Supplemental for: Assessing Human Judgment Forecasts in the Rapid Spread of the Mpox Outbreak: Insights and Challenges for Pandemic Preparedness

Thomas McAndrew,[1] Maimuna S. Majumder,[2,3] Andrew A. Lover,[4] Srini Venkatramanan,[5] Paolo Bocchini,[6] Tamay Besiroglu,[7,8] Allison Codi,[9] Gaia Dempsey,[8] Sam Abbott,[10] Sylvain Chevalier,[8] Nikos I. Bosse,[10,8] Juan Cambeiro,[11,8] and David Braun[12]

[1] *Department of Community and Population Health, College of Health, Lehigh University, Bethlehem, Pennsylvania, United States of America**

[2] *Boston Children's Hospital, Boston, MA, United States of America*

[3] *Harvard Medical School, Boston, MA, United States of America*

[4] *Dept. of Biostatistics and Epidemiology, School of Public Health and Health Sciences, University of Massachusetts Amherst, Amherst, Massachusetts, United States of America*

[5] *Biocomplexity Institute and Initiative, University of Virginia, Charlottesville, Virginia, United States of America*

[6] *Department of Civil and Environmental Engineering, P.C. Rossin College of Engineering and Applied Science, Lehigh University, Bethlehem, Pennsylvania, United States of America*

[7] *MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, Massachusetts, United States of America*

[8] *Metaculus, Santa Cruz, California, United States of America*

[9] *Department of Biostatistics, Rollins School of Public Health, Emory University, Atlanta, Georgia, United States of America*

[10] *Centre for Mathematical Modelling of Infectious Diseases, London School of Hygiene & Tropical Medicine, Keppel St, London, WC1E 7HT, UK*

[11] *Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, United States.*

[12] *Department of Psychology, Lehigh University, Bethlehem, Pennsylvania, United States of America*

Appendix A: Supplemental

1. Forecast submission

A forecaster on the Metaculus platform can submit a predictive density f as a convex combination of up to five logistic distributions

$$f(x) = \sum_{i=1}^{5} \pi_k \, g(x|\mu_k, s_k) \tag{A1}$$

$$g(x) = \frac{\exp\left(-(x-\mu)/s\right)}{s(1+\exp\left(-(x-\mu)/s\right))} \tag{A2}$$

where π_k is a non-negative weight associated with the kth logistic distribution $g(x|\mu_k, s_k)$ and all weights are constrained to sum to unity. By default, a forecaster is presented with a single logistic density with a slider bar underneath this density that contains a square and two circles (see supplemental figure S1). A forecaster can change the value of μ by shifting the square to the left for lower values and to the right for higher values. The parameter s is adjusted by sliding the circles left and right independently. If they wish, a forecaster may press "add a component" to add additional logistic distributions (up to 5 component distributions in total). A second slider will appear under the first and two additional sliders that control the weights (π) associated with the first and second logistic distribution. The predictive density a user creates is continually updated in the browser to facilitate iteratively building their forecast for submission.They are also shown the Community Forecast [needs to be defined] along with a Current Points calculation reflecting the points they may received based on the resolution (see `https://www.metaculus.com/help/faq/#howscore`

* mcandrew@lehigh.edu

for more details). When a forecaster is satisfied with their prediction, they press the submit button. After the first submission, a forecaster may revise their original prediction as many times as they choose until the question is closed on the platform or the ground truth is reported.

Forecasters construct their predictive density over a bounded interval, presented either on a linear scale or on a log scale (base 10). The choice of intervals, and linear vs. log scale, was made by forecast question developers at Metaculus. After 10 predictions from 10 unique forecasters were submitted to the Metaculus platform, an ensemble of these predictions was revealed to the community. In addition to a formatted method for building a forecast, users were also able to post comments that could be viewed by all other users of the Metaculus system.

2. Ensemble building

An ensemble predictive density f in response to a question posed on the Metaculus platform was created as a weighted combination of all M individual forecaster densities

$$f(x) = \sum_{m=1}^{M} \pi_m f_m(x) \tag{A3}$$

where f_m is a predictive density submitted by user m and the weights assigned to forecaster m are annotated as π_m. Weights could be a function of the forecaster's accuracy on past questions and of how recently they submitted a prediction, but in this study we chose to weight all forecasts equally (i.e., we assigned $\pi_m = 1/M \ \forall m$). The above ensemble strategy is used for questions where a forecaster submits a predictive density over a closed interval (a continuum). For questions entailing answers in the form of a scalar number, such as when forecasters are asked to submit a single probability of a particular event (see question (vi) above), the ensemble strategy build an empirical cumulative density function from predicted probabilities (See supplemental figure S2 bottom right panel for an example).

Question	Link to Metaculus question	Data source	Scale
Total cases in Europe	`https://www.metaculus.com/questions/10978/total-monkeypox-cases-in-europe-july-1-2022/`	Global health	Logarithmic
Total cases in US	`https://www.metaculus.com/questions/10979/total-monkeypox-cases-in-usa-on-july-1-2022/`	Global health	Logarithmic
Total cases in Canada	`https://www.metaculus.com/questions/11039/monkeypox-cases-in-canada-on-july-1-2022/`	Global health	Logarithmic
Countries with at least one case	`https://www.metaculus.com/questions/10975/countries-with-monkeypox-by-july-31-2022/`	Global health	Natural
Number of US states	`https://www.metaculus.com/questions/10981/-us-states-with-monkeypox-cases-on-july-1/`	CDC	Natural
WHO declaration of PHEIC	`https://www.metaculus.com/questions/10977/who-declares-that-monkeypox-is-a-pheic/`	WHO	Natural

TABLE (S1) List of questions, links to forecast solicitation on Metaculus, data source, and forecast scale. In-depth criteria to determine the ground truth can be found under the heading 'Resolution criteria' for each Metaculus link.

These questions were: (i) what will be the total number of confirmed and suspected monkeypox cases in Europe as of July 1, 2022?, (ii) what will be the total number of confirmed and suspected monkeypox cases in the United States as of July 1, 2022?, (iii) what will be the total number of confirmed and suspected monkeypox cases in Canada as of July 1, 2022? (iv) how many countries will confirm at least one case of monkeypox by July 31, 2022?, (v) how many states in the United States will report having a case of monkeypox as of July 1, 2022?, and (vi) will the World Health Organization declare the spread of monkeypox a Public Health Emergency of International Concern before 2023? We received a total of 1275 probabilistic predictions (initial forecasts or subsequent updates) from 442 individual forecasters.

FIG. (S1) Example of the Metaculus platform used by forecasters to generate probabilistic predictions of questions posed about monkeypox.

Appendix B: Figures

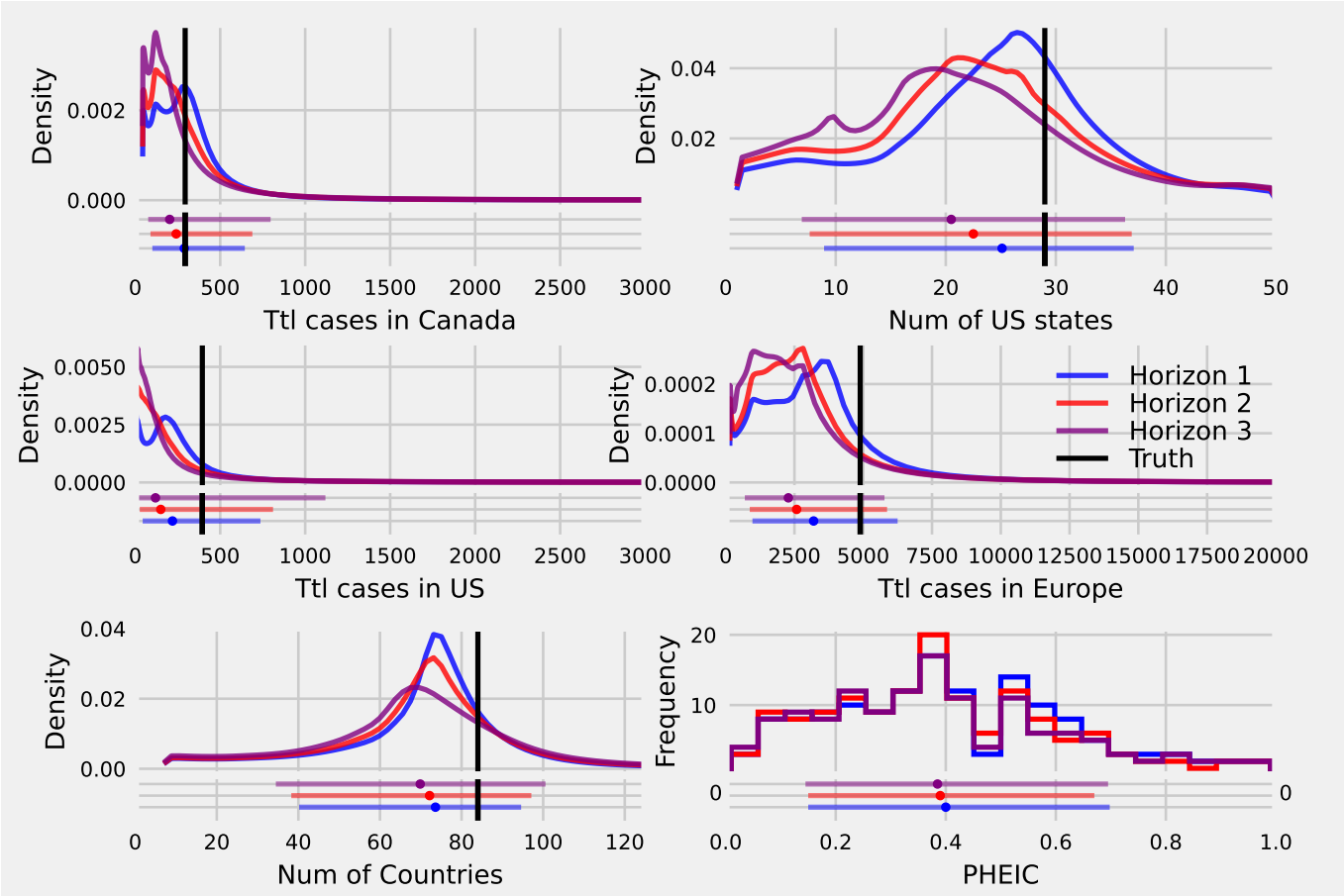

FIG. (S2) Human judgment ensemble predictive densities one (blue), two (red), and three (purple) weeks ahead of the date of the true value (black vertical line). Below each plot is presented a 25th quantile (leftmost point of the line), 50th quantile (dot), and 75th quantile (rightmost point of the line).

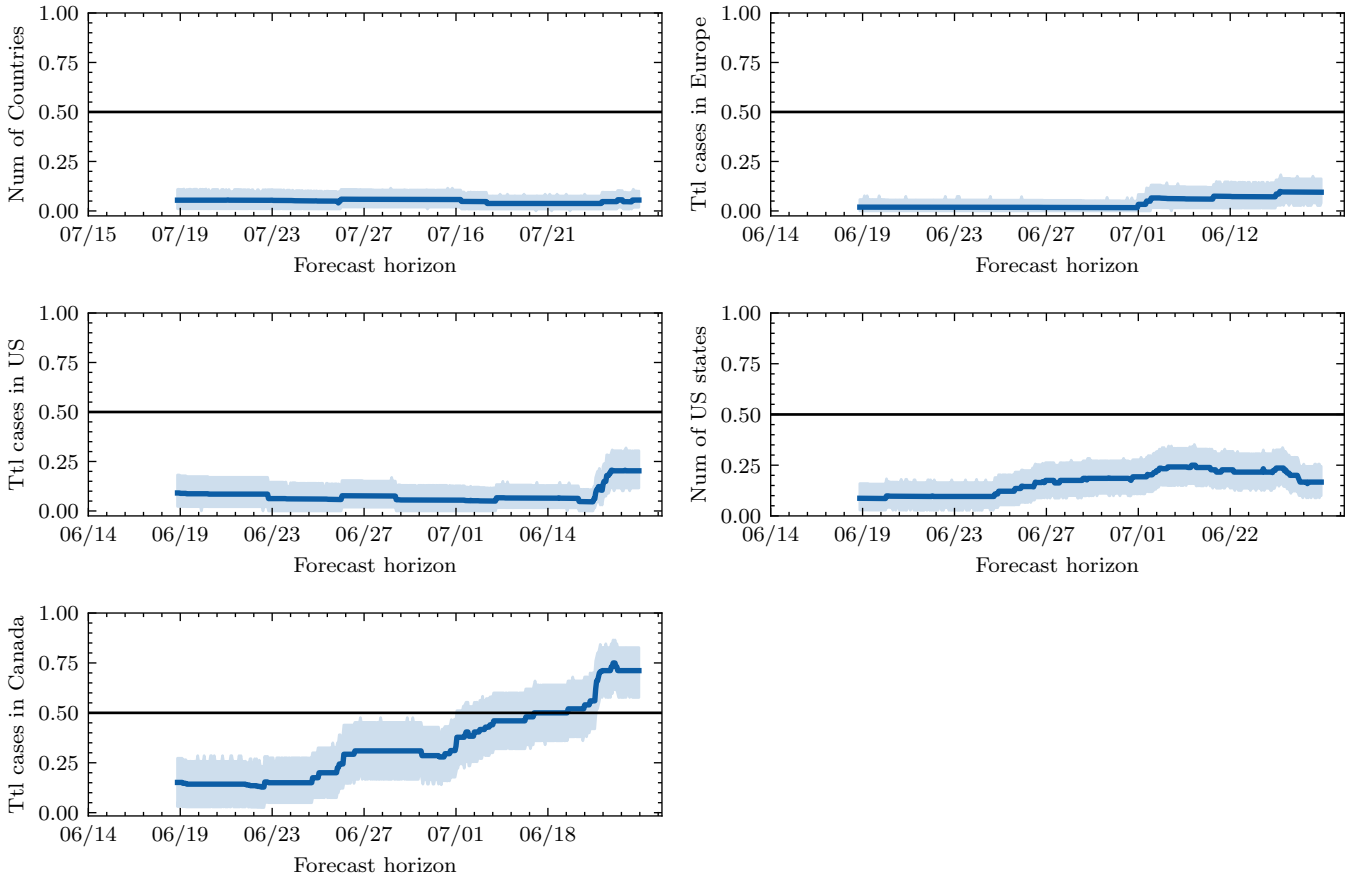

FIG. (S3) The proportion plus 95% confidence intervals of individual median predictions that were greater than the ground truth for five continuous targets from three weeks before the ground truth to one week before the ground truth in intervals of one hour. The majority of human judgment forecasters submitted densities with a median below the ground truth. Only when asked to predict the total number of cases in Canada did human forecasters over estimate the severity of the mpox outbreak.

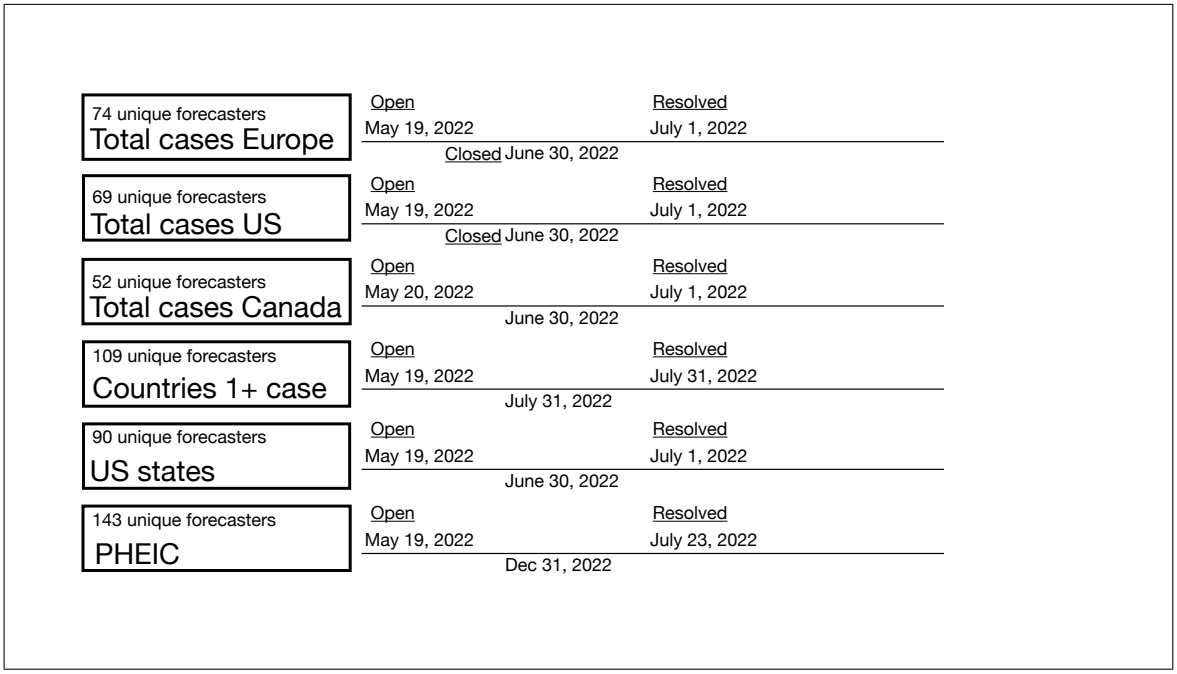

FIG. (S4) Visual timeline that shows: when the question opened for forecasting, when the question closed, and when the ground truth was resolved. In addition, we present the number of unique forecasters that submitted at minimum one forecast.