Meta and consensus forecast of COVID-19 targets

Allison Codi,[1] Damon Luk,[1] Juan Cambeiro,[2] David Braun,[1] Tamay Besiroglu,[2]
Eva Chen,[3] Luis Enrique Urtubey de Cèsaris,[3] and Thomas McAndrew[1, *]

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

[2] *Metaculus*

[3] *Good Judgment Inc.*

(Dated: August 11, 2021)

SUMMARY FINDINGS

As of April, 15, 2021 there were a reported 31,552,732 cumulative confirmed COVID-19 cases and 569,539 cumulative deaths in the United States according to the COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. On April, 15, 2021 there were 5,796 US previous day adult and pediatric admissions to the hospital attributed to 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, trained forecasters from Metaculus, and generalist forecasters from Good Judgment Open—a forecasting platform open to any interested member of the public—called a *consensus forecast* (Fig. 1). The second forecast is a combination of our consensus forecast and 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 *metaconsensus* (Fig. 4).

A consensus of forecasters from Metaculus and forecasters from Good Judgment Open (no experts volunteered predictions) for the week beginning Apr 25 and ending May 1 predicted 38,463 new pediatric and adult hospital admissions attributed to COVID-19 (80% CI = [28,657, 46,380]), 457,432, incident confirmed cases of COVID-19 (80% CI = [345,793, 561,624]), and 4,473 incident deaths due to COVID-19 (80% CI: [3,179, 5,487]). Generalist forecasters from Good Judgment Open were aligned with forecasts from Metaculus (see Fig. 1).

Predicting the number of people in the US who are vaccinated can support hospital supply chain management and help better understand the trajectory of incident cases and deaths. A consensus of trained forecasters from Metaculus and Good Judgment Open predict 157,035,640 people in the US will have one or more doses of a COVID-19 vaccine by Apr 30 (80% CI = [149,930,824, 166,819,623]). Of those 157,035,640 people, the consensus predicts 107,666,015 will be fully vaccinated by Apr 30 (80% CI = [99,596,816, 114,945,273]).

The consensus median prediction of cumulative deaths attributed to COVID-19 in the US by Dec 31, 2021 is 693,940 (80% CI = [608,501, 873,493]), approximately 80,000 more deaths than have occurred at the time of this publication.

Information about the spread of the B.1.1.7 variant of concern—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 the potential to impact decisions about non-pharmaceutical interventions and changes in the pace of vaccination. A consensus of forecasters from Metaculus predict the rolling 7-day average of % B.1.1.7 in the US on Apr 30 will be 69% (80% CI = [49, 89]).

Of even higher threat than variants of concern are variants of high consequence (VOHC)—SARS-COV-2 variants with clear evidence that prevention measures or medical countermeasures (MCMs) have significantly reduced effectiveness relative to previously circulating variants. Currently, there are no VOHCs. Forecasters from Metaculus predict that the CDC will not announce that they are tracking any VOHCs until the Fall of 2022 (median=89, 80% CI = [89, 89]).

Metaforecasts were generated for the number of incident confirmed cases and incident deaths for the week of Apr 25 (Fig. 4). A metaforecast can combine forecasts from computational models and a consensus of human judgment. It has a level of uncertainty between that of computational models and a consensus of human judgment.

---

[*] 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 Apr 25, 2021 and ending May 1, 2021 was 38,463 (80% CI:[28,657, 46,380]).

- The median consensus prediction of the number of new US deaths due to COVID-19 beginning Apr 25, 2021 and ending May 1, 2021 was 4,473 (80% CI:[3,179, 5,487]).

- The median consensus prediction of the number of new US confirmed COVID-19 cases beginning Apr 25, 2021 and ending May 1, 2021 was 457,432 (80% CI:[345,793, 561,624]).

- The median consensus prediction of the cumulative number of people who will have received one or more doses of a COVID-19 vaccine in the U.S by Apr 30, 2021 was 157,035,640 (80% CI:[149,930,824, 166,819,623]).

- The median consensus prediction of the cumulative number of people who will be fully vaccinated in the U.S by Apr 30, 2021 was 107,666,015 (80% CI:[99,596,816, 114,945,273]).

- The median consensus prediction of the cumulative number of deaths in the US due to COVID-19 by Dec 31, 2021 wa 693,940 (80% CI:[608,501, 873,493]).

- The median consensus prediction of the seven day rolling average of the % B.1.1.7 in the US on Apr 30, 2021 was 69 (80% CI:[49, 89]).

- A consensus of experts and trained forecasters predicted the CDC will announce they are monitoring a variant of high consequence by 89 (80% CI:[89, 89])
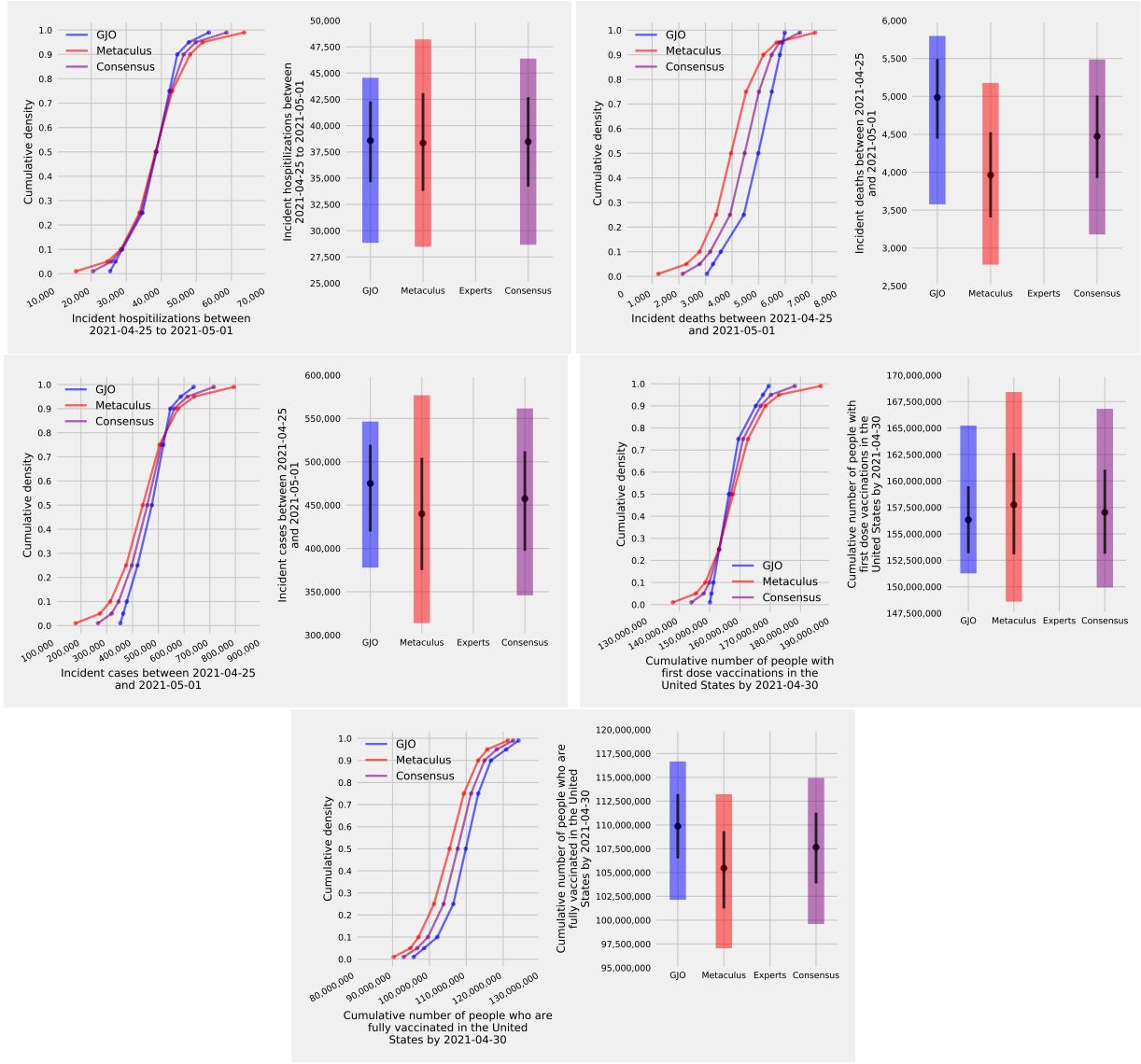
CONSENSUS PREDICTIVE DENSITIES

FIG. 1: (Left) Cumulative density functions of a consensus of trained forecasters from Metaculus and generalist forecasters from Good Judgment Open, and an equally weighted combination of all three consensus predictions (a consensus). (Right) An 80% prediction interval (the 10th and 90th percentile), 50% prediction interval (25th and 75th percentile), and the median (50th) percentile for all three consensus models.

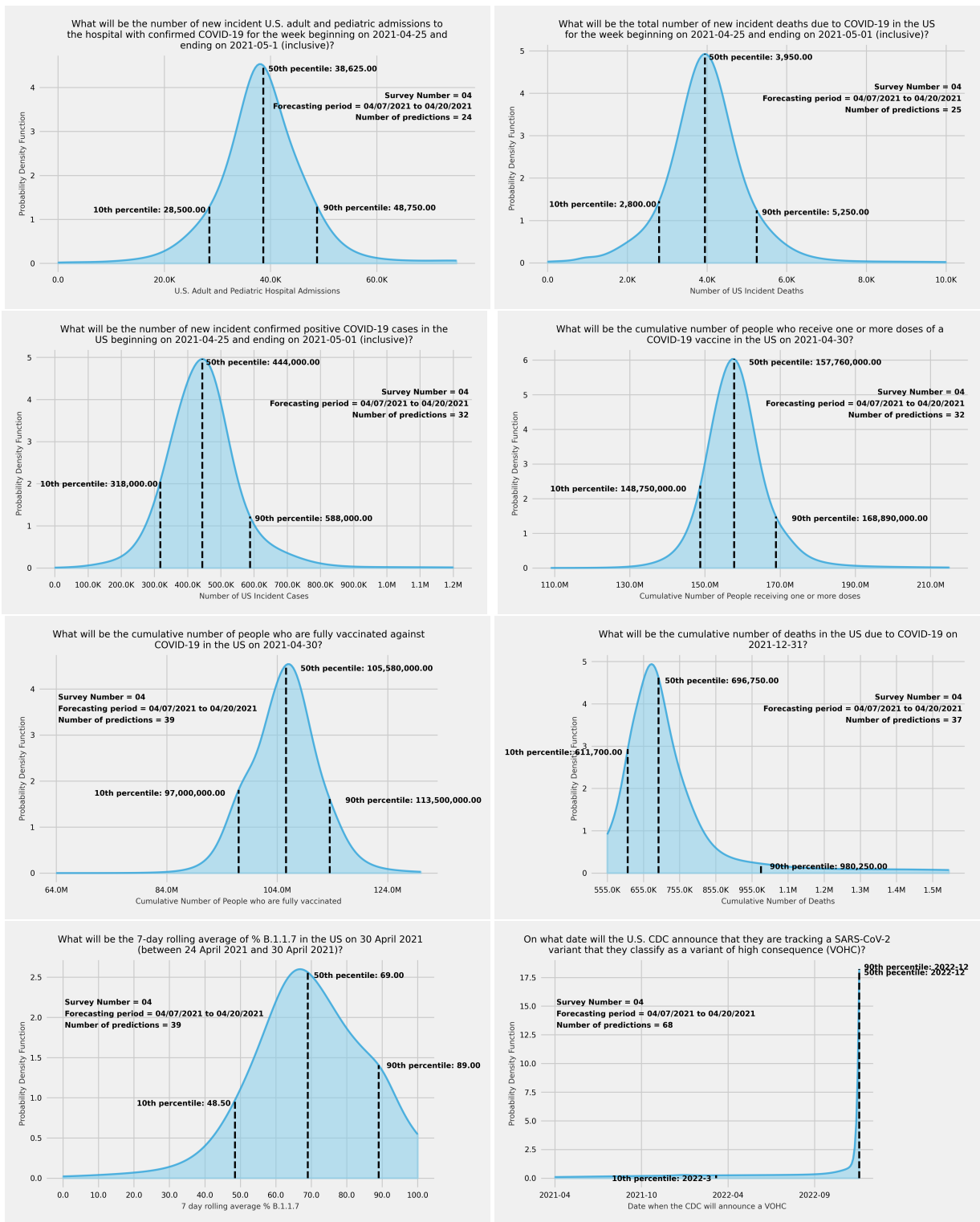# CONSENSUS PREDICTIVE DENSITIES (METACULUS PLATFORM)

FIG. 2: Consensus predictive densities of eight targets of the US COVID-19 outbreak submitted by forecasters from Metaculus. A consensus density is an equally weighted average of individual densities proposed by forecasters. 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.

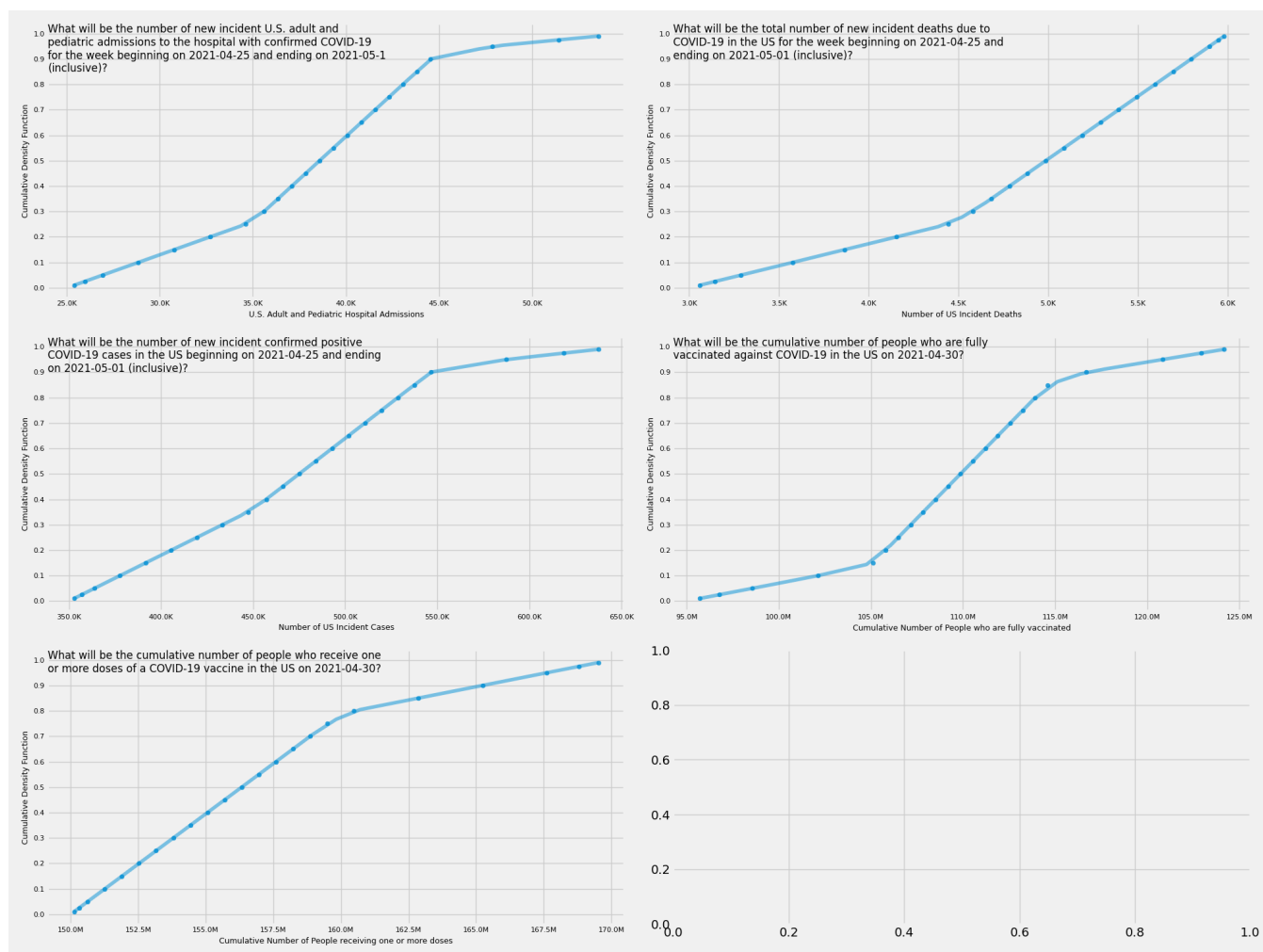# CONSENSUS PREDICTIVE DENSITIES (GOOD JUDGMENT PLATFORM)

FIG. 3: Cumulative density functions of five targets of the US COVID-19 outbreak. A cumulative density function is the probability that the variable takes a value less than or equal to x. Plots are available for download at https://github.com/computationalUncertaintyLab/aggStatModelsAndHumanJudgment_PUBL/tree/main/data/GJOforecasts.
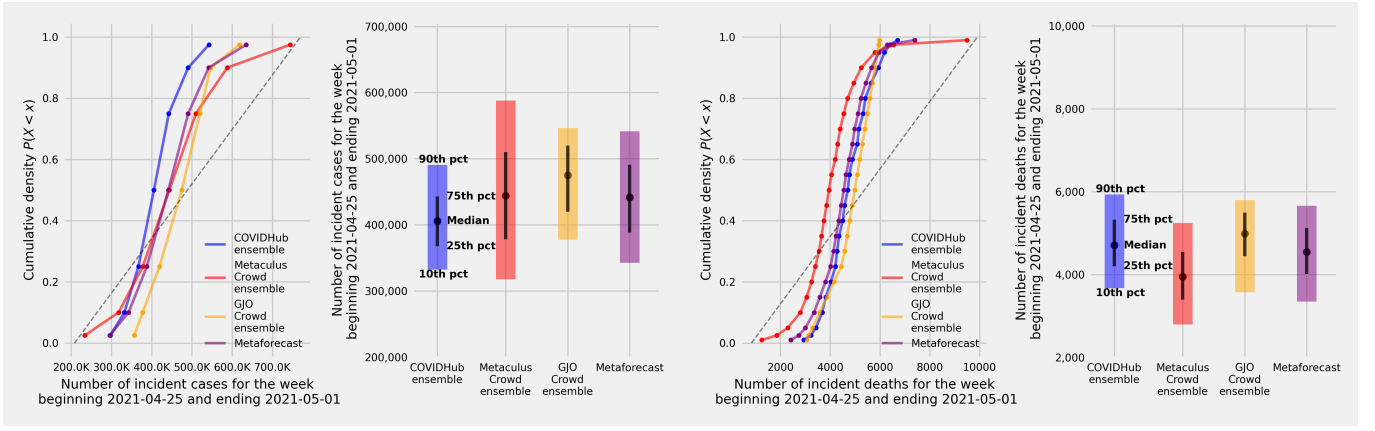
FIG. 4: Metaforecasts of the number of incident cases and incident deaths for the week beginning on Apr 25, 2021 and ending May 1, 2021. A metaforecast is a combination of an ensemble of computational models from the COVID-19 Forecasthub and a consensus of predictions from trained forecasters and experts. (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_survey4.png

MECHANICS OF A CONSENSUS FORECAST

To build a forecast we (i) pose questions to a crowd of subject matter experts and 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 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 US adult and pediatric admissions to the hospital with confirmed COVID-19 for the week beginning on 2021-04-25 and ending on 2021-05-01 (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 Services report of COVID-19 reported patient impact and hospital capacity for the dates from 2021-04-25 to 2021-05-01, corresponding to the number of hospitalizations from 2021-04-25 to 2021-05-01. Daily updates are provided by the Department of Health and Human Services. 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 and stored in Lehigh University's Computational Uncertainty Lab Github data repository. 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 no sooner than (2021-09-05).

  - Range:[0-75K]

  - Metaculus URL: https://pandemic.metaculus.com/questions/6985/new-us-covid-hospital-admissions-25-apr-1-may/

  - GJO URL: https://www.gjopen.com/questions/1976

- Question 2

- – What will be the total number of new incident deaths due to COVID-19 in the US for the week beginning on 2021-04-25 and ending on 2021-05-01 (inclusive)?

- – Resolution Criteria: This question will resolve as the number of new deaths due to confirmed COVID-19 for the week beginning on 2021-04-25 and ending on 2021-05-01 (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 number of deaths for the week beginning on 2021-04-25 will be computed by adding the number of new deaths from the 2021-04-25 up to, and including, 2021-05-01. The report will be accessed no sooner than 9 May 2021.

- – Range:[0-10k]

- – Metaculus URL: https://pandemic.metaculus.com/questions/6986/new-us-covid-deaths-25-apr-1-may/

- – GJO URL: https://www.gjopen.com/questions/1977

- Question 3

  - – Question: What will be the number of new incident confirmed positive COVID-19 cases in the US beginning on 2021-04-25 and ending on 2021-05-01 (inclusive)?

  - – Resolution Criteria: This question will resolve as the number of new confirmed cases beginning on 2021-04-25 and ending on 2021-05-01 (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 no sooner than 2021-05-09.

  - – Range:[0-1.2M]

  - – Metaculus URL: https://pandemic.metaculus.com/questions/6987/new-us-covid-cases-25-apr-1-may/

  - – GJO URL: https://www.gjopen.com/questions/1978

- Question 4

  - – Question: What will be the cumulative number of people who receive one or more doses of a COVID-19 vaccine in the US on 2021-04-30?

  - – Resolution Criteria: This question will resolve as the cumulative number of people who have received one or more doses of a vaccine on 2021-04-30 as recorded by the Centers for Disease Control COVID-19 Data tracker under Vaccine Trends. The dashboard is updated daily at 8pm ET and will be accessed no sooner than 9 May 2021.

  - – Range:[109M-215M]

  - – Metaculus URL https://pandemic.metaculus.com/questions/6988/cumulative-1st-dose-us-vaccinations-30-april/

  - – GJO URL: https://www.gjopen.com/questions/1979

- Question 5

  - – Question: What will be the cumulative number of people who are fully vaccinated against COVID-19 in the US on 2021-04-30?

  - – Resolution Criteria: This question will resolve as the cumulative number of people who receive one or more doses of a COVID-19 vaccine on 2021-04-30 as recorded by the Centers for Disease Control COVID-19 Data tracker in the column "People Fully Vaccinated". The dashboard is updated daily at 8pm ET and will be accessed no sooner than 9 May 2021.

  - – Range:[64M-130M]

  - – Metaculus URL: https://pandemic.metaculus.com/questions/6989/cumulative-fully-vaccinated-in-us-on-30-april/

  - – GJO URL: https://www.gjopen.com/questions/1980

- Question 6

- Question: What will be the cumulative number of deaths in the US due to COVID-19 on 2021-12-31?
- Resolution Criteria: This question will resolve as the number of cumulative deaths due to confirmed COVID-19 on 2021-12-31 as recorded in the Johns Hopkins University (JHU) CSSE Github data repository. This file records the daily number of deaths by county. The number of cumulative deaths at the end of the year will be computed by adding the cumulative number of deaths across states. This data, 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 no sooner than 9 January 2022.
- Range:[555K-1.5M]
- Metaculus URL: `https://pandemic.metaculus.com/questions/6990/cumulative-us-covid-deaths-by-end-of-2021/`

- Question 7

  - Question: What will be the 7-day rolling average of % B.1.1.7 in the US on 30 Apr 2021 (between 24 Apr 2021 and 30 Apr 2021)?
  - Resolution Criteria: This question will resolve as the 7-day rolling average of % B.1.1.7 in the US (in other words, the frequency of B.1.1.7 as a percentage of all sequenced SARS-CoV-2 cases) on 30 Apr 2021 (the average percentage between 24 Apr 2021 and 30 Apr 2021) at the following website: `https://outbreak.info/situation-reports?pango=B.1.1.7&loc=USA&selected=USA`. This percentage will be accessed no sooner than 9 May 2021.
  - Range:[0-100]
  - Metaculus URL: `https://pandemic.metaculus.com/questions/6991/-b117-in-us-on-30-april/`

- Question 8

  - Question:On what date will the United States CDC announce that they are tracking a SARS-CoV-2 variant that they classify as a variant of high consequence (VOHC)?
  - Resolution Criteria: This question will resolve as the date that a SARS-CoV-2 variant is categorized under the "Variant of High Consequence" section on the CDC's SARS-CoV-2 Variant Classifications and Definitions page.

    If no variant is classified as a VOHC before 1 January 2023, then this resolves as ¿ 31 December 2022.
  - Range:[2021-04-07 - >2022-12-31]
  - Metaculus URL: `https://pandemic.metaculus.com/questions/6992/variant-of-high-consequence-before-2023`