

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 C  sar  s,³ and Thomas McAndrew^{1,*}

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

²Metaculus

³Good Judgment Inc.

(Dated: August 11, 2021)

SUMMARY FINDINGS

As of May 31, 2021 there were a reported 34,828,023 cumulative confirmed COVID-19 cases and 607,852 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 July 24, 2021 there were 5,585 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. 2). 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 *metaconsensus* (Fig. 4).

A consensus of forecasters from [Metaculus](#) and forecasters from [Good Judgment Open](#) (no experts volunteered predictions) for the week beginning May 30 and ending Jun 5 predicted 21,296 new pediatric and adult hospital admissions attributed to COVID-19 (80% CI = [12,539, 28,370]), 171,353 incident confirmed cases of COVID-19 (80% CI = [93,697, 266,508]), and 2,664 incident deaths (80% CI: [93,697, 266,508]) due to COVID-19. Compared to the end of April, predictions for incident cases, hospitalizations, and deaths decreased. 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 predicted 165,461,443 people in the US will have one or more doses of a COVID-19 vaccine by May 31 (80% CI = [161,492,442, 169,742,916]). Of those 165,461,443 people, 140,123,256 would be fully vaccinated by May 31 (80% CI = [132,508,875, 147,920,239]).

Despite the vaccination effort, the consensus median prediction of cumulative deaths attributed to COVID-19 in the US by Dec 31, 2021 was 707,345 (80% CI = [628,353, 806,574]).

As of Apr 30, there are five variants that cause "reduced neutralization by convalescent and post-vaccination sera" - these are: P.1, B.1.351, B.1.427, B.1.429, and B.1.526. Information about high risk variants in the US has the potential to impact decisions about non-pharmaceutical interventions and changes in the pace of vaccination. A consensus of subject matter experts and trained forecasters predicted that over the two week period of 23 May to 05 Jun, the prevalence of SARS-Cov-2 variants that are thought to partially escape immunity would be 15.6% (80% CI = [1.6, 41.4]).

Of even higher threat than the aforementioned variants of interest and consequence 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 predicted that the CDC will not announce any VOHCs until after 2024 (median=806,574, 80% CI = [806,574, 806,574]).

Metaforecasts were generated for the number of incident confirmed cases and incident deaths for the week of May 30 (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 May 30, 2021 and ending Jun 5, 2021 was 21,296 (80% CI:[12,539, 28,370]).
- The median consensus prediction of the number of new US deaths due to COVID-19 beginning May 30, 2021 and ending Jun 5, 2021 was 2,664 (80% CI:[1,912, 3,743]).
- The median consensus prediction of the number of new US confirmed COVID-19 cases beginning May 30, 2021 and ending Jun 5, 2021 was 171,353 (80% CI:[93,697, 266,508]).
- The median consensus prediction of the cumulative number of people who will have received one of more doses of a COVID-19 vaccine in the U.S by May 31, 2021 was 165,461,443 (80% CI:[161,492,442, 169,742,916]).
- The median consensus prediction of the cumulative number of people who will be fully vaccinated in the U.S by May 31, 2021 was 140,123,256 (80% CI:[132,508,875, 147,920,239]).
- The median consensus prediction of the cumulative number of deaths in the US due to COVID-19 by Dec 31, 2021 was 707,345 (80% CI:[628,353, 806,574]).
- The median consensus prediction of the prevalence of SARS-CoV-2 variants thought to partially escape immunity for the two-week period of May 23 to Jun 05 2021 was 15.6 (80% CI:[1.6, 41.4]).
- A consensus of experts and trained forecasters predicted the CDC will announce they will not announce they are monitoring a variant of high consequence until after 806,574 (80% CI:[806,574, 806,574]).

CONSENSUS PREDICTIVE DENSITIES

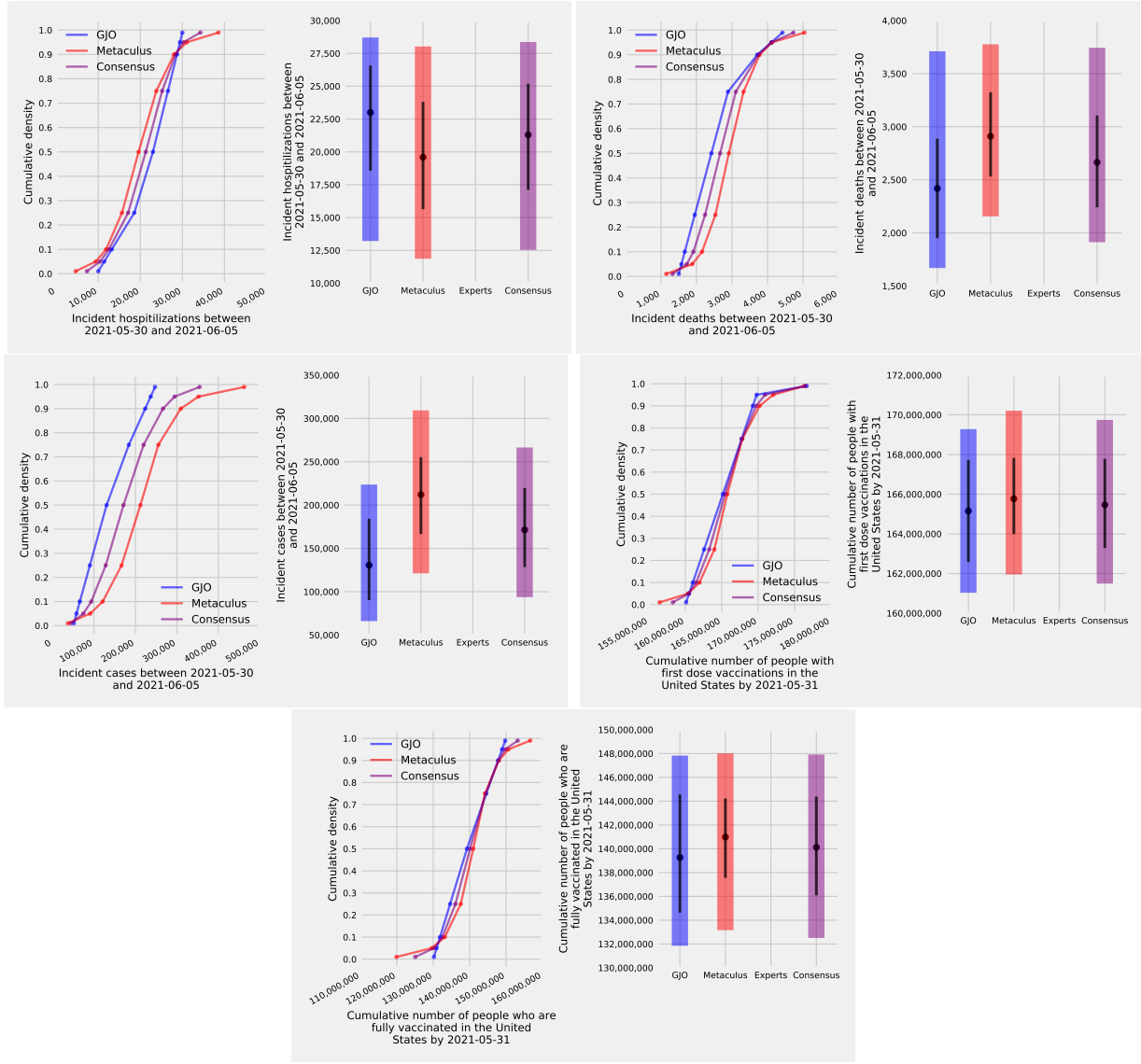


FIG. 1: (Left) Cumulative density functions of a consensus of trained forecasters from Metaculus and generalist forecasters from Good Judgment Open, and an equally weighted combination of all three consensus predictions (a consensus). (Right) An 80% prediction interval (the 10th and 90th percentile),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

CONSENSUS PREDICTIVE DENSITIES (METACULUS PLATFORM)

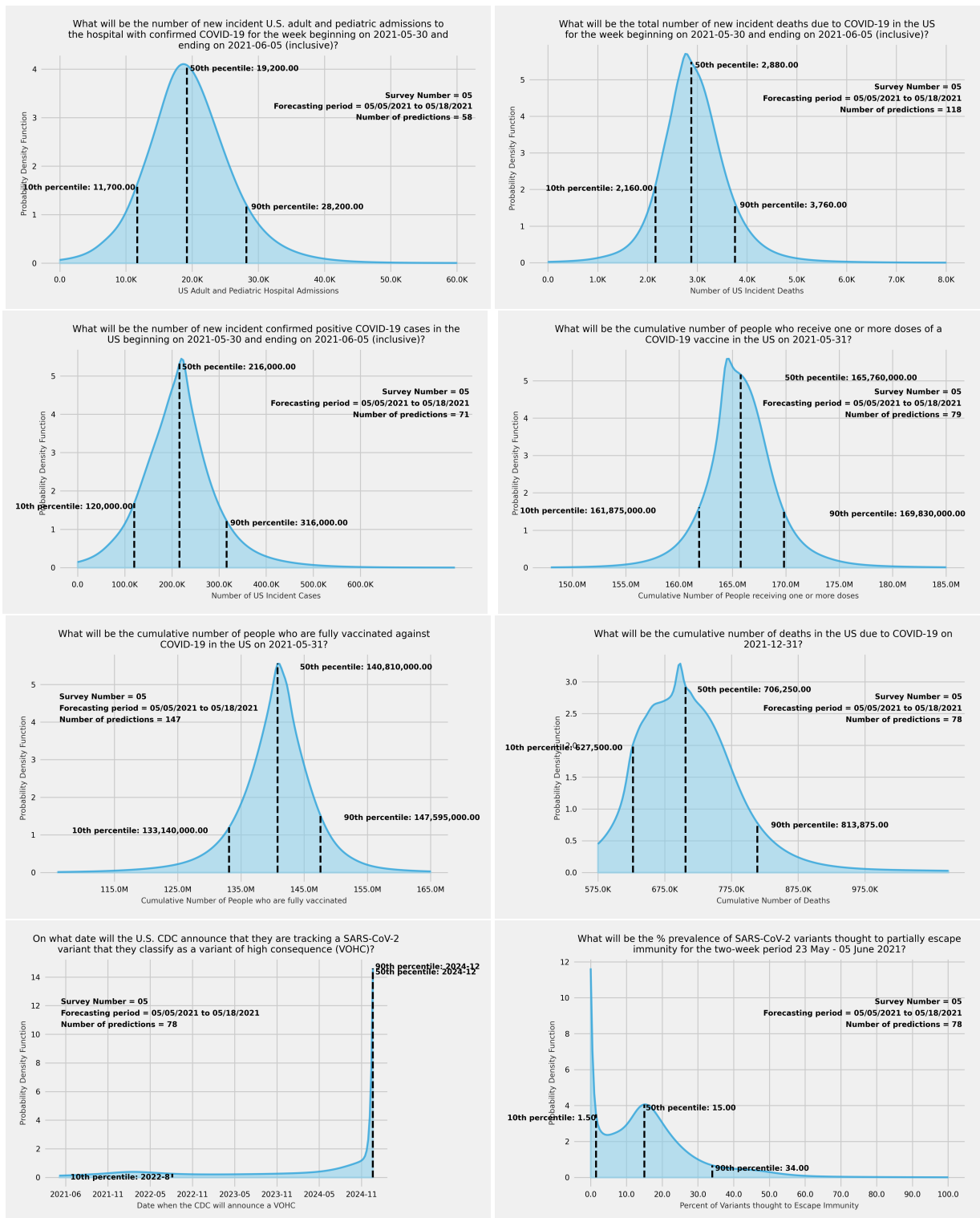


FIG. 2: Consensus predictive densities of eight targets of the US COVID-19 outbreak submitted by forecasters from Metaculus. A consensus density is an equally weighted average of individual densities proposed by trained forecasters. Black dotted lines represent (from left to right) the tenth, fiftieth, 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)

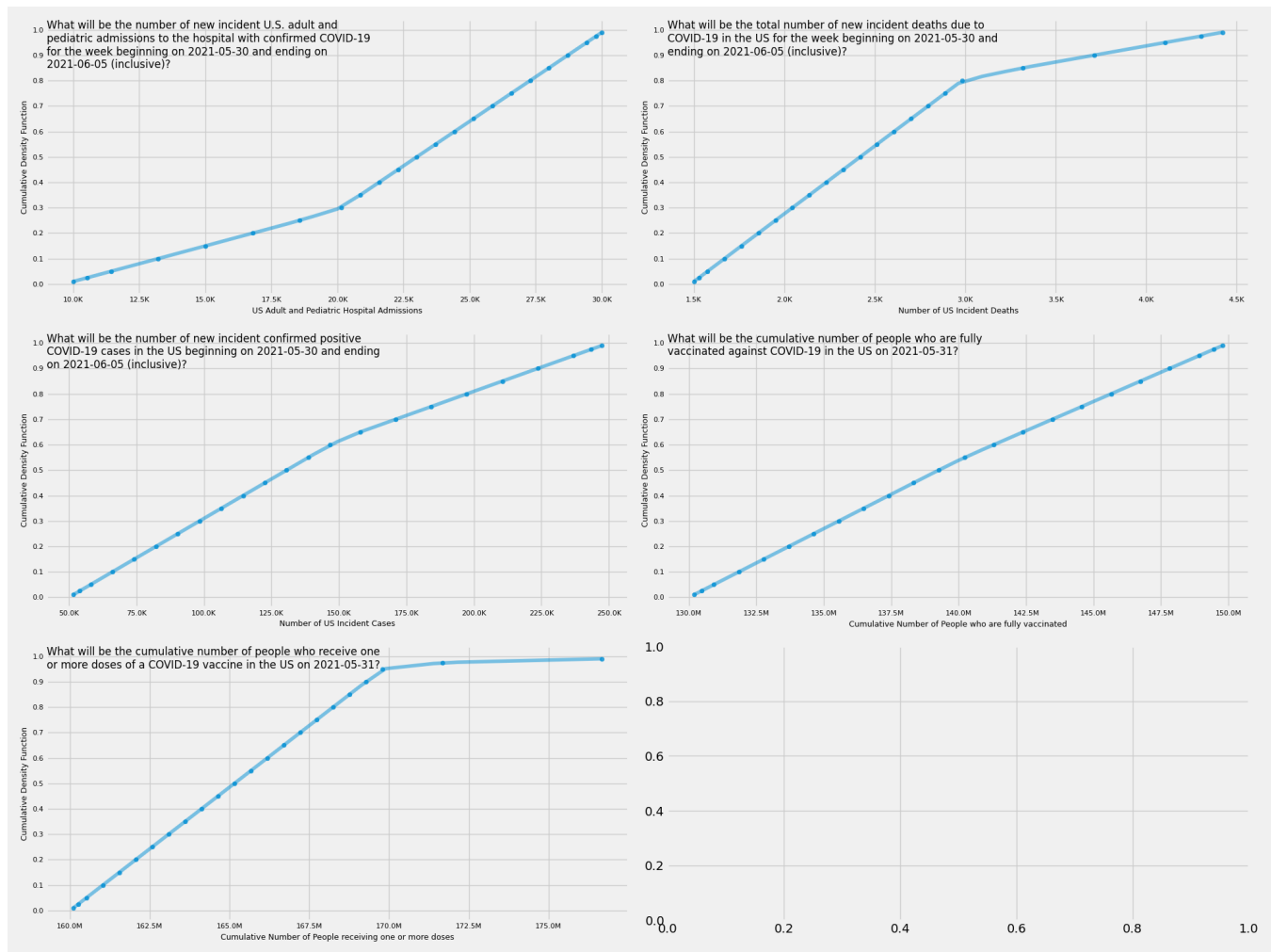


FIG. 3: Cumulative density functions of five targets of the US COVID-19 outbreak. A cumulative density function is the probability that the variable takes a value less than or equal to x . Plots are available for download at https://github.com/computationalUncertaintyLab/aggStatModelsAndHumanJudgment_PUBL/tree/main/data/GJOforecasts.

METAFORECASTS

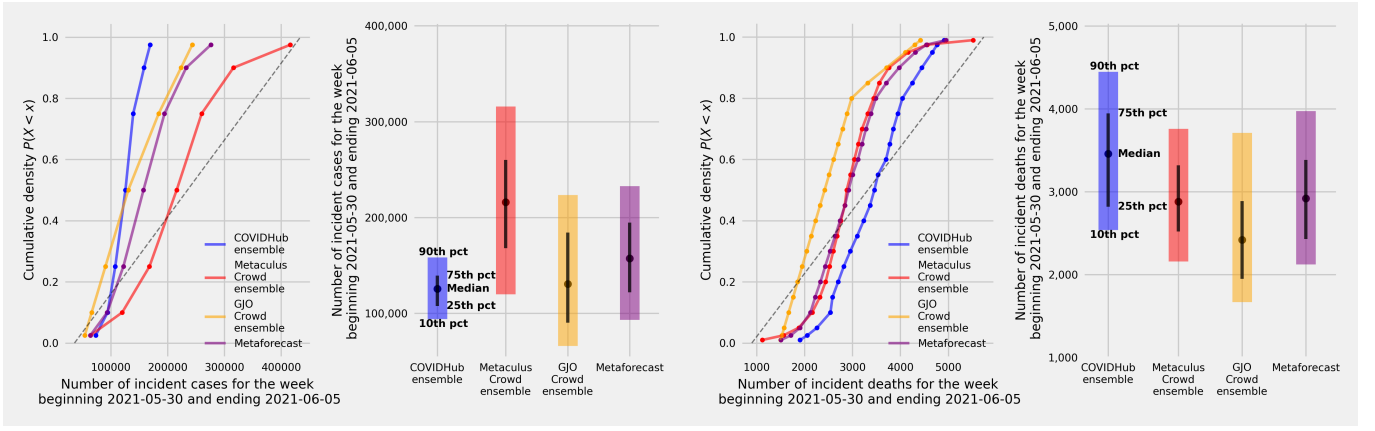


FIG. 4: Metaforecasts of the number of incident cases and incident deaths for the week beginning on May 30, 2021 and ending Jun 5, 2021. A metaforecast is a combination of an ensemble of computational models from the [COVID-19 Forecasthub](https://github.com/computationalUncertaintyLab/aggStatModelsAndHumanJudgment_PUBL/blob/main/metaforecasts/cases_survey4.png) 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 $F = \{f_1, f_2, f_3, \dots, f_C\}$ of predictive densities from C members of the crowd. Our consensus predictive density ($f_{\text{consensus}}$) is generated as

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

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

where f_i is the i th 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 ($\pi_i = \frac{1}{C}$) 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) monotonic 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 U.S. adult and pediatric admissions to the hospital with confirmed COVID-19 for the week beginning on 2021-05-30 and ending on 2021-06-05 (inclusive)?
- Range:[0 – 60K]
- Metaculus URL: <https://pandemic.metaculus.com/questions/7156/new-us-hospital-admissions-30-may-5-june/>
- GJO URL: <https://www.gjopen.com/questions/1996>

• Question 2

- What will be the total number of new incident deaths due to COVID-19 in the US for the week beginning on 2021-05-30 and ending on 2021-06-05 (inclusive)?
- Resolution Criteria: This question will resolve as the number of new deaths due to confirmed COVID-19 for the week beginning on 2021-05-30 and ending on 2021-06-05 (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 2021-05-30 will be computed by adding the number of new deaths from the 2021-05-30 up to, and including, 2021-06-05. The report will be accessed no sooner than 2021-06-14.
- Range:[0 – 8K]

- Metaculus URL: <https://pandemic.metaculus.com/questions/7157/new-us-covid-deaths-30-may-5-june/>
- GJO URL: <https://www.gjopen.com/questions/1997>

- Question 3

- Question: What will be the number of new incident confirmed positive COVID-19 cases in the US beginning on 2021-05-30 and ending on 2021-06-05 (inclusive)?
- Resolution Criteria: This question will resolve as the number of new confirmed cases beginning on 2021-05-30 and ending on 2021-06-05 (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 on 2021-06-14.
- Range:[0 – 800K]
- Metaculus URL: <https://pandemic.metaculus.com/questions/7158/new-us-covid-cases-30-may-5-june/>
- GJO URL: <https://www.gjopen.com/questions/1998>

- 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 2021-05-31?
- Resolution Criteria: This question will resolve as the cumulative number of people who have received one or more doses of a vaccine on 2021-05-31 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 2021-06-14
- Range:[148M – 185M]
- Metaculus URL <https://pandemic.metaculus.com/questions/7159/cumulative-1st-dose-us-vaccinations-31-may/>
- GJO URL: <https://www.gjopen.com/questions/1999>

- Question 5

- Question: What will be the cumulative number of people who are fully vaccinated against COVID-19 in the US on 2021-05-31?
- Resolution Criteria: This question will resolve as the cumulative number of people who receive one or more doses of a COVID-19 vaccine on 2021-05-31 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 2021-06-14.
- Range:[106M – 165M]
- Metaculus URL: <https://pandemic.metaculus.com/questions/7160/cumulative-fully-vaccinated-in-us-on-31-may/>
- GJO URL: <https://www.gjopen.com/questions/2000>

- 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 2022-01-09.
- Range:[575K – 1.1M]
- Metaculus URL: <https://pandemic.metaculus.com/questions/7161/cumulative-us-covid-deaths-by-end-of-2021/>

- Question 7

- Question: What will be the % prevalence of SARS-CoV-2 variants thought to partially escape immunity for the two-week period 23 May - 05 Jun 2021?
- Resolution Criteria: This question will resolve on the basis of the first update that shows figures for the two-week period ending 05 Jun of the "Weighted Estimates of Proportions of SARS-CoV-2 Lineages" table on the U.S. CDC's "[Variant Proportions](#)" page. The percentages of variants that cause "reduced neutralization by convalescent and post-vaccination sera" will be added up. If between now and 05 Jun there are additional variants classified by the CDC as variants that cause "reduced neutralization by convalescent and post-vaccination sera," these will count toward the total percent figure. Likewise, if any of the variants that are currently classified as causing partial immune escape are removed from being classified as such, they will no longer count toward the total percent figure.
- Range:[0 – 100]
- Metaculus URL: <https://pandemic.metaculus.com/questions/7164/prevalence-of-immune-evading-variants-5-june/>

- 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 2025, then this resolves as 31 December 2024.
- Range:[2021 – 05 – 01 – > 2024 – 12 – 31]
- Metaculus URL: <https://pandemic.metaculus.com/questions/7163/variant-of-high-consequence-before-2025/>