2022-2023 DoD AFHSD-IB CLI Forecasting Collaboration Guidelines

07 September 2022

# Introduction

The Armed Forces Health Surveillance Division, Integrated Biosurveillance Branch (AFHSD-IB) is hosting a COVID-like Illness (CLI) forecasting collaboration, in order to provide forecasts for an indicator of COVID-19 disease activity in the DoD population in select US Military Health Service (MHS) markets. The collaboration will utilize ICD coded medical encounters for CLI provided by AFHSD-IB. Each Wednesday, starting October 12th, 2022, participants will provide US market-level forecasts for specific targets related to the burden and trajectory of CLI.

# Eligibility

All are welcome to participate in this challenge. Participants must complete a Defense Health Agency (DHA) Memorandum of Agreement (MOA) to receive weekly datasets. All datasets will be de-identified and aggregated to the MHS market-level and contain data from January 2020 through the current week. Participants do not need to provide forecasts for all locations or targets to participate. Historical and weekly datasets will be available on Github. An example of the weekly dataset along with a data dictionary have been provided along with this guideline.

# Overview

COVID-19 disease, caused by the novel 2019 coronavirus, SARS-CoV-2, was declared a pandemic by the WHO on March 11, 2020, with known global spread since January 2020. To provide insight on the burden and trajectory of COVID-19 outbreaks in DoD-relevant locations in the US and globally, AFHSD will undertake a collaborative comparison of CLI medical encounter forecasts. For each week, participants will provide forecasts for the 15 largest MHS markets using DoD service member and beneficiary data. The short-term targets are the percent of outpatient, MTF encounters experiencing CLI one week, two weeks, three weeks, and four weeks ahead from the week of most recently provided data. Forecast results will be compared to the observed weekly CLI percent based on data from the Electronic Surveillance System for the Early Notification of Community-based Epidemics (ESSENCE).

# Forecasting Period

The forecast period will run from MMWR week 41 (week beginning Sunday, October 9th, 2022) through MMWR week 20 (week ending Saturday, May 27th, 2022). The first forecasts will be based on weekly data through MMWR week 40, while final forecasts will be based on weekly data through MMWR week 20. The end of the forecasting period may be altered based on the characteristics of the outbreak. Missed or late submissions will not preclude participation, however we ask that participants aim to be as timely as possible in their submissions to increase the utility of the results. AFHSD-IB reserves the right to internally compare the submitted forecasts to observed values from ESSENCE, but forecast evaluation results will not preclude participation in the collaboration. Timely submissions will be incorporated into an ENSEMBLE model.

Forecasts should be submitted on Wednesdays by 11:59pm ET. The first submission will be due by Wednesday, October 12th, 2022, but the first forecasting deadline may be extended so that teams can adjust to the novelty of the datasets and targets. Teams may start submitting forecasts at any time during the forecasting period, but early participation is highly encouraged. ESSENCE data will be shared with participants on Mondays by 5:00pm ET, pending any system delays.

Scores will be generated each week during the forecasting period to assess the utility of each model by comparing the observed weekly CLI values to the forecasted values for each MHS market. The final observed weekly CLI values for internal forecast evaluation will be determined by the reported ESSENCE data as of 6 weeks after the last forecasting target (after EW30) and used to determine final scores for the forecasting challenge.

# Forecasting Targets

* 1 through 4-week ahead CLI percentages: Number of CLI encounters over total medical outpatient encounters for the projected week, multiplied by 100.
  + Target details:
    - Type of target: continuous
    - Description: The reported CLI% for {1, 2, 3, 4} week(s) after the most recently released ESSENCE data.
    - Units: percent, a real number in [0, 100]
    - Quantiles: {0.01, 0.025, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, 0.975, 0.99}.

# Forecasting Locations

Forecasts may be provided for any of the 15 largest MHS market locations (list provided at the end of this document):

Locations are mapped to the data by MTF DMISID and grouped by market name. A list of DMISIDs in each market is provided for participants. A list of counties within 30 miles of each MHS market can also be provided.

# ESSENCE Datasets and CLI% Definition

All participants will sign a memorandum of agreement (MOA) to receive the aggregated data elements weekly during the forecasting period. Once the MOA is in place, AFHSD-IB will notify participants by email to download the DoD data from GitHub**.**

AFHSD defines CLI% as the total number of weekly CLI outpatient encounters divided by the weekly number of total outpatient encounters and multiplied by 100, for a given location. An observed values, or ‘truth’ file with calculated CLI % through the most recent week will be provided to participants each week. The truth file will be used by AFHSD to evaluate accuracy of forecast submissions for each target.

# Forecasts and Data Formatting

## Teams and Models

Teams interested in participating in the DoD AFHSD-IB CLI Forecasting Collaboration can submit forecasts from multiple models. Teams are encouraged to provide forecasts for locations and targets for which they feel their models are well-suited.

## Forecast file storage and submission

Submitted forecasts will be stored in the CDC GitHub repository: <https://github.com/cdcepi> (the repository will be created later). Each team will have a subdirectory within the data-forecasts directory where weekly forecasts will be stored. The subdirectory name will have the format “team-model”, where team is the name of your team and model is the name of your model. The team and model names should be less than 15 characters, not include any hyphens, and be unique from any other model in the project.

Prior to the first submission for a given model, the submitting team must provide a metadata file with structured information about the model. Each submitting team must choose a full name and an abbreviation for both their team and their model to uniquely identify their submissions.

The metadata file for each model must be named `metadata-[teamabbr]-[modelabbr].txt` and include the following variables:

* team\_name: Name of your team (<50 characters)
* model\_name: Name of your model (<50 characters)
* model\_abbr: Abbreviated name for your model in format of [team\_abbr]-[model\_abbr] where each of the [team\_abbr] and [model\_abbr] are text strings of less than 15 alphanumeric characters and do not include hyphens or white space
* model\_contributors: List of all individuals involved in the forecasting efforts, and email address. At least one contributor needs to have a valid email address. The syntax of this field should be:

name1 (affiliation 1) <user@address>, name2 (affiliation 2) <user2@address2>

* website\_url: A url to a web site with additional info or data about your model or team
* license
* team\_model\_designation: Upon initial submission this field should be one of “primary”, “proposed” or “other”. For teams submitting only one model, this should be “primary”. For each team, only one model can be designated as “primary”. Primary means the model will be scored in evaluations, eligible for inclusion in the ensemble, and visualized. Proposed means the team would like the model to be considered as a "secondary" model rather than an "other" model. Secondary means the forecasts will be visualized and eligible for inclusion in the ensemble, but will not be scored in evaluations. Other means the forecasts will not be visualized, included in the ensemble, nor scored in evaluations.
* methods: Brief description of your forecasting methodology that is less than 200 characters
* ensemble\_of\_hub\_models: a boolean value (true or false) that indicates whether a model combines multiple hub models into an ensemble.

We request that all metadata and forecast submissions for each team will be submitted via a GitHub pull request. We will provide instructions for submitting via pull request if this process is new for a team.

## Forecast file format

In what follows, we refer to a “forecast” as a collection of quantitative predictions that are specific to a location and target specified above. One forecast can be submitted for a given model on or before the forecast “due date” specified in the provided dates table. A forecast consists of a single plain-text file, in a particular format, which encapsulates the set of predictions for all or a subset of locations and targets.

Forecast submission template details are provided below. In general, a prediction for a specific location and target will be specified by point forecasts and a quantile representation of a probability distribution. We will refer to these two representations as “point forecasts” and “quantile forecasts”.

Forecasts should provide quantile forecasts (i.e., 0.01, 0.025, 0.05, etc...) as well as point predictions for each target. Forecast files must be a comma-separated value (csv) file with the appropriate columns. An example of the weekly forecast file along with a data dictionary have been provided along with this guideline. The value for each single quantile or point prediction should be non-negative. Each forecast file within a subdirectory should have the following format: YYYY-MM-DD-team-model.csv, where the date YYYY-MM-DD is the forecast\_date, team is the name of the team, and model is the name of your model. The team and model in this file must match the team and model in the directory this file is in.

# Data Sources

Once the MOA is signed and executed, historical ESSENCE surveillance data may be used for training and model development. Teams are welcome and encouraged to utilize additional data beyond the provided ESSENCE data.

# Contact Info

Additional questions may be addressed to [dha.ncr.health-surv.mbx.dodflucontest@mail.mil](mailto:dha.ncr.health-surv.mbx.dodflucontest@mail.mil). Please refer to the CDC Flusight Forecast Github page (<https://github.com/cdcepi/Flusight-forecast-data>) for additional info on file formats. **Note**: A separate repository for the DoD AFHSD-IB CLI Forecasting Collaboration will be created at a later date.

# Appendix

**Table A1.** ICD-10 codes comprising ESSENCE CLI definition

|  |  |
| --- | --- |
| ICD-10 CODE for COVID-like Illness | DESCRIPTION |
| *B34.2* | Coronavirus infection, unspecified |
| *B97.21* | SARS-associated coronavirus as the cause of diseases classified elsewhere |
| *B97.29* | Other coronavirus as the cause of diseases classified elsewhere |
| *J00* | Acute nasopharyngitis [common cold] |
| *J06.9* | Acute upper respiratory infection, unspecified |
| *J12.81* | Pneumonia due to SARS-associated coronavirus |
| *J12.89* | Other viral pneumonia |
| *J12.9* | Viral pneumonia, unspecified |
| *J16.8* | Pneumonia due to other specified infectious organisms |
| *J17* | Pneumonia in diseases classified elsewhere |
| *J18.0* | Bronchopneumonia, unspecified organism |
| *J18.1* | Lobar pneumonia, unspecified organism |
| *J18.8* | Other pneumonia, unspecified organism |
| *J18.9* | Pneumonia, unspecified organism |
| *J20.8* | Acute bronchitis due to other specified organisms |
| *J20.9* | Acute bronchitis, unspecified |
| *J40* | Bronchitis, not specified as acute or chronic |
| *J22* | Unspecified acute lower respiratory infection |
| *J80* | Acute respiratory distress syndrome |
| *R05* | Cough |
| *R50.9* | Fever, unspecified |
| *R06.0* | Dyspnea |
| *R06.00* | Dyspnea unspecified |
| *R06.02* | Shortness of breath |
| *R06.03* | Acute respiratory distress |
| *R06.09* | Other forms of dyspnea |
| *U07.1* | COVID-19 |
| *R43.0* | Anosmia |
| *R43.2* | Parageusia |
| *J84.111* | Idiopathic interstitial pneumonia not otherwise specified |

**Table A2.** 15 Largest MHS Markets

|  |  |
| --- | --- |
| mARKET nAME (fORMAL) | mARKET nAME (fOR fILES) |
| National Capital Region | National\_Capital\_Region |
| Tidewater | Tidewater |
| San Antonio | San\_Antonio |
| Hawaii | Hawaii |
| San Diego | San\_Diego |
| Bragg/Pope | Bragg\_Pope |
| Hood | Hood |
| Colorado Springs | Colorado\_Springs |
| Bliss/White Sands | Bliss\_White\_Sands |
| Lejeune/Cherry Point | Lejeune\_Cherry\_Point |
| Lewis/McCord | Lewis\_McChord |
| Pensacola/Eglin | Pensacola\_Eglin |
| Benning | Benning |
| Jacksonville | Jacksonville |
| Campbell | Campbell |

**Table A3.** Graph of COVID-19 cases and CLI encounters (using definition in Table A1) in MHS beneficiaries over time

