Meta and consensus forecast of COVID-19 targets

Thomas McAndrew,[1, *] Juan Cambeiro,[2] David Braun,[1] Tamay Besiroglu,[2] Damon Luk,[1] and Allison Codi[1]

[1] *College of Health, Lehigh University, Bethlehem, Pennsylvania, United States of America*

[2] *Metaculus*

(Dated: January 27, 2021)

SUMMARY FINDINGS

As of Jan 25, 2021 there are a reported 26,566,968 cumulative confirmed COVID-19 cases, 430,552 cumulative deaths due to COVID-19 according to the COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University, and on Jan 24, 2021 there were 11,795 US previous day adult and pediatric admissions to the hospital because of COVID-19 as reported by the Department of Health and Human Services.

Our aim is to provide two types of probabilistic forecasts of the evolving US COVID-19 outbreak to support public health officials, infectious disease modeling groups, and the general public. The first forecast is an aggregate of probabilistic predictions from experts in the modeling of infectious disease and trained forecasters called a *consensus forecast* (Fig. 1). The second forecast is a combination of our consensus forecast and the the COVIDhub-ensemble, an ensemble of computational models hosted by the COVID-19 Forecast Hub. This combination of predictions from computational models and from experts and trained forecasters is called a *metaforecast*.

A consensus of experts and trained forecasters predicts, for the week beginning Jan 24th and ending Jan 30th, an increase in the number of pediatric and adult hospital admissions (median = 132,500; 80% CI = [105,000, 168,750]), increase in the number of new confirmed cases of COVID-19 (median = 1,700,000; 80% CI = [1,309,000, 2,074,000]), and an increase in the number of new deaths due to COVID-19 (median = 22,400; 80% CI: [17,600, 27,200]).

In an effort to support predictions from computational models maintained by the COVID-19 Forecast Hub, we asked forecasters to assess the COVIDhub-ensemble's 4 week ahead prediction of US incident deaths made on Jan 4th, 2021. The consensus expects the COVIDhub's median prediction on Jan 4th, 2020 of 17,936 will underestimate the true number of incident deaths for the week beginning Jan 24th (median = 1.24; 80% CI = [0.99, 1.51]). A COVIDhub adjusted prediction of incident deaths is 17,936 × 1.24 = 22,241 and inline with the expert and trained forecaster consensus prediction of the number of new deaths (22,400).

If greater than or equal to 50% of the US population initiates vaccination by March 1, 2021 the consensus median prediction of the cumulative number of deaths by Dec. 31, 2021 is 520,000 (80% CI = [385,000, 690,000]). In contrast, if less than 50% of the US population initiates vaccination the consensus median prediction is 645,000 (80% CI = [385,000, 690,000]). A consensus of subject matter experts and trained forecasters predict 125,000 (difference between two medians above) fewer deaths due to COVID-19 if at least 50% of the population was vaccinated by March 1. 2021 and highlights the importance of increasing the rate of vaccinations throughout the US.

Information about the spread of the B.1.1.7 variant—the variant that was identified in the United Kingdom and found in more than 98% of UK SARS-CoV-2 samples sent for genomic sequencing—in the US has important implications for decisions about non-pharmaceutical interventions and changes in the pace of vaccination. A consensus of subject matter experts and trained forecasters predict 87% of US samples sent for genomic sequencing in the first two weeks of Feb. that have an S-gene dropout (present in all B.1.1.7 samples) will be identified as the B.1.1.7 variant.

Metaforecasts were generated for the number of incident confirmed cases and incident deaths for the week of Jan 24th, 2021 (Fig. 2). The Metaforecast's uncertainty intervals for incident cases and deaths were in between the uncertainty intervals of the COVIDhub ensemble of computational models and consensus forecast of experts and trained forecasters. Consensus forecasts tended to have wider uncertainty intervals compared to the COVIDhub ensemble, and so uncertainty intervals for the metaforecast were larger than the COVIDhub ensemble but smaller than the consensus forecast.

---

* mcandrew@lehigh.edu

CONSENSUS FORECASTS

Consensus predictions for posed questions

- The median consensus prediction of the number of new US COVID-19 hospital admissions beginning Jan 24, 2021 and ending Jan 30, 2021 is 132,500 (80% CI:[105,000, 168,750])

- The median consensus prediction of the number of new US deaths due to COVID-19 beginning Jan 24, 2021 and ending Jan 30, 2021 is 22,400 (80% CI:[17,600, 27,200])

- The median consensus prediction of the number of new US confirmed COVID-19 cases beginning Jan 24, 2021 and ending Jan 30, 2021 is 1,700,000 (80% CI:[1,309,000, 2,074,000])

- The median consensus prediction of the factor to multiply the COVID-19 Forecast Hub's 4 week ahead median forecast of incident deaths made on Jan 4th, 2021 (a forecast for the week beginning Jan 24, 2021 and ending Jan 30, 2021) by so that the prediction equals the true number of reported deaths is 1.24 (80% CI:[0.99, 1.51]). The median factor of 1.24 would shift the number of deaths forecasted by the COVID-19 Forecast Hub from 17,936 to 22,241.

- If less than 50% of the US population receives one or more vaccine doses by March 1st, 2021 then the median consensus prediction of the number of cumulative deaths due to COVID-19 on December 31st, 2021 is 645,000 (80% CI:[430,000, 870,000])

- If greater than or equal to 50% of the US population receives one or more vaccine doses by March 1st, 2021 then the median consensus prediction of the number of cumulative deaths due to COVID-19 on December 31st, 2021 is 520,000 (80% CI:[385,000, 690,000])

- The median consensus prediction of the percent of S-gene dropout samples that are submitted for sequencing contain the B.1.1.7 variant is 87% (80% CI:[62%, 97%]).
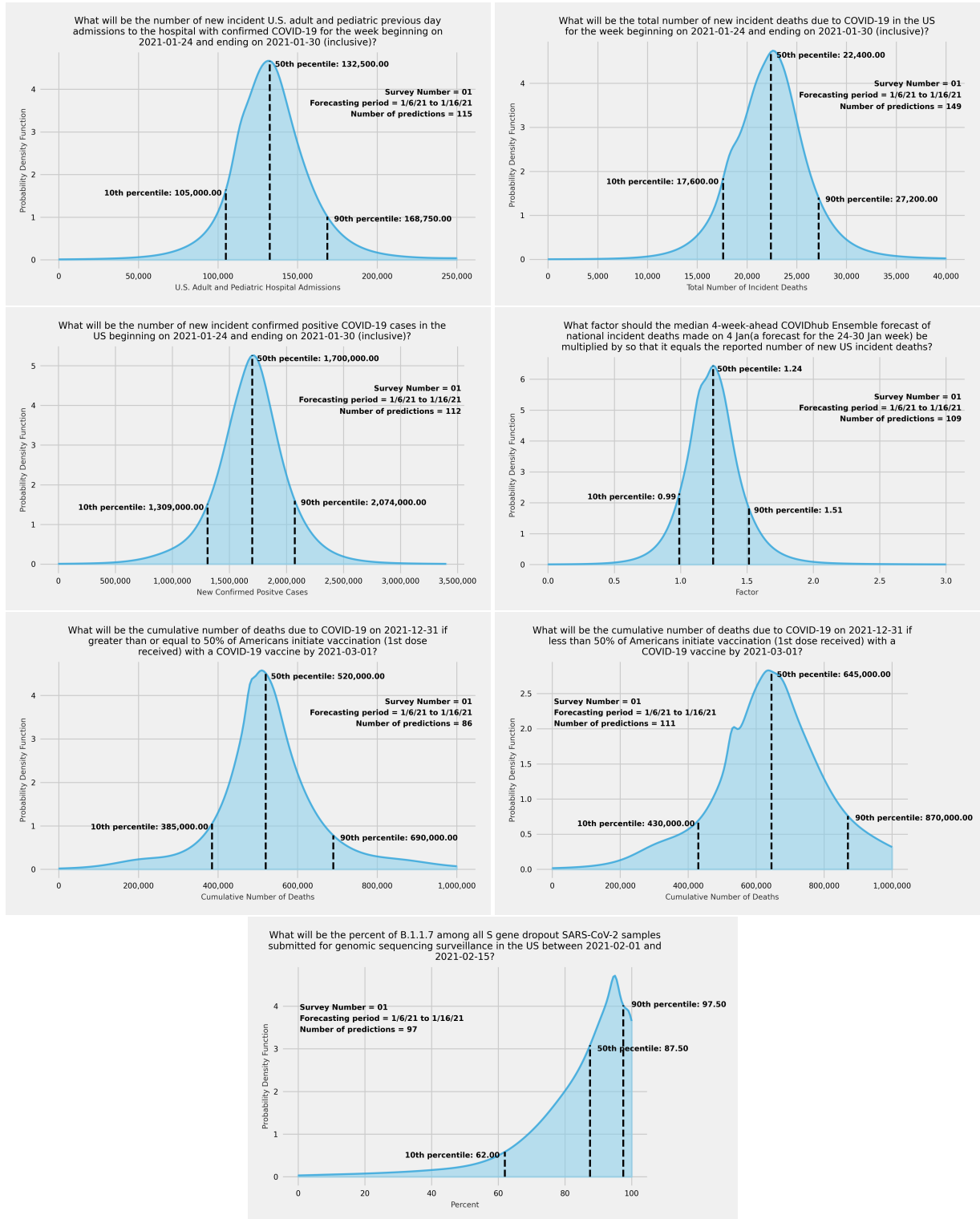
FIG. 1: Consensus predictive densities of seven targets of the US COVID-19 outbreak. A consensus density is an equally weighted average of individual densities proposed by trained forecasters and subject matter experts. Black dotted lines represent (from left to right) the tenth, fiftieth, and ninetieth percentiles of the distribution. Individual plots are available for download at https://github.com/computationalUncertaintyLab/aggStatModelsAndHumanJudgment_PUBL/tree/main/densities.
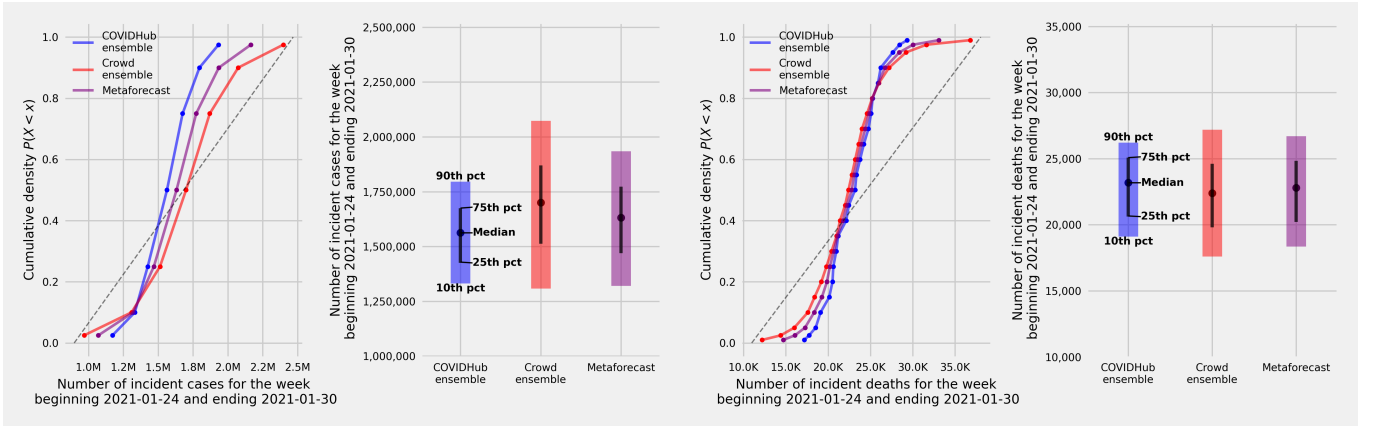
FIG. 2: Metaforecasts of the number of incident cases and incident deaths for the week beginning on Jan. 24th, 2021 and ending on Jan 30th, 2021. A metaforecast is a combination of an ensemble of computational models from the COVID-19 Forecasthub and a consensus of predictions from trained forecasters and experts. (A) Cumulative density functions of an ensemble of computational models, consensus of predictions from trained forecasters and experts, and a metaforecast. (B) An 80% prediction interval (the 10th and 90th percentile), 50% prediction interval (25th and 75th percentile), and the median (50th) percentile for all three ensemble models. This plot is available for download at https://github.com/computationalUncertaintyLab/aggStatModelsAndHumanJudgment_PUBL/blob/main/metaforecasts/cases_survey1.png .

MECHANICS OF A CONSENSUS FORECAST

To build a forecast we (i) pose questions to a crowd of subject matter experts and trained generalist forecasters, (ii) collect predictive densities over potential future values from each member of the crowd, and (iii) we aggregate this set of predictive densities into a single *consensus forecast*. We expect predictions from subject matter experts and trained forecasters to be accurate and calibrated because they have access to structured data, the same data computational models use, and because they have access to subjective, unstructured data often unavailable to computational models.

Suppose we collect a set $F = \{f_1, f_2, f_3, \cdots, f_C\}$ of predictive densities from C members of the crowd. Our consensus predictive density ($f_{\text{consensus}}$) is generated as

$$f_{\text{consensus}}(v) = \sum_{i=1}^{C} \pi_i f_i(v)$$

$$\text{such that } \sum_{i=1}^{C} \pi_i = 1$$

where f_i is the ith individual density from a crowd member, C is the number of individual densities, and π_i is the corresponding weight assigned to that member of the crowd. As of now, this summary report assigns equal weights $\left(\pi_i = \dfrac{1}{C} \right)$ to each member who proposed a predictive density f_i

MECHANICS OF A METAFORECAST

A metaforecast is built by combining an ensemble of computational models and a consensus of predictions from humans, in our case, predictions from trained forecasters and experts. The COVID-19 forecast hub ensemble cumulative

predictive density is divided into 23 quantiles for incident deaths and 7 quantiles for incident cases. We can extract the same number of quantiles from our consensus forecast and compute quantiles for our metaforecast as

$$Q_{\text{metaforecast}} = \frac{1}{2}Q_{\text{COVID-19 ensemble}} + \frac{1}{2}Q_{\text{Consensus forecast}}$$

where Q_x is a quantile from distribution x.

ONGOING METAFORECASTING WORK

There is ongoing work to generate a full density for the meta forecast by (i) estimating a probability density for the COVID-19 ensemble from quantile information and (ii) combining predictive densities generated by the COVID-19 ensemble and consensus forecast. We plan to generate a cumulative density function (cdf) $F_{\text{COVID-19 ensemble}}$ for the COVID-19 ensemble using (i) monotic cubic interpolation [1] and (ii) fitting a mixture model that minimizes the sum square error between the fitted distribution and COVID-19 ensemble quantiles. We will combine the COVID-19 ensemble cumulative density and a linearly interpolated cumulative density from the consensus forecast as

$$F_{\text{metaforecast}}(v) = \pi_{\text{COVID-19 ensemble}} F_{\text{COVID-19 ensemble}} + \pi_{\text{Consensus forecast}} F_{\text{Consensus forecast}}$$

where π is the weight assigned to each ensemble. We can use the linearity of the derivative and our ability to differentiate the above interpolated cdfs to generate a probability density function for the metaforecast as

$$f_{\text{metaforecast}}(v) = \frac{dF_{\text{metaforecast}}}{dv} = \pi_{\text{COVID-19 ensemble}} \frac{dF_{\text{COVID-19 ensemble}}}{dv} + \pi_{\text{Consensus forecast}} \frac{dF_{\text{Consensus forecast}}}{dv}$$

QUESTIONS AND RESOLUTION CRITERIA

Each month, we store all question and resolution text, the launch and close data to make forecasts, and urls corresponding to question in a dataset available for download at https://github.com/computationalUncertaintyLab/aggStatModelsAndHumanJudgment_PUBL/blob/main/questions/questionData.csv.

- Question 1

  - Question: What will be the number of new incident U.S. adult and pediatric previous day admissions to the hospital with confirmed COVID-19 for the week beginning on 2021-01-24 and ending on 2021-01-30 (inclusive)?

  - Resolution Criteria:This question will resolve as the total number of adult plus pediatric previous day admissions with confirmed COVID-19 as recorded in the Department of Health and Human Service's report of COVID-19 reported patient impact and hospital capacity. The total previous day admissions is computed using two variables in this report: previous_day_admission_adult_covid_confirmed and previous_day_admission_pediatric_covid_confirmed. This report, and the resolution criteria, includes data on all 50 US states, Washington DC, Puerto Rico, and the US Virgin Islands (53 states and territories). The report will be accessed one week after the end of the month (2021-02-06).

  - Range:[0-240K]

  - Question URL: https://pandemic.metaculus.com/questions/6160/new-us-hospital-admissions-24-30-january/

- Question 2

  - Question: What will be the total number of new incident deaths due to COVID-19 in the US for the week beginning on 2021-01-24 and ending on 2021-01-30 (inclusive)?

- – Resolution Criteria: This question will resolve as the number of new deaths due to confirmed COVID-19 for the week beginning on 2021-01-24 and ending on 2021-01-30 (inclusive) as recorded in the Johns Hopkins University (JHU) CSSE Github data repository. This file records the daily number of deaths by county. From this file deaths are summed across all counties and aggregated by week to generate the number of new deaths per week. The report will be accessed one week after the end of the month (2021-02-06).
  - – Range:[0-40k]
  - – Question URL: https://pandemic.metaculus.com/questions/6161/new-us-covid-deaths-24-30-january/

- Question 3

  - – Question: What will be the number of new incident confirmed positive COVID-19 cases in the US beginning on 2021-01-24 and ending on 2021-01-30 (inclusive)?
  - – Resolution Criteria: This question will resolve as the number of new confirmed cases beginning on 2021-01-24 and ending on 2021-01-30 (inclusive) recorded in the Johns Hopkins University (JHU) CSSE Github data repository. This file records the daily number of cases by county. From this file cases are summed across all counties and aggregated by week to generate the number of new cases per week. The report will be accessed one week after the end of the month (2021-02-06).
  - – Range:[0-3M]
  - – Question URL: https://pandemic.metaculus.com/questions/6162/new-us-covid-cases-24-30-january/

- Question 4

  - – Question: What factor should the median 4-week-ahead COVIDhub Ensemble forecast of national incident deaths made on 4 Jan(a forecast for the 24-30 Jan week) be multiplied by so that it equals the reported number of new US incident deaths?
  - – Resolution Criteria: This question will resolve as the factor that the The COVID-19 Forecast Hub's "COVIDhub" Ensemble median forecast of the US national number of incident deaths 4 weeks into the future should be multiplied by (the reported number of incident deaths divided by the forecasted median) to equal the reported number of US national incident deaths as reported by the Johns Hopkins University (JHU) CSSE Github data repository.
  - – Range:[0-3]
  - – Question URL https://pandemic.metaculus.com/questions/6163/factor-covidhub-forecast-to-be-multiplied-by/

- Question 5

  - – Question: What will be the cumulative number of deaths due to COVID-19 on 2021-12-31 if less than 50% of Americans initiate vaccination (1st dose received) with a COVID-19 vaccine by 2021-03-01?
  - – Resolution Criteria: The percent of the population that received a COVID-19 vaccine on or before 2021-03-01 will be computed by dividing the number of individuals who have initiated vaccine (1st dose taken) provided by the CDC COVID data tracker by the current US population which on 2021-01-04 was reported to be 330,782,991 and multiplying this fraction by 100. The CDC COVID data tracker that counts the number of individuals who have initialized vaccination will be accessed when data is available after and as close as possible to 2021-03-01.

    To resolve deaths, we will use the cumulative number of deaths due to confirmed COVID-19 as recorded in the Johns Hopkins University (JHU) CSSE Github data repository. This file records the daily number of deaths by county. From this file deaths are summed across all counties and aggregated to week to generate the number of new deaths per week. The report will be accessed one week after 2021-12-31.

    9 January edit: This question will resolve ambiguously if greater than or equal to 50% of Americans are vaccinated by 2021-03-01.
  - – Range:[0-1M]
  - – Question URL: https://pandemic.metaculus.com/questions/6165/number-of-us-covid-deaths-if-50-vaccinated/

- Question 6

– Question: What will be the cumulative number of deaths due to COVID-19 on 2021-12-31 if greater than or equal to 50% of Americans initiate vaccination (1st dose received) with a COVID-19 vaccine by 2021-03-01?

– Resolution Criteria: The percent of the population that received a COVID-19 vaccine on or before 2021-03-01 will be computed by dividing the number of individuals who have initiated vaccine (1st dose taken) provided by the CDC COVID data tracker by the current US population which on 2021-01-04 was reported to be 330,782,991 and multiplying this fraction by 100. The CDC COVID data tracker that counts the number of individuals who have initialized vaccination will be accessed when data is available after and as close as possible to 2021-03-01.

To resolve deaths, we will use the cumulative number of deaths due to confirmed COVID-19 as recorded in the Johns Hopkins University (JHU) CSSE Github data repository. This file records the daily number of deaths by county. From this file deaths are summed across all counties and aggregated to week to generate the number of new deaths per week. The report will be accessed one week after 2021-12-31.

9 January edit: This question will resolve ambiguously if less than 50% of Americans are vaccinated by 2021-03-01.

– Range:[0-1M]

– Question URL: https://pandemic.metaculus.com/questions/6164/number-of-covid-deaths-if-50-vaccinated/

• Question 7

– Question: What will be the percent of B.1.1.7 among all S gene dropout SARS-CoV-2 samples submitted for genomic sequencing surveillance in the US between 2021-02-01 and 2021-02-15?

– Resolution Criteria: This NextStrain page on S:N501Y as a proportion of overall S gene dropout samples in the US will be consulted for resolution. To access this data, scroll to the bottom of the page and download "selected metadata." In this Excel sheet, please consult the "Pangolin Lineage" column to find B.1.1.7 samples. Samples that have a "collection data" date between 2021-02-01 and 2021-02-15 (inclusive) will be considered. We will compute the proportion by dividing the number of B.1.1.7 samples by the total number of samples (which are all 69del samples) in the spreadsheet. NextStrain will be accessed in the last week of February (the week ending on 2021-02-27) for resolution.

– Range:[0-100]

– Question URL: https://pandemic.metaculus.com/questions/6166/-b117-among-all-s-gene-dropout-samples/

---

[1] Frederick N Fritsch and Ralph E Carlson. Monotone piecewise cubic interpolation. *SIAM Journal on Numerical Analysis*, 17(2):238–246, 1980.