**Forecast the 2018–2019 Influenza Season Collaborative Challenge**

**Objectives:**

To improve influenza forecasting, we will undertake a collaborative comparison of forecasts for the 2018-2019 influenza season. For each week during the season, participants will be asked to provide national and regional probabilistic forecasts for the overall influenza season (seasonal targets) and for the next four weeks ahead (short-term targets). The seasonal targets are the onset week, the peak week, and the peak intensity of the 2018-2019 influenza season. The short-term (week–ahead) targets are the percent of outpatient visits experiencing influenza-like illness (ILI) one week, two weeks, three weeks, and four weeks ahead from date of the forecast. All forecasts will be compared to the weighted values from the U.S. Outpatient Influenza-like Illness Surveillance Network (from the ILINet system: <http://www.cdc.gov/flu/weekly/overview.htm>). Participants can submit forecasts for seasonal targets, short-term targets, or both. If discussing the forecasting challenge on social media, teams are encouraged to use the hashtag #CDCflusight to promote visibility of the challenge.

**Eligibility:**

All are welcome to participate in this collaborative challenge, including individuals or teams that have not participated in previous CDC forecasting challenges.

**Dates:**

The Challenge Period will begin October 29, 2018 and will run until May 13, 2019. Participants must submit weekly forecasts by 11:59PM Eastern Standard Time each Monday. Missed or late submissions will not preclude participation in this challenge but will adversely affect submission scores.

**Forecasting Targets:**

* Seasonal Targets
  + The onset of the season is defined as the MMWR surveillance week (<http://wwwn.cdc.gov/nndss/script/downloads.aspx>) when the percentage of visits for influenza-like illness (ILI) reported through ILINet reaches or exceeds the baseline value for three consecutive weeks. Forecasted onset week values should be for the first week of that three week period.
  + The peak week is defined as the MMWR surveillance week that the weighted ILINet percentage is the highest for the 2018-2019 influenza season.
  + The peak intensity is defined as the highest numeric value that the weighted ILINet percentage reaches during the 2018-2019 influenza season.
* Short-term Targets
  + One- to four-week ahead forecasts will be defined as the weighted ILINet percentage for each target week.

The ILINet national and regional baseline values for the 2018-19 season are shown in the table below. These data will also be available at <http://www.cdc.gov/flu/weekly/overview.htm>.

|  |  |
| --- | --- |
| **Table 1. ILINet national and regional baselines, United States, 2018-19** | |
| **Group Name** | **2018-19 Baseline** |
| National | 2.2 |
| Region 1 | 1.8 |
| Region 2 | 3.1 |
| Region 3 | 2.0 |
| Region 4 | 2.2 |
| Region 5 | 1.8 |
| Region 6 | 4.0 |
| Region 7 | 1.6 |
| Region 8 | 2.2 |
| Region 9 | 2.3 |
| Region 10 | 1.1 |

ILINet values will be rounded to one decimal point for determining all forecast targets. In the case of multiple peak weeks (i.e. there is an identical peak ILINet value in two or more weeks within a geographic region), both weeks will be considered the peak week.

**Forecast Submission:**

Forecasts should provide probabilistic forecasts (i.e., 50% peak will occur on week 2; 30% chance on week 3) as well as the point prediction for each of the three seasonal targets and four week-ahead targets. The probabilities for each target prediction should be non-negative and sum to 1. If the sum is greater than 0.9 and less than 1.1, the probabilities will be normalized to 1.0. If any probability is negative or the sum is outside of the 0.9-1.1 range, the forecast will be discarded. Short-term forecast submissions should be relative to the most recent week of ILINet data released. For example, ILINet data for week 43 will be posted on Friday, November 2 at 12:00PM Eastern Standard Time. Each short-term forecast (1- , 2- , 3- , and 4-week ahead) submitted on Monday, November 5 should include predictions for ILINet values for weeks 44-47. Forecasts must be provided at both the national level and the HHS region level.

A description of methodology should be submitted to CDC by November 16 using the methodology form provided. This form captures key model factors, such as data source(s) and model type(s) in a standardized way. Model methodology and source data may be changed during the course of the challenge, but teams should submit a new methodology form as soon as possible after the change. Please submit the completed form and forward any questions to [flucontest@cdc.gov](mailto:flucontest@cdc.gov).

*Submission Structure*

All forecasts should be structured to match the attached spreadsheet (named “Weekly\_Submisson\_Spreadsheet”). **The structure of the spreadsheet (e.g., the column or row locations) should not be modified in any way.** For onset, the “none” field in the spreadsheet is to indicate if no influenza season is forecasted (e.g., the ILINet value never reaches or exceeds the baseline for at least three consecutive weeks during the season). Peak intensity and week-ahead forecasts should be given in the provided 0.1 percentage intervals labeled “bin\_start\_incl” on the submission sheet. For example, the bin for 3.1% represents the probability that the rounded ILI equals 3.1%. The probability assigned to the final bin labeled 13% includes the probability of ILINet values greater than or equal to 13%.

Forecasts should be submitted online through the FluSight website (<https://predict.cdc.gov/>). Instructions for submission will be provided once testing on the new website is complete. In the event forecasts cannot be submitted online, they may be emailed to [flucontest@cdc.gov](mailto:flucontest@cdc.gov) using the provided .csv spreadsheet. For an email submission, the file name should be modified to the following standard naming convention: a forecast submission using week 43 surveillance data submitted by John Doe University on November 5, 2018, should be named “EW43-JDU-2018-11-05.csv” where EW43 is the latest week of ILINet data used in the forecast, JDU is the name of the team making the submission (e.g., John Doe University), and 2018-11-05 is the date of submission.

At some point during the season, teams may be able to submit their forecasts using an application programming interface (API). Additional guidance will be provided at that time.

**Evaluation Criteria:**

*Log Score*

All forecasts will be evaluated using the weighted observations pulled from the ILINet system for MMWR week 28 of 2019, and the logarithmic scoring rule will be used to measure the accuracy of the probability distribution of a forecast. If is the set of probabilities for a given forecast, and   is the probability assigned to the observed outcome , the logarithmic score is:   
For onset and peak week, the probability assigned to the correct bin (based on the weighted ILINet value) plus the probability assigned to the immediately preceding and proceeding bins will be summed to determine the probability assigned to the observed outcome. For example, if onset occurs during week 47, the probabilities assigned to bins 46-48 will be summed. If onset is never reached during the season, only the probability assigned to the bin for “none” will be scored. In the case of multiple peak weeks, the probability assigned to the bins containing each peak week and the respective preceding and proceeding bins will be summed.

For peak intensity and short-term forecasts, the probability assigned to the correct bin plus the probability assigned to the five preceding and five proceeding bins will be summed to determine the probability assigned to the observed outcome. For example, if the correct peak ILINet value is 6.5%, the probabilities assigned to all bins ranging from 6.0% to 7.0% will be summed to determine the probability assigned to the observed outcome.

For all targets, if the correct bin is near the first or last bin, the number of bins summed will be adjusted accordingly. No bin farther than one bin (onset and peak week) or five bins (peak intensity and short-term, i.e. percentage forecasts) away from the correct bin will contribute to the score. For example, if the correct ILINet percentage for a given week is 0.3%, probabilities assigned to bins ranging from 0% to 0.8% will be summed. Undefined natural logs (which occur when the probability assigned to the observed outcome is 0) will be assigned a value of -10. Forecasts which are not submitted (e.g., if a week is missed) or that are incomplete (e.g., sum of probabilities greater than 1.1) will also be assigned a value of -10.

**Example:** A forecast predicts there is a probability of 0.2 (i.e., a 20% chance) that the flu season starts on week 44, a 0.3 (30%) probability that it starts on week 45, and a 0.1 (10%) probability that it starts on week 46, with the other 0.4 (40%) distributed across other weeks according to the forecast. Once the flu season has started, the prediction can be evaluated, and the ILINet data show that true onset was on week 45. The probabilities for week 44, 45, and 46 would be summed, and the forecast would receive a score of log(0.6) = -0.51. If the season started on another week, the score would be calculated on the probability assigned to that week plus the values assigned to the immediately preceding and proceeding weeks.

*Absolute Error*

Forecast accuracy will be measured by log score only. Nonetheless, forecasters are requested to continue to submit point predictions, which should aim to minimize the absolute error (AE). Absolute error (AE) is the absolute difference between a prediction  and an observation such that: . If a point prediction is not provided, CDC will estimate the point prediction using the median of the submitted distribution.

**Example:** A forecast predicts that the flu season will start on week 45; flu season actually begins on week 46. The AE of the prediction is |45-46| = 1 [week]. For season onset, if the point prediction is for no onset, please report a point prediction of “NA”.

**Method to Determine Overall Team Rankings**

Logarithmic scores for seasonal and short-term forecasts will be averaged across different submission time periods and locations to provide both specific and generalized measures of model accuracy. The overall team rankings at the end of the season will be determined by averaging scores across all of the national- and regional-level targets over their respective evaluation periods as described below. Teams that do not provide all seven seasonal and short-term targets for all locations will be ineligible to be named the overall top performing team; however, they will still be ranked for the targets they provided. Although teams may choose to participate in more than one challenge (e.g. FluSight described here and the hospitalization and state challenges described below), rankings for one challenge will not influence rankings for another, and an overall top-score will not be determined.

The evaluation period will vary by forecasting target and geographic region, representing the weeks when the forecasts are most useful. For all seasonal targets, the evaluation period will begin with the first forecast submission. The evaluation period for season onset will end six weeks after the observed onset week; the evaluation periods for peak week and intensity will end after ILINet is observed to go below baseline for the final time during an influenza season. For short-term forecasts, the evaluation period will begin four weeks prior to the observed onset week and will end three weeks after ILINet is observed to go below baseline for the final time during an influenza season.

**Data Sources**

Historical national surveillance data may be used for training and model development, and are available at <http://gis.cdc.gov/grasp/fluview/fluportaldashboard.html>. These data are updated every Friday at noon Eastern Standard Time. The “cdcfluview” package for R can be used to retrieve these data automatically. In addition, the archive of historical CDC regional baselines have also been utilized and are available at <https://github.com/cdcepi/FluSight-forecasts/blob/master/wILI_Baseline.csv>.

Teams are welcome to utilize additional data beyond ILINet - additional potential data sources include but are not limited to: Carnegie Mellon University’s Epidata API ([Delphi group](http://delphi.midas.cs.cmu.edu/) <<http://delphi.midas.cs.cmu.edu/>> and <https://github.com/undefx/delphi-epidata>) and Health Tweets (<http://www.healthtweets.org/>).

**Publication of Forecasts:**

All participants provide consent for their forecasts to be published in real-time on the CDC’s Epidemic Prediction Initiative website (<https://predict.cdc.gov/>), GitHub page (<https://github.com/cdcepi>), and, after the season ends, in a scientific journal describing the results of the challenge. The forecasts can be attributed to a team name (e.g., John Doe University) or anonymous (e.g., Team A) based on individual team preference. Team names should be limited to 25 characters for display online. The team name registered with the EPI website will be displayed alongside a team’s forecasts – any team that wishes to remain anonymous should contact CDC to obtain an anonymous team name to use.. No participating team may publish the results of another team’s model in any form without the team’s consent. The manuscript describing the accuracy of forecasts across teams will be coordinated by a representative from CDC.

**Ensemble Model and Null Models:**

Starting with the 2015-2016 influenza season, CDC created a simple average ensemble of forecasts to use as the basis of CDC’s communication of influenza forecasts. While this method is consistently one of the top performing forecasts among those submitted, the FluSight Network was created in 2017 to improve upon this ensemble by using performance from past years to weight the models’ contribution. This collaborative ensemble approach was implemented and submitted to CDC on a weekly basis during the 2017–18 season, and this model was one of the top-performing forecasting models overall, exceeding the accuracy of the simple average ensemble. This ensemble network is planning to continue in the 2018–19 season, and it is open to any teams who wish to contribute forecasts. Nick Reich at UMass is the lead. Please reach out to him at [nick@schoolph.umass.edu](mailto:nick@schoolph.umass.edu) if you are interested in joining this year. A draft of the guidance is available https://groups.google.com/d/forum/flusightnework.

In addition, forecasts will be displayed alongside the output of one null model for comparison, which is based solely on the historical distribution of the value of interest (i.e., onset week, peak week, peak percentage, or ILI percentage in a given MMWR week), excluding the 2009/2010 H1N1 pandemic season.

**Hospitalization Rates**

Teams interested in participating in the second year of the FluSurv-Net Hospitalization Rate Forecasting Challenge should contact CDC at [flucontest@cdc.gov](mailto:cdccontest@cdc.gov). Historical surveillance data of influenza hospitalization rates from FluSurv-Net are available at <http://gis.cdc.gov/GRASP/Fluview/FluHospRates.html>.

**State-based ILINet**

Teams interested in participating in the second year of the State-based ILINet Forecast Challenge should contact CDC at [flucontest@cdc.gov](mailto:cdccontest@cdc.gov).