**Forecast Influenza-related Hospitalizations During the 2018–2019 Influenza Season: Elective Collaborative Challenge Year 2**

**Objectives:**

Influenza-related hospitalizations are a major contributor to the overall burden of influenza in the United States. Accurate prediction of hospitalization rates could help ensure an appropriate public health response during an influenza season. To work towards this goal, CDC will coordinate a collaborative forecasting challenge for influenza-related hospitalizations during the 2018-2019 influenza season. For each week during the season, participants will be asked to provide national probabilistic forecasts for seasonal and short-term targets. The seasonal targets are the peak week and peak weekly rate of hospitalizations; the short-term targets are weekly hospitalization rates one, two, three, and four weeks ahead from the date of forecast. Forecasts will be compared with values from the Influenza Hospitalization Surveillance Network (FluSurv-NET: <https://www.cdc.gov/flu/weekly/#S6>).

**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 cannot be determined ahead of time because FluSurv-Net data are not published until enough cases have been reported. Teams will be alerted the first Friday that data are available and the first submission will be due the second Monday after the first publication (i.e., 10 days later). 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 peak week is defined as the MMWR surveillance week that the overall FluSurv-NET hospitalization rate is the highest for the 2018-2019 influenza season.
  + The peak weekly rate is defined as the highest numeric value that the overall FluSurv-NET hospitalization rate reaches during the 2018-2019 influenza season.
* Short-term Targets
  + One- to four-week ahead forecasts are defined as the weekly overall FluSurv-NET hospitalization rate for the target week.

FluSurv-NET rates 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 FluSurv-NET value in two or more weeks), both weeks will be considered the peak week.

**Forecast Submission**

Forecasts should provide probabilistic forecasts (i.e., 50% peak week will occur on week 2; 30% chance on week 3) for each of the two seasonal targets and four short-term 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 FluSurv-NET data released. For example, FluSurv-NET data for week 48 will be posted on Friday, December 7 at 11:00AM Eastern Time. Each short-term forecast (1- , 2- , 3- , and 4-week ahead) submitted on Monday, December 10 should include predictions for FluSurv-NET values for weeks 49-52.

A description of methodology should be submitted to CDC within two weeks of the start of the challenge 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 using the provided submission template (named “FluSurv\_Submission\_Template.csv”). **The structure of the spreadsheet (e.g. the column or row locations) should not be modified in any way to ensure accurate forecast upload.** Peak weekly rate and short-term forecasts should be given in the provided “0.1 case per 100,000” intervals labeled as “bin\_start\_incl” on the submission sheet (e.g. the bin for 3.1 represents probability that the rounded FluSurv-NET rate = 3.1). The probability assigned to the final bin (labeled 13 for all age groups except 65+, where it is labeled 60) includes the probability of all FluSurv-NET rates greater than or equal to that value.

Forecasts should be submitted online through the Epidemic Prediction Initiative (EPI) website (<https://predict.cdc.gov/>). Instructions for submission will be provided once testing on the new website is complete. To conform to the current structure of the Epidemic Prediction Initiative website’s submission system, some labels on the submission template are not aligned with the information presented. First, for this challenge, the submission template column labeled “location” should instead be used for the specific age groups. Second, while FluSurv-NET data are measured in rates per 100,000 population, the current units for short-term and peak weekly rate targets are labeled as “percent” – disregard and enter data as “X per 100,000.” Finally, the peak weekly rate target is labeled as “Season peak percentage.” These minor discrepancies are necessary to allow the submitted forecasts to be visualized correctly on the EPI website. Please note the relevant areas on the template and ensure submitted forecasts are properly labeled.

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 48 surveillance data submitted by John Doe University on December 10, 2018, should be named “EW48-JDU-Hosp-2018-12-10.csv” where EW48 is the latest week of FluSurv-NET data used in the forecast, JDU is the name of the team making the submission (e.g., John Doe University), and 2018-12-10 is the date of submission.

**Evaluation Criteria:**

*Log Score*

All forecasts will be evaluated using the hospitalization rates pulled from the FluSurv-NET system during MMWR week 28 of 2019. 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 peak week, the probability assigned to tet correct bin (based on the FluSurv-NET values) plus the probability assigned to the immediately preceding and proceeding weeks will be summed to determine the probability assigned to the observed outcome. For example, if peak week occurs during week 1, the probabilities assigned to bins 52, 1, and 2 will be summed. 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 weekly rate and short-term forecasts, the probability assigned to the correct bin plus the probability assigned to a variable number of preceding and proceeding bins (predetermined based on the specific age group being forecast) will be summed to determine the probability assigned to the observed outcome, with a minimum of one preceding and one proceeding bin. For both overall rates and rates for each age group, bins including up to ±10% of the observed value will be applied to each side of the observed value, rounded to the nearest 0.1. For example, if the observed overall peak weekly hospitalization rate is 3.3 per 100,000, the bins encompassing 10% of this value (0.3) above and below the observed value of 3.3 will be included. Therefore, the probabilities assigned to all bins ranging from 3.0 to 3.6 would be summed to determine the probability assigned to the observed outcome.

In the case of observed values < 1.0 per 100,000, a minimum of one bin preceding and proceeding the observed bin will be included with the observed bin. For example, if the observed weekly hospitalization rate is 0.2 per 100,000, the probabilities assigned to bins representing 0.1 and 0.3 will also be included. For all targets, if the correct bin is near the first or last bin, the number of bins will be truncated at the respective boundary.

For example, if the observed overall hospitalization rate for a given week is in the lowest bin (i.e., 0.0 per 100,000), probabilities assigned to bins ranging from 0.0 to 0.1 will be summed. If the observed rate is in the highest bin (i.e., ≥13.0 per 100,000), only those bins with values contained in the calculated ±10% window based on the value of the observed rate will be summed. Meaning, if the correct maximum overall hospitalization rate is 13.2 per 100,000, the window will be ±1.3 and only the probabilities assigned to bins ranging from 11.9 to ≥13.0 will be summed.

Undefined natural logs (which occur when the probability assigned to the observed outcome was 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 less than 0.9 or greater than 1.1) will also be assigned a value of -10.

**Example:** At the conclusion of the season, FluSurv-NET showed that the overall weekly hospitalization rate peaked at 5.4 per 100,000. The window of ±0.54 rounds to 0.5, so the probability bins that would be considered accurate and summed for scoring range from 4.9 to 5.9. If the submitted forecast predicted there is a 0.1 probability (i.e., a 10% chance) that hospitalization rates peak at 5.4 per 100,000, with an additional 0.3 probability (i.e., a 30% chance) assigned to multiple bin values between 4.9 and 5.3, and a 0.2 probability (i.e., a 20% chance) assigned to multiple bin values between 5.5 and 5.9, then the forecast would receive a score of log(0.6) = -0.51.

*Absolute Error*

Forecast accuracy will be measured by log score only. Nonetheless, forecasters are requested to continue to submit point predictions to aid in communication and dissemination of forecasts, 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 value of the submitted distribution.

**Example:** A forecast predicts that the hospitalization rates will peak in week 4; hospitalization rates actually peak on week 6. The AE of the prediction is [weeks].

**Method to Determine Overall Team Rankings**

Logarithmic scores for seasonal and short-term targets will be averaged across different submission periods 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 the two seasonal targets and the four short-term targets for the submitted age groups over the entire forecast period. Teams that do not provide all six targets for all age groups will be ineligible to be named the overall top performing team; however, they will still be ranked for the ages and targets provided. Although teams may choose to participate in more than one challenge (e.g. FluSight and the hospitalization challenge described here), rankings for one challenge will not influence rankings for another, and an overall top-score will not be determined.

**Data Sources**

Historical national surveillance data for training and model development are available at <https://gis.cdc.gov/GRASP/Fluview/FluHospRates.html>. These data are updated every Friday at noon Eastern Standard Time. Teams will be responsible for downloading weekly updates to FluSurv-NET surveillance data during the challenge. The “cdcfluview” R package can also be used to retrieve these data automatically. FluSurv-NET backfill data is available to participating teams from the 2013-2014 season onwards. Please contact [flucontest@cdc.gov](mailto:flucontest@cdc.gov) to receive the data. Teams are welcome to utilize additional data beyond FluSurv-NET - 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. If discussing the forecasting challenge on social media, teams are encouraged to use the hashtag #CDCflusight to promote visibility of the challenge.

**Ensemble Model and Historical Model:**

Participant forecasts will be combined into an unweighted average ensemble forecast to be published in real-time along with the participant forecasts. In addition, forecasts will be displayed alongside the output of a null model for comparison, which is based solely on the historical distribution of the value of interest (i.e., peak week, peak weekly rate, or hospitalization rate in a given MMWR week), excluding the 2009-2010 H1N1 pandemic season.

**FluSight Challenge**

Teams interested in participating in the sixth year of the FluSight Forecast the Influenza Season Collaborative Challenge should contact CDC at [flucontest@cdc.gov](mailto:flucontest@cdc.gov). Historical national surveillance data from ILINet are available at <http://gis.cdc.gov/grasp/fluview/fluportaldashboard.html> and archived national and regional ILI baseline values are available at <https://github.com/cdcepi/FluSight-forecasts/blob/master/wILI_Baseline.csv>.

**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).