**Forecast the 2019–2020 Influenza Season Collaborative Challenge**

**Objectives:**

CDC will again host the FluSight collaborative comparison of forecasts for the 2019-2020 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 four weeks ahead of publication (short-term targets). The seasonal targets are the onset week, the peak week, and the peak intensity of the 2019-2020 influenza season. The short-term targets are the percent of outpatient visits experiencing influenza-like illness (ILI) one week, two weeks, three weeks, and four weeks ahead of publication 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 28, 2019 and will run until May 11, 2020. 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 (rounded to the nearest 0.1) 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 2019-2020 influenza season.
  + The peak intensity is defined as the highest numeric value that the weighted ILINet percentage reaches during the 2019-2020 influenza season.
* Short-term Targets
  + One- to four-week ahead forecasts will be defined as the weighted ILINet percentage for each target week.

Updated 2019–2020 ILINet baseline values for the US and each HHS region will be available at <http://www.cdc.gov/flu/weekly/overview.htm> the week of October 7, 2019. Baseline ILI% values for past seasons can be found at <https://github.com/cdcepi/FluSight-forecasts/blob/master/wILI_Baseline.csv>

ILINet values will be rounded to the nearest 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), all weeks with the identical peak value 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. A forecast that is later discarded does not disqualify teams from participating but will be assigned a score of -10. 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 1 at 12:00PM Eastern Standard Time. Each short-term forecast (1- , 2- , 3- , and 4-week ahead) submitted on Monday, November 4 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 15, using a form which will be released to teams soon. 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 “FluSightILI\_submission\_template\_2019\_2020.csv”). The structure of the spreadsheet (e.g., the column or row locations) should not be modified in any way. The functions “*verify\_entry*” and “*verify\_entry\_file*” from the FluSight R package can be used to verify that columns are named and ordered correctly and that probabilities are non-negative and sum to a value between 0.9 and 1.1. 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 13.0%.

Forecasts should be submitted online through the FluSight website (<https://predict.cdc.gov/>). Instructions for submission will be listed in an appendix. The appendix for the 2018-19 challenge is included, and teams will receive an updated version soon. 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 4, 2019, should be named “EW43-JDU-2019-11-04.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 2019-11-04 is the date of submission.

**Evaluation Criteria:**

*Log Score*

Once initially published, ILINet values may change as additional reports are received or revised. The Epidata API includes weekly surveillance data as they were first published and in their most up-to-date version following backfilling (see “Data Sources” section below). All forecasts will be evaluated using the weighted observations pulled from the ILINet system for MMWR week 28 of 2020, 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 each forecast of each target,  will be set to the probability assigned to the single bin containing the observed outcome (based on the rounded weighted ILINet value). If onset is never reached during the season, only the probability assigned to the bin for “none” will be scored. For the peak week target, in the case of multiple peak weeks, the probability assigned to the bins containing each peak week 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.

In addition to the final scores, CDC may provide interim score reports to participants on a semi-regular basis during the season. Interim scores will not impact final team standings.

**Example:** A forecast predicts there is a probability of 0.3 (i.e., a 30% chance) that the flu season starts on week 45, with the remaining 0.7 probability 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 probability assigned to week 45, 0.3, would be derived, and the forecast would receive a score of log(0.3) = -1.20. If the season started on another week, the score would be calculated on the probability assigned to that week.

*Absolute Error*

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. While official team rankings will only be based on log scores, CDC may report on the accuracy of point predictions in manuscripts and analyses.

**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 at least one week during the challenge 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 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.

Teams are free to submit as many systems as they wish, but these systems should all be substantially different from one another, reflecting materially different approaches to the forecasting problem.

**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/>). The Epidata API includes weekly surveillance data as they were first published and in their most up-to-date version following backfilling. **If teams know of additional data that they would like to highlight, please email** [**flucontest@cdc.gov**](mailto:flucontest@cdc.gov) **so this information can be included in an updated version of this document.**

**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/>), CDC’s Epidemic Prediction Initiative 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 each models’ contribution. This collaborative ensemble approach was implemented and submitted to CDC on a weekly basis during the 2017–18 and 2018-19 seasons, and these models were among the top-performing forecasting models overall in both seasons, exceeding the accuracy of the simple average ensemble. This ensemble network will continue in the 2019–20 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 at <https://github.com/FluSightNetwork/cdc-flusight-ensemble/blob/master/guidelines.md>.

In addition, forecasts will be displayed alongside the output of one null model for comparison. In this model, a smooth gaussian kernel density function was fit to historical observations 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. This null historical model is described in greater detail [here](https://www.nature.com/articles/s41598-018-36361-9#Sec8).

**State-based ILINet**

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

**Department of Defense Influenza Foresting Challenge**

The Department of Defense (DoD) Influenza Forecasting Challenge Guidance, developed by the Armed Forces Health Surveillance Branch (AFHSB), aims to accurately forecast the onset, the week when influenza-like-illnesses peak, and peak intensity of the influenza season, in order to inform public health professionals and policy makers for more effective and targeted interventions. Teams interested in participating in the first year of the DoD Influenza Forecasting Challenge should contact [dha.ncr.health-surv.mbx.dodflucontest@mail.mil](mailto:dha.ncr.health-surv.mbx.dodflucontest@mail.mil).

**Influenza Hospitalization Challenge**

The Influenza Hospitalization Forecasting Challenge will be put on hold during the

2019-20 season while we explore the feasibility and value of different forecasting projects in this area.