consequences, including the loss of credibility, wasted and misdirected resources, and, in the worst case, increases in morbidity or mortality. There are at least two possible reasons for the lack of accuracy; the first is the use of digital data. Multiple published scientific studies show that digital surveillance data, such as Twitter, Wikipedia search queries, and Google Flu Trends data, have correlation with influenza activity as measured by existing traditional influenza surveillance programs and have been used successfully to track vaccine sentiment and monitor disease outbreaks [11–26], often in a timelier manner than traditional infectious disease surveillance data. However, not all digital data are equally accurate, and the algorithms and methodologies underpinning these data require constant upkeep to maintain their accuracy [27]. Several contestants in this challenge utilized Google Flu Trends to inform their forecasts, which has been shown to have overestimated influenza activity in recent seasons [28, 29]. Influenza forecasting models informed by digital data are subject to the biases and errors of their underlying source data, and few contestants in this study utilized more than one digital data set, which could make influenza forecasts more robust to data biases.

The second reason that forecasts may not have been accurate relates to the model methodologies and specifications themselves. While substantial work has advanced the field of infectious disease forecasting, it is still in the early years of development. Recent reviews of influenza forecasting found the use of a variety of statistical and

infectious disease modeling approaches, including time series models, SEIR models, and agent-based models [6, 7]. While the advantages and limitations of these approaches have been described, the impact on the relative accuracies of these approaches are unknown, and optimal forecasting methods are still under study [7]. The results of this contest, while helpful to gauge the general accuracy of influenza forecasting, cannot be used to recommend the best forecasting methodology because participants in this contest used different data sources and methodological approaches to make their forecasts. Forecasts could have been improved by the use of more precise digital data, better forecasting methodology, or both. Additionally, the challenge ran for only one influenza season. Because of the unpredictable timing and intensity of influenza seasons in the United States, assessments of forecasting methods and accuracy will require a review over multiple influenza seasons to ensure consistency in performance.

This challenge identified a number of barriers limiting forecasting model development and application, adoption by decision-makers, and the eventual public health impact of forecasts. First, interaction between model developers and public health decision-makers has been limited, leading to difficulties in identifying and specifying relevant prediction targets that would best inform decision-making. Second, data for making and evaluating predictions are often difficult to share, obtain, and interpret. Third, no common standards exist for evaluating models, either against each other or against metrics

relevant to decision-makers, making it difficult to evaluate the accuracy and reliability of forecasts. Lastly, the presentation of the forecast confidence varied between teams, making comparison and interpretation by decision makers difficult. CDC and challenge contestants continue to work together through collaborative challenges to forecast the 2014–15 and 2015–16 influenza seasons in order to improve data availability and interpretation, develop standardized metrics to assess forecast accuracy and standardized ways to communicate forecasts and their uncertainty, and identify the types of decisions best aided by forecasts [30]. The identified improvements will not only increase the utility of forecasts for public health decision making in influenza but will be relevant to the forecasting of other infectious diseases, which face similar challenges. The best practices and lessons learned from influenza have already been shared with other government-run infectious disease forecasting challenges [31, 32].

The Predict the Influenza Season Challenge represents a successful utilization of the COMPETES Act, which authorized U.S. government agencies to host challenges to improve government and encourage innovation [33]. By hosting this challenge, CDC was able to receive and evaluate influenza season forecasts from 9 teams based on a variety of digital data sources and methodologies [34–38]. These forecasts were submitted by teams that were affiliated with a diverse set of organizations including universities and private industry, and some of the teams had never produced an influenza forecast before participation in the challenge. The high number of forecasts received through this challenge is in contrast to the number of forecasts that would have been received if a more traditional method of outside engagement available at CDC was utilized (e.g., traditional contracts or grants). The challenge mechanism allowed CDC to establish the overall goal of accurately forecasting influenza season milestones without specifying the forecasting methodologies and allowed CDC to evaluate forecasts for accuracy and quality prior to awarding of the challenge prize.

Conclusions

CDC currently monitors influenza activity each year using routine influenza surveillance systems that do not forecast influenza activity [3]. To help promote the continued advancement of influenza forecasting, CDC held the first nationally coordinated challenge to forecast the influenza season in the United States. Nine teams representing academia, private industry, and public health made forecasts about the start, the duration, and the intensity of the 2013–14 influenza season using a variety of digital data sources and methods. While no team accurately forecasted all influenza season milestones, the

results of this challenge indicate that reasonably accurate forecasts can be made in the short term. Further work to refine and adapt forecasting methodology and identify the best uses of forecasts for public health decision making is required before the results of influenza forecasting can be used by policy makers and the public to inform the selection and implementation of prevention and control measures. Nevertheless, forecasting holds much promise for seasonal influenza prevention and control and pandemic preparedness.

Additional file

Additional file 1: Median predicted start week, peak week, peak % ILINet, and duration of season for the 10 HHS Regions, and national-level forecasts that were within 1 week or 1 percent of the start week, peak week, peak % ILINet, and duration of season, CDC's Predict the 2013–2014 Influenza Season Challenge. Tables S1–10 contains the median submitted milestone forecasts and the milestones as calculated from ILINet for the 10 HHS regions (Tables S1–S10). Tables S11–14 contain the national-level accuracy results for the 13 predictions received by each milestone. (DOCX 47 kb)

Abbreviations

CDC, centers for disease control and prevention; HHS, health and human services; ILI, influenza-like illness; ILINet, U.S. outpatient influenza-like illness surveillance network; MMWR, morbidity and mortality weekly report; SEIR, susceptible-exposed-infected-recovered; SIRS, susceptible-infected-recovered-susceptible; US, United States

Acknowledgements

We thank Dr. Lance Waller, Professor, Rollins School of Public Health, Emory University; Dr. Andrew Hill, Statistician, National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention; and Ben Lopman, Epidemiologist, National Center for Immunization and Respiratory Diseases, who served as judges for this contest. We also thank Carrie Reed; Lynette Brammer; Justin O'Hagan; Scott Epperson; Joseph Bresee; and Juliana Cyril, of the Centers for Disease Control and Prevention, for their help with the development of this contest and the judging criteria.

The members of the Influenza Forecasting Contest Working Group are as follows:

4Sight team: Prithwish Chakraborty¹, Patrick Butler¹, Pejman Khadivi¹, Naren Ramakrishnan¹, Bryan Lewis², Jiangzhuo Chen², Chris Barrett², Keith Bisset², Stephen Eubank², VS Anil Kumar², Kathy Laskowski², Kristian Lum², Madhav Marathe², Susan Aman³, John S Brownstein³, Ed Goldstein⁴, Sumiko R Mekar⁵, and Elaine O. Nsoesie⁶;

Basagni team: Stefano Basagni⁷, Paola Velardi⁸, Giovanni Stilo⁸, Francesco Gesualdo⁹, and Alberto E. Tozzi⁹;

Delphi team: Ronald Rosenfeld¹⁰, Logan Brooks¹⁰, David Farrow¹⁰, Sangwon Hyun¹⁰, and Ryan J. Tibshirani¹⁰;

Johns Hopkins University team: David Broniatowski¹¹, Mark Dredze¹², and Michael Paul¹²;

Los Alamos National Laboratory team: Sara Y Del Valle¹³, Alina Deshpande¹³, Geoffrey Fairchild¹³, Nicholas Generous¹³, Reid Priedhorsky¹³, Kyle S. Hickman¹⁴, and James M. Hyman¹⁴;

Northeastern team: Alessandro Vespignani¹⁵, Qian Zhang¹⁵, and Nicola Perra¹⁵;

Precision Health Data Institute team: David Alper¹⁶, Priyadarshini Chandra¹⁶, Hemchandra Kaup¹⁶, Ramesh Krishnan¹⁶, Satish Madhavan¹⁶, Ashirwad Markar¹⁶, and Bryanne Pashley¹⁶;

San Diego State University team: Ming-Hsiang Tsou¹⁷, Christopher Allen¹⁷, Anoshé Aslam¹⁷, and Anna Nagel¹⁷;

Shaman team: Jeffrey Shaman¹⁸, Wan Yang¹⁸, Alicia Karspeck¹⁹, and Marc Lipsitch²⁰;

University of Georgia-Georgia Southern University team: Zion Tsz-Ho Tse²¹, Yuchen Ying²¹, Issac Chun-Hai Fung²², Lurii Bakach²², Yi Hao²², Braydon J. Schable²², Jessica K. Sexton²², and Manoj Gambhir²³:

University of Texas at Austin team: Spencer Fox²⁴, Lauren Ancel Meyers²⁴, Rosalind Eggo²⁴, Jette Henderson²⁴, Anurekha Ramakrishnan²⁴, James Scott²⁴, Bismark Singh²⁴, Ravi Srinivasan²⁴, and Sam Scarpino²⁵:

¹Discovery Analytics Center, Department of Computer Science, Virginia Tech, Blacksburg, Virginia, USA;

²Network Dynamics and Simulation Science Laboratory, Virginia Bioinformatics Institute, Virginia Tech, Blacksburg, Virginia, USA;

³Childrens Hospital Informatics Program, Boston Childrens Hospital, Boston, Massachusetts; Department of Pediatrics, Harvard Medical School, Boston, Massachusetts;

⁴Harvard School of Public Health, Boston, Massachusetts;

⁵Childrens Hospital Informatics Program, Boston Childrens Hospital, Boston, Massachusetts;

⁶Network Dynamics and Simulation Science Laboratory, Virginia Bioinformatics Institute, Virginia Tech, Blacksburg, Virginia, USA; Childrens Hospital Informatics Program, Boston Childrens Hospital, Boston, MA; Department of Pediatrics, Harvard Medical School, Boston, Massachusetts;

⁷Northeastern University, Boston, Massachusetts;

⁸Sapienza University of Roma, Rome, Italy;

⁹Bambino Gesù Children Hospital, Rome, Italy;

¹⁰Carnegie Mellon University, Pittsburgh, Pennsylvania;

¹¹George Washington University, Washington, D.C.;

¹²Johns Hopkins University, Baltimore, Maryland;

¹³Los Alamos National Laboratory, Los Alamos, New Mexico;

¹⁴Tulane University, New Orleans, Louisiana; Los Alamos National Laboratory, Los Alamos, New Mexico;

¹⁵Northeastern University, Boston, Massachusetts;

¹⁶Everyday Health, New York City, New York;

¹⁷San Diego State University, San Diego, California;

¹⁸Columbia University, New York City, New York;

¹⁹National Center for Atmospheric Research, Climate & Global Dynamics Division, Boulder, Colorado;

²⁰Harvard School of Public Health, Boston, Massachusetts;

²¹University of Georgia, Athens, Georgia;

²²Georgia Southern University, Statesboro, Georgia;

²³Monash University, Melbourne, Australia;

²⁴University of Texas at Austin, Austin, Texas;

²⁵Santa Fe Institute, Santa Fe, New Mexico;

Funding

BL: Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under award number 5U01GM070694-13 and the Defense Threat Reduction Agency Comprehensive National Incident Management System Contract HDTRA1-11-D-0016-0001.

ICHF received salary support from the Centers for Disease Control and Prevention (15IPA1509134). This paper is not related to his CDC-funded projects.

JS: NIH grants GM100467, GM110748 and 1U54GM088558

MHT: This material is partially based upon work supported by the National Science Foundation under Grant No. 1416509, project titled "Spatiotemporal Modeling of Human Dynamics Across Social Media and Social Networks".

SF: Received funding through National Institute of General Medical Sciences MIDAS grant (U01GM087719).

Availability of data and materials

The data supporting the conclusions of this article are included within the article and as Additional_File_1_Tables_S1_S14.

Authors' contributions

MB led the data collation, analysis, and the writing of the article. DA, MD, SF, ICHF, KH, BL, RR, JS, MHT, PV, and AV led the forecasting teams and data collection, assisted with data interpretation, and reviewed multiple drafts of the article. LF contributed to the design of the study and reviewed multiple drafts of the article. All authors read and approved the final version of the manuscript.

Competing interests

The authors declare that they have no competing interests.

Ethics approval and consent to participate

This study did not involve human participants, data, or tissue. Institutional review board approval was not required.

Disclaimer

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views or official position of the National Science Foundation, the Centers for Disease Control and Prevention, or the National Institutes of Health.

Author details

¹Epidemiology and Prevention Branch, Influenza Division, Centers for Disease Control and Prevention, Atlanta, Georgia, USA. ²Everyday Health, New York, New York, USA. ³Johns Hopkins University, Baltimore, Maryland, USA.

⁴University of Texas at Austin, Austin, Texas, USA. ⁵Georgia Southern

University, Statesboro, Georgia, USA. ⁶Los Alamos National Laboratory, Los

Alamos, New Mexico, USA. ⁷Tulane University, New Orleans, Louisiana, USA.

⁸Virginia Tech, Blacksburg, Virginia, USA. ⁹Carnegie Mellon University,

Pittsburgh, Pennsylvania, USA. ¹⁰Columbia University, New York, New York,

USA. ¹¹San Diego State University, San Diego, California, USA. ¹²Sapienza

University of Roma, Rome, Italy. ¹³Northeastern University, Boston,

Massachusetts, USA.

Received: 1 December 2015 Accepted: 28 June 2016

Published online: 22 July 2016

References

- Centers for Disease Control and Prevention. Influenza activity—United States, 2012–13 season and composition of the 2013–14 influenza vaccine. *MMWR Morb Mortal Wkly Rep.* 2013;62(23):473–9.
- Centers for Disease Control and Prevention. Update: influenza activity - United States, 2011–12 season and composition of the 2012–13 influenza vaccine. *MMWR Morb Mortal Wkly Rep.* 2012;61(22):414–20.
- Epperson S, Blanton L, Kniss K, Mustaqim D, Steffens C, Wallis T, et al. Influenza activity - United States, 2013–14 season and composition of the 2014–15 influenza vaccines. *MMWR Morb Mortal Wkly Rep.* 2014; 63(22):483–90.
- Centers for Disease Control and Prevention: FluView Interactive [http://www.cdc.gov/flu/weekly/fluviewinteractive.htm] (2014). Accessed 8/11/2014.
- Centers for Disease Control and Prevention: Overview of Influenza Surveillance in the United States [http://www.cdc.gov/flu/weekly/overview.htm] (2014). Accessed 9/25/2014.
- Chretien JP, George D, Shaman J, Chitale RA, McKenzie FE. Influenza forecasting in human populations: a scoping review. *PLoS One.* 2014;9(4): e94130.
- Nsoesie EO, Brownstein JS, Ramakrishnan N, Marathe MV. A systematic review of studies on forecasting the dynamics of influenza outbreaks. *Influenza Other Respir Viruses.* 2014;8(3):309–16.
- Announcement of Requirements and Registration for the Predict the Influenza Season Challenge, 78 Fed. Reg. 70303–70305 (November 25, 2013).
- Brammer L, Blanton L, Epperson S, Mustaqim D, Bishop A, Kniss K, et al. Surveillance for influenza during the 2009 influenza A (H1N1) pandemic—United States, April 2009–March 2010. *Clin Infect Dis.* 2011;52 Suppl 1:S27–35.
- Centers for Disease Control and Prevention: CDC Announces Winner of the 'Predict the Influenza Season Challenge' [http://www.cdc.gov/flu/news/predict-flu-challenge-winner.htm] (2014). Accessed 3/16/2016.
- Broniatowski DA, Paul MJ, Dredze M. National and local influenza surveillance through Twitter: an analysis of the 2012–2013 influenza epidemic. *PLoS One.* 2013;8(12):e83672.
- Yuan Q, Nsoesie EO, Lv B, Peng G, Chunara R, Brownstein JS. Monitoring influenza epidemics in china with search query from baidu. *PLoS One.* 2013; 8(5):e64323.
- Mclver DJ, Brownstein JS. Wikipedia usage estimates prevalence of influenza-like illness in the United States in near real-time. *PLoS Comput Biol.* 2014;10(4):e1003581.

14. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. *Nature*. 2009;457(7232):1012–4.
15. Cook S, Conrad C, Fowlkes AL, Mohebbi MH. Assessing Google flu trends performance in the United States during the 2009 influenza virus A (H1N1) pandemic. *PLoS One*. 2011;6(8):e23610.
16. Polgreen PM, Chen Y, Pennock DM, Nelson FD. Using internet searches for influenza surveillance. *Clin Infect Dis*. 2008;47(11):1443–8.
17. Lamb A, Paul M, Dredze M. Separating Fact from Fear: Tracking Flu Infections on Twitter. North American Chapter of the Association for Computational Linguistics (NAACL), 2013; 789–795.
18. Paul MJ, Dredze M. You are what you Tweet: Analyzing Twitter for public health. Fifth International AAAI Conference on Weblogs and Social Media. 2011, 265–272.
19. Velardi P, Stilo G, Tozzi AE, Gesualdo F. Twitter mining for fine-grained syndromic surveillance. *Artif Intell Med*. 2014;61(3):153–63.
20. Salathe M, Bengtsson L, Bodnar TJ, Brewer DD, Brownstein JS, Buckee C, et al. Digital epidemiology. *PLoS Comput Biol*. 2012;8(7):e1002616.
21. Salathe M, Freifeld CC, Mekaru SR, Tomasulo AF, Brownstein JS. Influenza A (H7N9) and the Importance of Digital Epidemiology. *N Engl J Med* 2013; 369(5):401–4.
22. Salathe M, Khandelwal S. Assessing vaccination sentiments with online social media: implications for infectious disease dynamics and control. *PLoS Comput Biol*. 2011;7(10):e1002199.
23. Barboza P, Vaillant L, Mawudeku A, Nelson NP, Hartley DM, Madoff LC, et al. Evaluation of epidemic intelligence systems integrated in the early alerting and reporting project for the detection of A/H5N1 influenza events. *PLoS One*. 2013;8(3):e57252.
24. Chunara R, Andrews JR, Brownstein JS. Social and news media enable estimation of epidemiological patterns early in the 2010 Haitian cholera outbreak. *Am J Trop Med Hyg*. 2012;86(1):39–45.
25. Chunara R, Freifeld CC, Brownstein JS. New technologies for reporting real-time emergent infections. *Parasitology*. 2012;139(14):1843–51.
26. Larson HJ, Smith DM, Paterson P, Cumming M, Eckersberger E, Freifeld CC, et al. Measuring vaccine confidence: analysis of data obtained by a media surveillance system used to analyse public concerns about vaccines. *Lancet Infect Dis*. 2013;13(7):606–13.
27. Althouse BM, Scarpino SV, Meyers LA, Ayers JW, Bargsten M, Baumbach J, et al. Enhancing disease surveillance with novel data streams: challenges and opportunities. *EPJ Data Science*. 2015;4(1):1–8.
28. Lazer D, Kennedy R, King G, Vespignani A. Big data. The parable of Google Flu: traps in big data analysis. *Science*. 2014;343(6176):1203–5.
29. Olson DR, Konty KJ, 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*. 2013;9(10):e1003256.
30. Centers for Disease Control and Prevention: Flu Activity Forecasting Website Launched | News (Flu) | CDC [http://www.cdc.gov/flu/news/flu-forecast-website-launched.htm] (2016). Accessed 3-16-2016.
31. National Oceanic and Atmospheric Administration: Dengue Forecasting [http://dengueforecasting.noaa.gov/] (2015). Accessed 3/16/2016.
32. Defense Advanced Research Projects Agency: CHIKV Challenge Announces Winners, Progress toward Forecasting the Spread of Infectious Diseases [http://www.darpa.mil/news-events/2015-05-27] (2015). Accessed 3/16/2016.
33. White House webpage: Congress Grants Broad Prize Authority to All Federal Agencies [http://www.whitehouse.gov/blog/2010/12/21/congress-grants-broad-prize-authority-all-federal-agencies] (2010). Accessed 8/11/2014.
34. Paul MJ, Dredze M, Broniatowski DA. Twitter Improves Influenza Forecasting. *PLoS Currents Outbreaks* 2014;10.1371/currents.outbreaks.1390b1379ed1370f1359bae1374ccaa1683a39865d39117.
35. Chakraborty P, Khadivi P, Lewis B, Mahendiran A, Chen J, Butler P, Nsoesie EO, Mekaru SR, Brownstein JS, Marathe MV, Ramakrishnan N. Forecasting a Moving Target: Ensemble Models for ILI Case Count Predictions. *Proceedings of the 2014 SIAM International Conference on Data Mining*. 2014, 262–270.
36. Columbia Prediction of Infectious Diseases: Columbia Mailman School of Public Health [http://cpid.iri.columbia.edu] (2016). Accessed 3/16/2016.
37. Hickmann KS, Fairchild G, Priedhorsky R, Generous N, Hyman JM, Deshpande A, et al. Forecasting the 2013–2014 influenza season using Wikipedia. *PLoS Comput Biol*. 2015;11(5):e1004239.
38. Brooks LC, Farrow DC, Hyun S, Tibshirani RJ, Rosenfeld R. Flexible Modeling of Epidemics with an Empirical Bayes Framework. *PLoS Comput Biol*. 2015; 11(8):e1004382.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

