Eliciting Beliefs about COVID-19 Prevalence and Mortality: Epidemiological Models Compared with The Street

by

Glenn W. Harrison, Andre Hofmeyr, Harold Kincaid, Brian Monroe, Don Ross, Mark Schneider and J. Todd Swarthout⌘

November 2020

ABSTRACT

Subjective belief elicitation about uncertain events has a long lineage in the economics and statistics literatures. Recent developments in the experimental elicitation and statistical estimation of subjective belief distributions allow inferences about whether these beliefs are biased, and the confidence with which they are held. Beliefs about COVID-19 prevalence and mortality interact with risk management efforts, so it is important to understand relationships between these beliefs and publicly disseminated statistics, particularly those based on evolving epidemiological models. The pandemic provides a unique setting over which to bracket the range of possible COVID-19 prevalence and mortality outcomes given the proliferation of estimates from epidemiological models. We rely on the epidemiological model produced by the Institute for Health Metrics and Evaluation together with the set of epidemiological models summarised by FiveThirtyEight to bound prevalence and mortality outcomes for one-month, and December 1, 2020 time horizons. We develop a new method to partition these bounds into intervals, and ask subjects to place bets on these intervals, thereby revealing their beliefs. The intervals are constructed such that if beliefs are consistent with epidemiological models, subjects are best off betting the same amount on every interval. We use an incentivised experiment to elicit beliefs about COVID-19 prevalence and mortality from 598 students at Georgia State University, using six temporally-spaced waves between May and November 2020. We find that beliefs differ markedly from epidemiological models, which has implications for public health communication about the risks posed by the virus.

Keywords: subjective beliefs, beliefs, COVID-19 mortality, COVID-19 prevalence

Declarations of interest: none

⌘ Department of Risk Management & Insurance and Center for the Economic Analysis of Risk (CEAR), Robinson College of Business, Georgia State University, USA (Harrison); School of Economics, University of Cape Town, South Africa (Hofmeyr, Kincaid); Research Unit in Behavioural Economics and Neuroeconomics (Hofmeyr, Kincaid, Ross); School of Philosophy and School of Economics, University College Dublin, Ireland (Monroe); School of Society, Politics and Ethics, University College Cork, Ireland, School of Economics, University of Cape Town, South Africa, and Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University, USA (Ross); Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University, USA (Schneider); and Department of Economics, Andrew Young School of Policy Studies, Georgia State University, USA (Swarthout). Harrison is also affiliated with the School of Economics, University of Cape Town. E-mail contacts: gharrison@gsu.edu, andre.hofmeyr@uct.ac.za, harold.kincaid@uct.ac.za, brian.monroe@ucd.ie, don.ross931@gmail.com, mschneider@gsu.edu, and swarthout@gsu.edu. We are grateful to the Center for the Economic Analysis of Risk for funding this research; in all other respects the funder had no involvement in the research project.
Corresponding author: Andre Hofmeyr (andre.hofmeyr@uct.ac.za)

1. Introduction

Beliefs that individuals hold about COVID-19 prevalence and mortality interact with efforts to manage the risks of the virus. A core concern is the relationships between these beliefs and publicly disseminated statistics, particularly statistics based on evolving epidemiological models. The COVID-19 pandemic provides an important setting to study this relationship because of the role that epidemiological models have played in public debate, and understandable biases in early editions of models that became evident over a relatively short period of time. Public awareness of the extent to which official statistics about COVID-19 in the United States (US) might be biased, due to political influences, poses an additional challenge when studying this relationship. To what extent did the beliefs of individuals evolve with the forecasts of epidemiological models? To what extent did the beliefs of individuals evolve with the official reports from the Centers for Disease Control and Prevention (CDC)? To what extent did these trends affect the confidence of individual beliefs over time?

We examine these issues by eliciting the subjective beliefs of 598 students at Georgia State University using incentivized forecasting tasks about expected COVID-19 prevalence and mortality in the US. Our methods are designed to bracket the range of possible beliefs that individuals have, and critically to assess their individual confidence in those beliefs. We also developed a method that allows us to directly identify whether beliefs tracked those of some epidemiological models, quite apart from the elicitation of beliefs to address the broader questions posed above. To ensure that we were able to observe changes over time, we administered six temporally-spaced waves between May and November 2020, with different respondents selected at random for each wave.

To anchor COVID-19 prevalence and mortality outcomes for the elicitation of beliefs over horizons of one month, and over horizons to December 1, 2020, we relied in part on one prominent epidemiological model, from the Institute for Health Metrics and Evaluation (IHME) at the University of Washington (http://www.healthdata.org). The IHME model has produced publicly disseminated daily forecasts of both infections and deaths throughout the course of the pandemic. We also made use of the evolving set of epidemiological models featured by FiveThirtyEight (fivethirtyeight.com), which ranged from 6 to 14 models over the course of our study, to complement the IHME model. We develop a method to partition the possible outcomes presented to subjects into intervals or *bins*, such that if a subject were to hold beliefs consistent with the epidemiological models, including allowance for statistical error, she would bet the same amount on every bin. A remarkable feature of our method is that it allows direct inferences about the extent to which distributions of expectations diverge from these model-based forecasts. More extensive inferences, beyond testing this specific null hypothesis, will require more structural statistical modelling, and will be undertaken in subsequent analyses.

We find that beliefs diverge markedly from the epidemiological models, with immediate implications for public health communication about the risks posed by the virus.

2. Material and Methods

The importance of eliciting subjective beliefs about uncertain events has long been clear across many disciplines. The earliest attempts to measure beliefs came from survey questions [1, 2]. These have become increasingly sophisticated, with researchers now trying to elicit whole belief distributions for non-binary events [3, 4], such as the levels of COVID-19 infections and deaths that are our focus. However, surveys do not incentivize the truthful revelation of beliefs, and there is substantial evidence that using hypothetical surveys can be unreliable [5]. Our use of an incentive-compatible mechanism to elicit beliefs makes our approach fundamentally different to survey responses, and more informative. The concept of subjective belief was formally developed in economics and decision theory as an extension of the notion of revealed preference [6]. Just as the strength of preferences for fine wine over plonk can be revealed by *purchase* decisions when the relative prices of the two types of wine are varied, beliefs can be revealed by *betting* decisions that depend on a particular outcome, such as the level of COVID-19 infections, reported by a certain source, such as the CDC, on a specific day in the future.

A key development in the reliable elicitation of subjective beliefs was operationalizing this notion of beliefs revealed by betting in terms of changes in betting decisions as the relative odds offered by bookies are varied. Imagine an array of bookies, lined up in terms of their odds that the COVID-19 infection rate will go up in the next month, rather than stay the same or go down. Some bookies offer great odds that it will go up, and some offer great odds that it will go down, and there are many bookies in between. Now allow someone to place a bet of

$1 with each bookie. If the bettor is risk neutral, the point at which they switch from betting that infections will go up to betting that they will not go up tells us the odds that this person places on these events, and from those odds we can infer the person's subjective probability of infections going up.

It is a small formal step to present this array of bookies in the form of a "scoring rule," which translates different bets into payoffs for the bettor, depending on the realized outcome or event [7, 8]. And in turn we can generalize these ideas to placing bets on several events, such as the event that infection rates go up by more than 1 percentage point, the event that they go up by between 0 and 1 percentage points, and the event that they go down. In this way we can elicit the subjective probability mass function over these events, or indeed the probability distribution function for continuous events [9]. Or we can divide a continuous event, such as the level of COVID-19 infections by June 30, 2020, into 10 bins that partition the event space over which we seek to elicit beliefs, as in the experiment on which we report. And since we are asking people to place bets with simulated bookies, with varying odds defined by a scoring rule, this is easy to do with real money, and thereby provide incentives for truthful revelation of beliefs *cum* bets by using "proper" scoring rules [10].

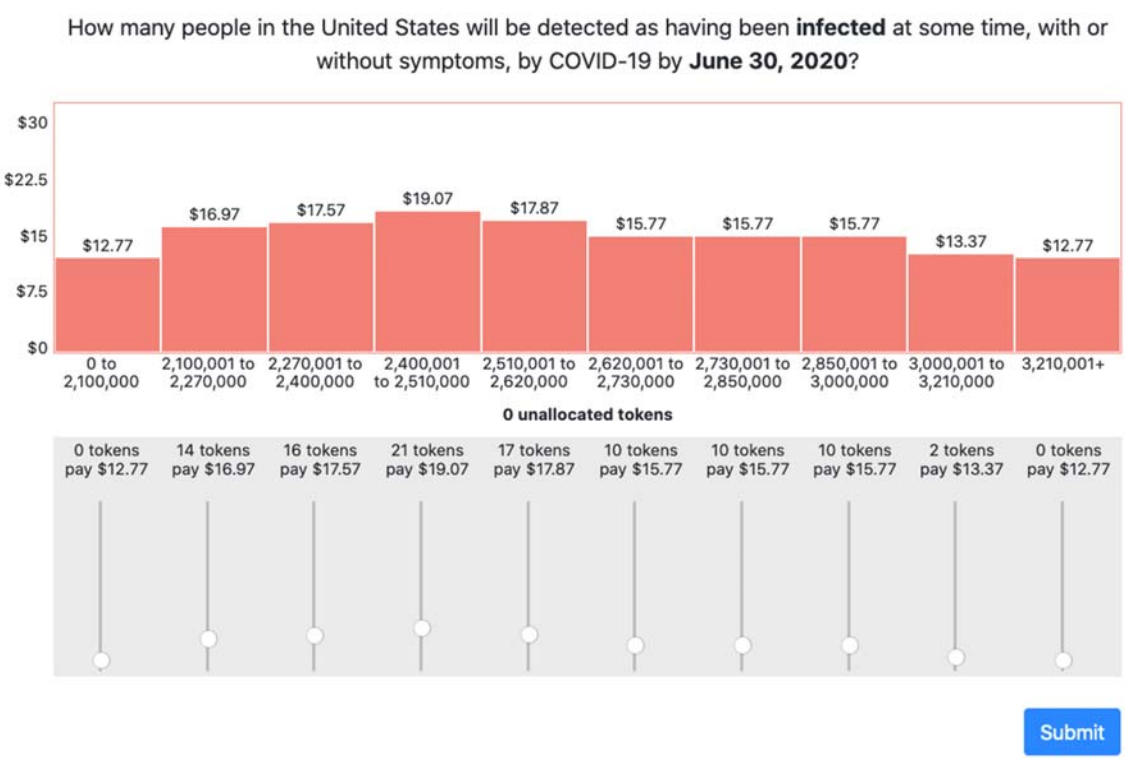

Figure 1: Subjective Belief Task Interface and Bets of Subject #183 on May 29, 2020

Figure 1 is a screenshot of the experimental software we developed to elicit the beliefs of each subject about COVID-19 prevalence and mortality. This subjective belief question was presented to subjects during Wave 1 of our study, which took place on May 29, 2020. Figure 1 shows the actual bets, in the form of a token allocation, of subject #183, and the amount to be paid depending on the answer to the question. The answer was verified using the first public report provided by the CDC *after* the date in the question, which was explained to subjects through audio-visual instructions before they completed the subjective beliefs task.

Armed with probability mass functions over ten events, as represented in Figure 1, which characterize subjective belief distributions over changes in COVID-19 prevalence and mortality, we can analyse the bias and confidence of those beliefs. Bias is just the familiar concept from statistical estimation: how different is the weighted average belief from the realized event, or the best available econometric or epidemiological model at the time [11], or the claims of political leaders? All are actually useful metrics for different reasons, so there is not just one measure of bias that is of interest. Confidence is just the familiar concept of imprecision from statistical estimation, most commonly captured by the variance of beliefs about their mean. We prefer to think of confidence more broadly to reflect the variability of beliefs, so we can also consider skewness and kurtosis, but the point is to pay attention to more than just the weighted average or mode of beliefs. One can only characterize bias and confidence if one elicits subjective belief *distributions* [12], which of course allow for the special case of degenerate beliefs held with certainty. The better, Bayesian epidemiological models of COVID-19 infections and deaths provide posterior predictive distributions of future levels, which can be used to also make determinations of whether subjective beliefs are "overconfident" or "insufficiently confident" [13].

Figure 2 shows the realized answer, as reported by the CDC, to the question from Figure 1, and hypothetical bets that vary according to whether they are biased relative to the number of infections by June 30, 2020, and the confidence with which these beliefs are held. Per the experimental protocol, the official reports from the CDC are treated as the correct answer that determine subject payments. The top left quadrant of the figure represents an unbiased, but relatively low confidence, set of bets, in the sense that the largest bet was placed on the correct answer, but bets were also made on other events. The bottom left quadrant also represents unbiased beliefs, but held with a degenerate level of confidence in the sense that all tokens were bet on the correct event. The two right quadrants represent biased beliefs because no tokens were allocated to the correct event, but clearly differ according to the strength with which beliefs were held.[1]

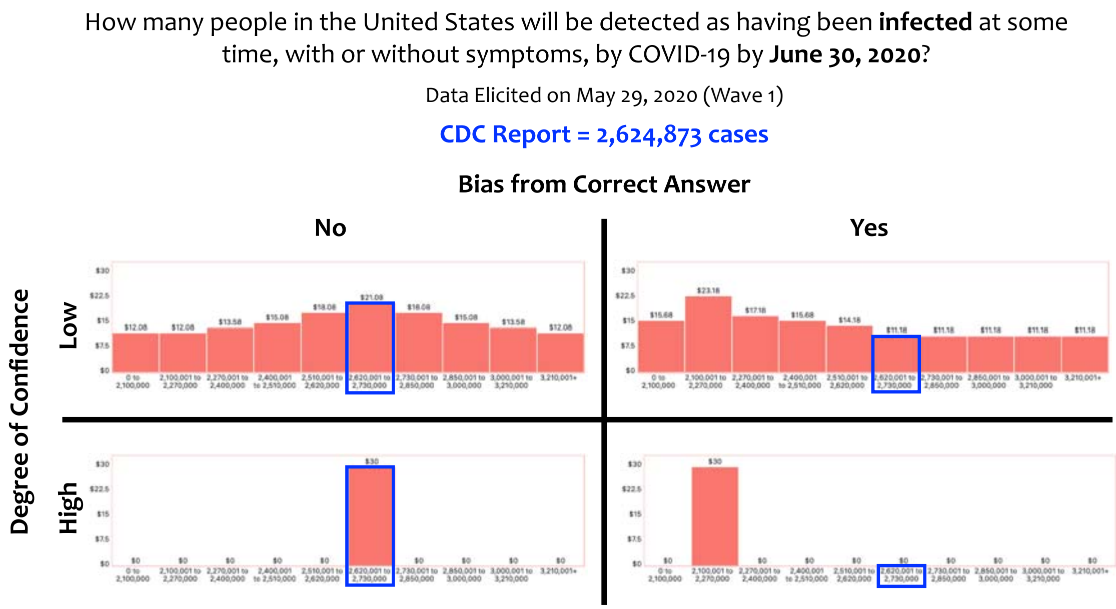

Figure 2: Bias and Confidence of Subjective Belief Distributions

A direct implication of incentivizing bets with a proper scoring rule is that if someone believes that each event, as represented by the bins in a task, is equally likely to occur, the person will bet exactly the same amount on each bin, as represented in Figure 3. Thus, when someone bets anything other than the same amount on every bin, this reveals that they do not consider every event as equiprobable. We constructed bins over which to elicit beliefs about the number of infections and deaths due to COVID-19 in the US either one month in the future or by December 1, 2020. These bins were constructed such that if a person's bets differ across bins, this reveals that the person's beliefs deviate from epidemiological models, described in more detail below, of infections and deaths due to COVID-19.

[1] We used a quadratic scoring rule (QSR) to incentivize truthful revelation of beliefs. As a proper scoring rule, the QSR provides the highest *expected* reward if risk neutral subjects report their true beliefs, and therefore penalizes subjects for betting on events to which they do not assign positive probability. Unless a subject reports degenerate beliefs, as in the bottom left or right quadrant of Figure 2, the QSR still provides payment for bins to which no tokens have been allocated, as in the top left or right quadrant of Figure 2.

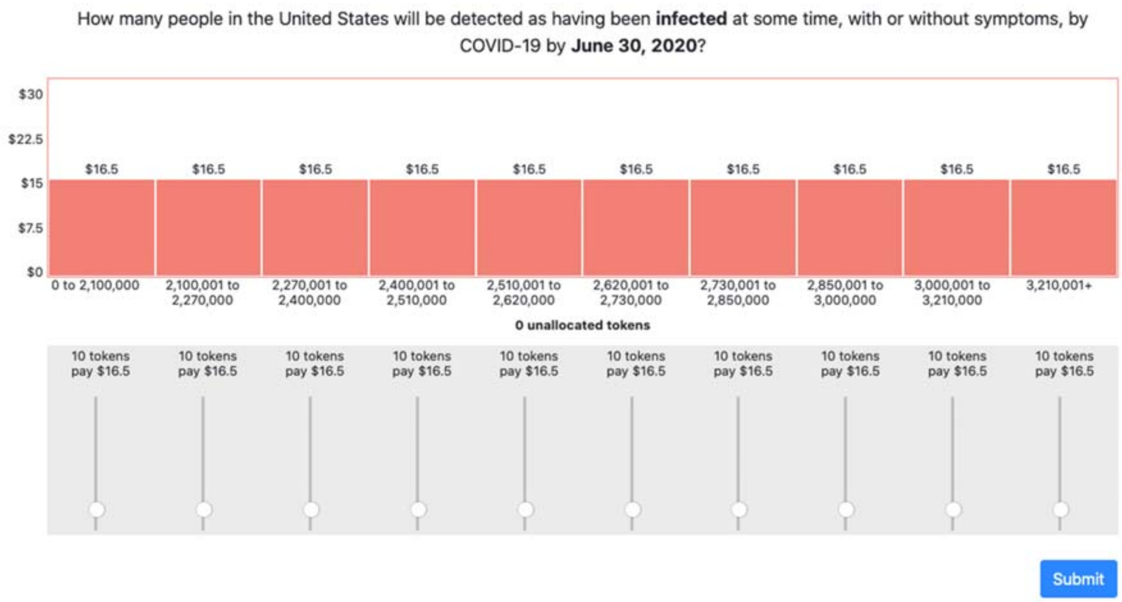

Figure 3: Bets for Equiprobable Events

The first step in constructing these bins is to define the distribution of underlying events. We assumed that deaths and infections, scaled to the population of the US, follow a beta distribution. The beta distribution is flexible enough for our purposes, has well-defined higher moments, and finite support over an interval. This latter property ensures that the number of people who will be infected or die due to COVID-19 cannot be negative or greater than the population of the US. In addition, the beta distribution is well suited to characterizing the bias and confidence of subjective belief distributions. Finally, it has two sufficient statistics, and therefore the shape of the distribution can be defined by two points, or *anchors*, along its cumulative distribution function (CDF), if the cumulative density at each anchor is known or imposed.

We therefore set out to define pairs of anchors that consist of a lower anchor, such that there is a probability of 0.1 that the true statistic would be less than this amount, and an upper anchor, such that there is a probability of 0.2 that the true statistic would be greater than this amount.[2] For each pair of anchors, a beta distribution was defined that uniquely satisfies these two sufficient statistics. This beta distribution was then used to define 10 bins such that each bin represented 10% of the full distribution's cumulative density. This bin construction exercise ensures that if a person's beliefs were the same as the distribution we defined, which

[2] Appendix A discusses the bin anchoring calculations we performed for each wave of the study.

was based on epidemiological models, they would be best off by betting the same amount on each bin.

Our method for designing which bins to present to subjects was intended to provide general information about the beliefs of individuals that reflected our hyper-priors about the underlying data-generating process. Our method also served to generate a sharp, direct test of the specific null hypothesis that the beliefs of individuals tracked those that come from epidemiological models. And by the beliefs from individuals and the models "tracking" each other, we mean much more than the weighted average: we insist that they also track each other in terms of level of confidence. This additional criterion allows us to determine if the evolution of epidemiological understanding and modeling, during the period of our elicitation, is matched by an evolution of individual beliefs.

To effect this test of the null hypothesis, we need some characterization of the beliefs that come from "epidemiological models." To do that we started with the IHME model, and used the forecasts that it provided to generate the bins we refer to as frame 0.[3] And, more specifically, our method generated bins that implied equal weight would be given to each bin, in terms of bets implemented with token allocations. Proper scoring rules incentivize the truthful revelation of beliefs of risk neutral bettors. There are deep theoretical, experimental, and statistical issues that arise when agents are *not* risk neutral, because then they can make bets to hedge against risk [10]. For example, an *extremely* risk averse decision maker might bet the same amount on every bin in a subjective beliefs task to ensure zero variance in payment, regardless of the event that is realized. Thus, if the subjects in our experiment all bet the same amount on every bin, we would be unable to directly infer whether this was due to high levels of risk aversion or to beliefs that are consistent with epidemiological models.[4] However, to the extent that subjects do not bet the same amount on every bin, this implies that their beliefs are not consistent with the epidemiological models *regardless* of their levels of risk aversion. This property is a powerful innovation in the methods applied: it is apparent that risk preferences of individuals *only* matter if subjects do not bet the same amount on

---

[3] See [14] for a review of the historical context, modeling assumptions, accuracy, and criticisms of the IHME model.

[4] During the experimental session, we also elicited the risk attitudes of each subject to account for the possibility of hedging in the elicitation and estimation of subjective beliefs. We do not focus on risk attitudes here, because they are unnecessary for our inferential objective.

every bin. Hence we are able to test this null hypothesis by directly comparing the token allocations we observe from individuals, without any adjustments for their risk preferences.

The specific epidemiological model used for frame 0 was then used as the basis for adjustments to generate the “hybrid, representative” epidemiological models reflected in frames 1, 2, and 3. Apart from allowing us to test for wholesale deviations from the hyper-priors reflected in frame 0, these frames themselves can be viewed as reflecting the beliefs of epidemiological models, with the constraint added by our method that someone holding beliefs consistent with those models would bet exactly the same amount on every bin. Although the bins for frames 1, 2, and 3 do not reflect a specific epidemiological model in the manner that the bin labels for frame 0 do, they should be viewed as systematically representing a wider range of epidemiological models. In this sense, our complete set of frames is designed to reflect “epidemiological models” as a whole, respecting the inevitable changes in the number of epidemiological models, and modeling assumptions of those models, over the course of the pandemic.

One frame per belief question was drawn randomly for each subject, so frames varied between subjects in the task. As discussed, the construction of these frames allows us to draw inferences about whether non-expert subjective beliefs differ from expert forecasts encoded in epidemiological models to the extent that bets vary from bin to bin. We investigate the distribution of subject bets on COVID-19 prevalence and mortality in the next section.

3. Results

Our full sample consists of 598 student subjects from Georgia State University whose beliefs were elicited monthly between May and November 2020. We focus here on subject bets implemented by token allocations for the one-month timeframe across waves 1–4 of our study. We limit our analyses to this timeframe and these waves for ease of exposition, and because we know the correct answers to these questions, as reported by the CDC. This sample consists of 458 subjects across the four waves, with 112 subjects in Wave 1, the data from whom are represented in the figures below.

Figure 4 shows the distribution of token allocations from May 29, 2020 (Wave 1) for the number of COVID-19 infections in the US by June 30, 2020. The distributions differ markedly across frames, which suggests that the way in which event spaces are anchored and partitioned affects subjects' token allocations, even though each set of anchors were defined by epidemiological models of the pandemic. However, to draw valid inferences about differences across frames both with respect to each other and the number of cases reported by the CDC, it is essential to account for the risk attitudes of subjects [10]. This is not the focus of our analyses here. Despite these apparent differences across frames, the crucial result is that subjects did not bet the same amount on every bin. We tested this formally using the nonparametric Epps-Singleton (ES) test [15]. ES tests of the null hypothesis that the token allocations in frames 0–3 are uniformly distributed are rejected at any standard level of statistical significance ($p < 0.001$ in all comparisons). Thus, subjects did not bet the same amount on every bin, which is a necessary condition for the beliefs of subjects to be consistent with epidemiological models of the spread of the virus. If some of the token allocation distributions were more flat than others, this would suggest that the epidemiological model associated with that frame was more closely aligned with the beliefs of subjects, but clearly no such inference is valid on the basis of the distributions in Figure 4.

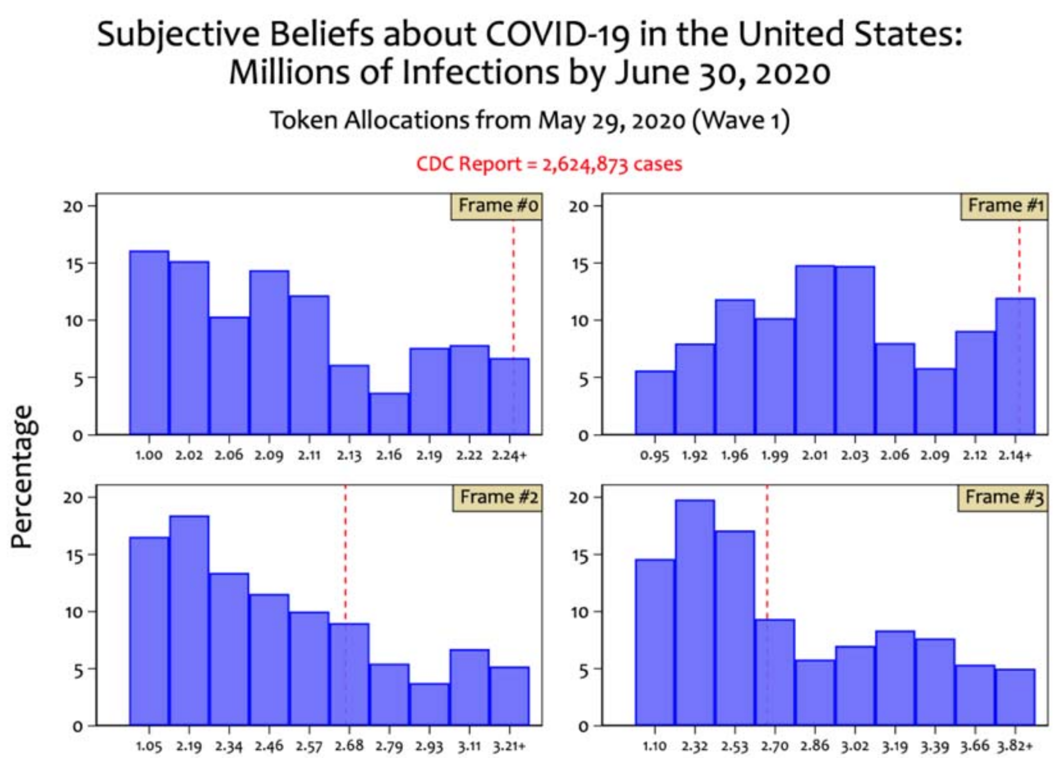

Figure 4: Beliefs about COVID-19 Infections in the US by June 30, 2020

Appendix B shows the distribution of token allocations elicited in waves 2–4 of our study of the number of COVID-19 infections in the US one month after the date of each wave. While there are some interesting differences across waves, which reflect the (rapid) evolution of the pandemic in the US, better scientific understanding of the spread of the virus, and the proliferation of epidemiological models that had more data to feed their predictions, the overall pattern is the same: subjects' beliefs differ significantly from epidemiological models of COVID-19 infections ($p < 0.001$ in all comparisons).

Figure 5 shows the distribution of token allocations from May 29, 2020 (Wave 1) for the number of COVID-19 deaths in the US by June 30, 2020. Unlike infections, the distributions across frames are similar, but formal tests of the extent to which they differ require adjustments for risk attitudes. Again, the crucial result is that subjects did not bet the same amount on every bin. Tests of the null hypothesis that the token allocations in frames 0–3 are uniformly distributed are rejected ($p < 0.001$ in all comparisons). Thus, the beliefs of subjects about COVID-19 deaths are not consistent with epidemiological models.

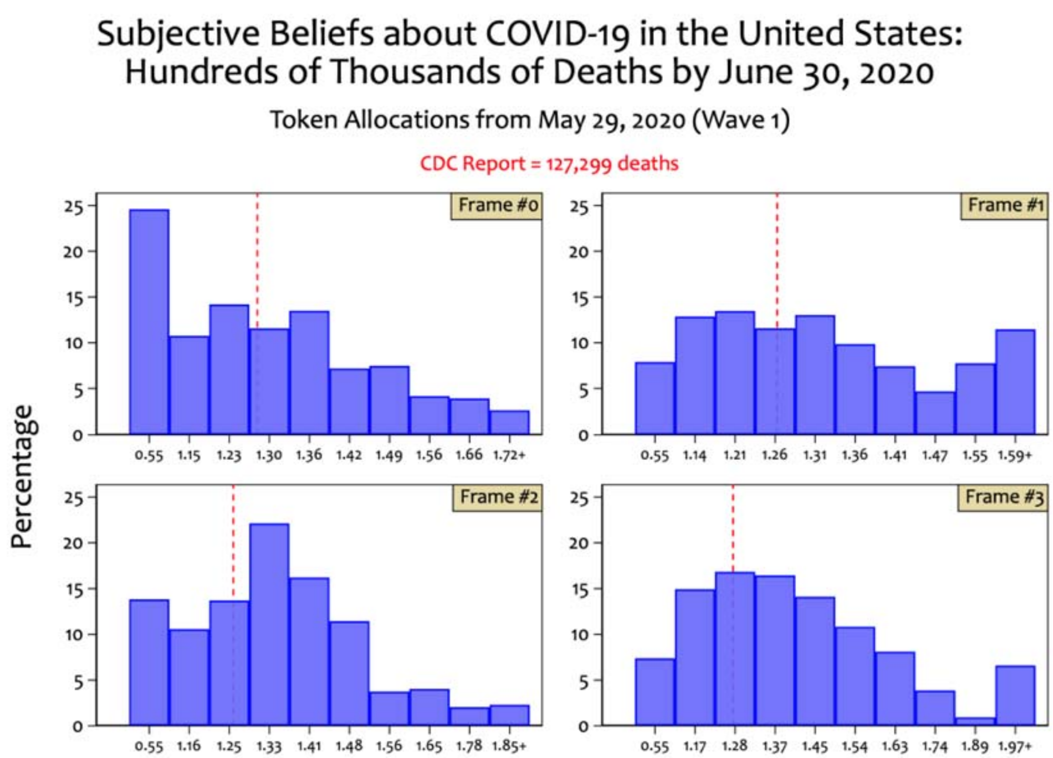

Figure 5: Beliefs about COVID-19 Deaths in the US by June 30, 2020

Appendix C shows the distribution of token allocations in waves 2–4 of the number of COVID-19 deaths in the US one month after the date of each wave. Differences across waves are less pronounced in comparison to beliefs about COVID-19 prevalence, but subjects' beliefs clearly differ from epidemiological models of deaths attributed to the virus ($p < 0.001$ in all comparisons).

4. Discussion

A general challenge implicit in our design was that the US has not, as we write, yet implemented large-scale randomized testing for COVID-19. Consequently, detected cases involve over-representation of infected people who presented with morbid symptoms. Furthermore, accurate tracking of prevalence and mortality in the US has been impeded by decentralized administration and politicized conflict [16]. Epidemiologists universally acknowledge that undetected cases with lower morbidity outnumber detected cases [17]. The implication is that the evidence-based forecasting in which we asked our subjects to engage was not directly of the disease itself, but rather of the evolution of the processes used by public health officials to arrive at announced statistics and projections. It is open to question to what extent people behaviorally manage their health risks by responding to expert forecasts, and to what extent they choose behavior on the basis of their own idiosyncratic representations of diseases.

Coupled with these issues were significant changes in epidemiological understanding of the virus over the time period of our study, which presumably also influenced people's risk mitigation efforts. These changes in epidemiological understanding can be summarized as follows: estimations of the frequency of fomite transmission declined; estimations of the frequency of aerosol transmission increased; estimations of the efficacy of widespread mask use against prevalence, morbidity, and mortality increased; estimations of the weight of behavioral variables, independent of public-health policy choices, increased; and estimation of the extent of path-dependence in transmission geography due to "super-spreading" events increased. While our study does not speak directly to this greater epidemiological understanding of the virus, the fact that we constructed a pseudo panel of participants means that we can track the evolution of beliefs about COVID-19 prevalence and mortality over time. This will allow us to determine whether beliefs became more or less biased, and whether the confidence with which these beliefs were held varied, as more information about the virus became available. We will proceed with this line of investigation in subsequent analyses.

Figures 4 and 5, together with the complementary figures in Appendices B and C, show that forecasting COVID-19 infections is fraught with difficulty, certainly in comparison to deaths. Figure 4 shows that the correct answer in frame 0 and frame 1 about the level of infections on June 30, 2020 fell into the last bin of the event space, despite the fact that frame 0 represented the forecast that the IHME and we considered most likely. By contrast, Figure 5 shows that the correct answer fell into the "middle" of the event space in every frame. This difference in the accuracy of forecasting infections and deaths is not particularly surprising: while there have been very large swings in daily infections, deaths are less variable and follow infections with a predictable lag. This is arguably one reason why our subjects' beliefs appear to be more closely calibrated to deaths than infections.

The potential implications of our research for educational interventions about COVID-19 are clear. While there is no single, well-confirmed consensus theory of health behavior or other-regarding behavior that can ground educational efforts [18], beliefs will play an important role in explaining health behavior, regardless of the specific approach adopted. Beliefs about risk to oneself and to others are fundamental factors in understanding behavior, and are potential levers, therefore, for educational intervention. Meta-analyses show that risk perception has a significant influence on behavior [19]. Moreover, the extent to which beliefs influence behaviors, and which beliefs are amenable to educational influence, depends in part on the confidence individuals have in their attitudes. There is also solid evidence that there is heterogeneity across groups in beliefs about health risks [20].

Our study provides rich data about beliefs and related factors that the literature suggests are necessary to ground educational interventions. Because we elicit incentivized measures of beliefs, as well as the spread of confidence in various COVID-19 outcomes, we have fine-grained detail that is seldom available in educational interventions. Individuals who have very focused beliefs, and discount alternative outcomes strongly, will respond to information differently than individuals who give more credibility to alternative degrees of COVID-19 risk. In Bayesian statistical terms, those with more diffuse priors should respond more strongly to new information than those with tighter priors. We will also be able to investigate how our participant's beliefs about COVID-19 vary according to demographic characteristics, and incentivized elicitations of risk attitudes and time discounting, which we also included in the study. These additional variables could allow one to target public health and educational interventions to particular groups on the basis of their beliefs, and the extent to which they are more or less receptive to new information about risks posed by the virus and attendant mitigation measures.

5. Conclusion

We conducted an incentivized, experimental study on people's beliefs about COVID-19 prevalence and mortality with six temporally-spaced waves between May and November, 2020. Our experimental design allows us to draw direct, simple inferences about whether those beliefs differ from leading epidemiological models of infections and deaths due to COVID-19. We find that people's beliefs about both infections and deaths differ markedly from epidemiological models. Our study has implications for the dissemination of scientific information, and could be used to tailor public health and educational interventions to people most receptive to risk mitigation efforts.

References

1. Galton F. Vox Populi. Nature. 1907;75(1949):450-1.
2. Juster FT. Consumer Buying Intentions and Purchase Probability: An Experiment in Survey Design. J Am Stat Assoc. 1966;61(315):658-96.
3. Delavande A, Giné X, McKenzie D. Measuring Subjective Expectations in Developing Countries: A Critical Review and New Evidence. Journal of Development Economics. 2011;94:151-63.
4. Manski CF. Measuring Expectations. Econometrica. 2004;72(5):1329-76.
5. Harrison GW. Hypothetical Surveys or Incentivized Scoring Rules for Eliciting Subjective Belief Distributions? 2019. Working Paper 2014-05, Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University
6. Savage LJ. The Foundations of Statistics (Second Edition). New York, NY: Dover Publications; 1972.
7. Brier GW. Verification of Forecasts Expressed in terms of Probability. Monthly Weather Review. 1950;78(1):1-3.
8. Savage LJ. Elicitation of Personal Probabilities and Expectations. J Am Stat Assoc. 1971;66:783-801.
9. Matheson JE, Winkler RL. Scoring Rules for Continuous Probability Distributions. Management Science. 1976;22:1087-96.
10. Harrison GW, Martínez-Correa J, Swarthout JT, Ulm ER. Scoring Rules for Subjective Probability Distributions. Journal of Economic Behavior & Organization. 2017;134:430-48.
11. Harrison GW, Phillips RD. Subjective Beliefs and the Statistical Forecasts of Financial Risks: The Chief Risk Officer Project. In: Andersen TJ, editor. Contemporary Challenges in Risk Management. New York, NY: Palgrave Macmillan; 2014.
12. Di Girolamo A, Harrison GW, Lau MI, Swarthout JT. Subjective Belief Distributions and the Characterization of Economic Literacy. Journal of Behavioral and Experimental Economics. 2015;59:1-12.
13. Harrison GW, Swarthout JT. Belief Distributions, Bayes Rule and Bayesian Overconfidence. 2020. Working Paper 2020-11, Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University
14. Avery C, Bossert W, Clark A, Ellison G, Ellison SF. An Economist's Guide to Epidemiology Models of Infectious Disease. Journal of Economic Perspectives. 2020;34(4):79-104.
15. Epps TW, Singleton KJ. An Omnibus Test for the Two-sample Problem using the Empirical Characteristic Function. Journal of Statistical Computation and Simulation. 1986;26:177-203.
16. Lin TPH, Wan KH, Huang SS, Jonas JB, Hui DSC, Lam DSC. Death Tolls of COVID-19: Where come the Fallacies and Ways to make them More Accurate. Global Public Health. 2020;15(10):1582-7.
17. Benatia D, Godefroy R, Lewis J. Estimating COVID-19 Prevalence in the United States: A Sample Selection Model Approach. The Lancet (preprint). 2020:DOI: https://dx.doi.org/10.2139/ssrn.3578760.
18. Glanz K, Bishop DB. The Role of Behavioral Science Theory in Development and Implementation of Public Health Interventions. Annual Review of Public Health. 2010;31:399-418.

19. Sheeran P, Harris PR, Epton T. Does Heightening Risk Appraisals Change People's Intentions and Behavior? A Meta-analysis of Experimental Studies. Psychological Bulletin. 2014;140(2):511-43.
20. Lunn PD, Belton CA, Lavin C, McGowan FP, Timmons S, Robertson DA. Using Behavioral Science to Help Fight the Coronavirus. Journal of Behavioral Public Administration. 2020;3:DOI: https://doi.org/10.30636/jbpa.31.147.

Appendix A
[Online Working Paper]

Our method to select pairs of anchors required us to focus on prospective outcomes for COVID-19 statistics at various future time points, on which subjects could then place bets. We aimed to base these anchors on credible epidemiological models. This presented several challenges. First, when COVID-19 was declared a pandemic, epidemiologists still had relatively little knowledge of its transmission vector, but this knowledge improved rapidly and steadily over the course of our study frame [1]. This improvement of knowledge resulted in changes in the specification structure of models as our study unfolded, and the addition of new models that were made available between waves of our study. Second, no single model supported forecasts of all of the outcomes on which we asked subjects to report beliefs. In the face of these challenges, and others on which we elaborated in the Discussion section, we adopted the following methodology for selecting anchors for COVID-19 prevalence and mortality in the US population.[5]

To begin, for each of our six waves we selected a *baseline distribution* (BD0) using forecasts from the IHME. We selected this model because among those that were available from the beginning of our study, it uniquely provided specific projections of both case and death numbers for every future date through our planned time course. However, we did not limit anchors to the baseline distribution for two reasons. First, we sought to reduce the likelihood that many subjects might place all of their tokens in one extreme bin or the other, thus failing to provide us with much information about the distribution of their beliefs. Second, we aimed to avoid being limited to broad bin ranges that would fail to provide subjects with opportunities to report relatively precise beliefs if they indeed held such beliefs. Finally, we did not want to end up with uninformative responses, which could occur with very wide bins if subjects bet all of their tokens on the same bin.

We therefore constructed three additional bin anchors (BA1 - BA3) that shifted forecasting anchors relative to the baseline. To connect these to expert observation and modeling, we drew information from additional epidemiological models. The data journalism website

[5] We also asked subjects to forecast prevalence and mortality rates among Americans aged 65 years and older, in light of the crucial role of their far higher mortality in driving policy responses. The construction of anchors for this part of our experiment involved special problems due to progressive decline in available data quality over the course of our study. This part of the study will be discussed elsewhere.

FiveThirtyEight consolidates models produced by leading public health research institutes. The number of these reported models varied during the course of our study between 6 on Wave 1 and 14 by Wave 6. We used the mortality forecasts of these additional models to directly constrain construction of bin anchors BA1 - BA3 for each wave. As these models, unlike the IHME model, do not forecast infections, in anchoring bounds for infections we imposed the case fatality rate (CFR) that prevailed at the time of the wave according to the CDC. We then assumed that this would converge linearly over time to the CFR of the IHME model for December 1, 2020, but for a different number of deaths, and, hence, cases implied by the models on FiveThirtyEight.

We established anchors for BA1 in each wave by replacing the mortality anchor for BD0 by the bottom of the forecast range for the most "optimistic" model in the FiveThirtyEight suite as of the wave in question, where "optimistic" means the model that forecast the lowest number of deaths The upper anchor was then adjusted so that the probability density function (PDF) would replicate the baseline distribution BD0 as closely as possible, subject to the constraint imposed by the assumption made above about the CFR. To avoid suggesting implausibly over-precise estimates to subjects, such as 50,123 deaths, all anchors were converted to integers rounded to the nearest multiple of 10. We constructed anchors for BA2 by replacing the upper anchor of BD0 by the upper end of the most "pessimistic" model in the FiveThirtyEight suite, and adjusting the bottom anchor by analogous restrictions as for the BA1 construction above. Finally, we constructed anchors for BA3 by setting the upper anchor to the top of the error range of the implied BA2 model for $p = 0.05$, then shifting up the bottom anchor by again maintaining the PDF constrained by the assumed CFR.

Thus, the ranges presented to study subjects were based on one set of bin anchors (BD0) representing the IHME forecast, one set of anchors (BA1) shifted in an "optimistic" direction that remained within the range of expert forecasts, and two sets of bin anchors (BA2 and BA3) shifted in a "pessimistic" direction, but also within the bounds of epidemiological modeling. The motivation for this asymmetry between optimistic and pessimistic representations was to reduce the risk that acceleration of rates of infection or mortality might cause reported statistics to "catch up" to lower bounds of the bin anchors during the 2-day intervals between programming and launching each wave, which would have the effect of reducing subjects' reasonable response ranges.

With this set of bin anchors for prevalence and mortality statistics over one-month and December 1, 2020 timeframes we used our beta distribution algorithm to partition the event spaces and define the set of bins for the task. Four sets of bin anchors produced four sets of bins per belief question, which we refer to as the *frames* for that question: BD0 defined the anchors for frame 0, and BA1, BA2, and BA3 defined the anchors for frames 1, 2, and 3, respectively.

Additional References

1. Xu B, Kraemer MUG, Open C-DCG. Open access epidemiological data from the COVID-19 outbreak. Lancet Infectious Diseases. 2020;20(5):534.

Appendix B
[Online Working Paper]

This appendix shows the distributions of subject reports about COVID-19 prevalence from waves 2, 3, and 4 of our study. The main text provides a general discussion of the results from these waves, and how they follow a similar pattern to that which we observed in wave 1.

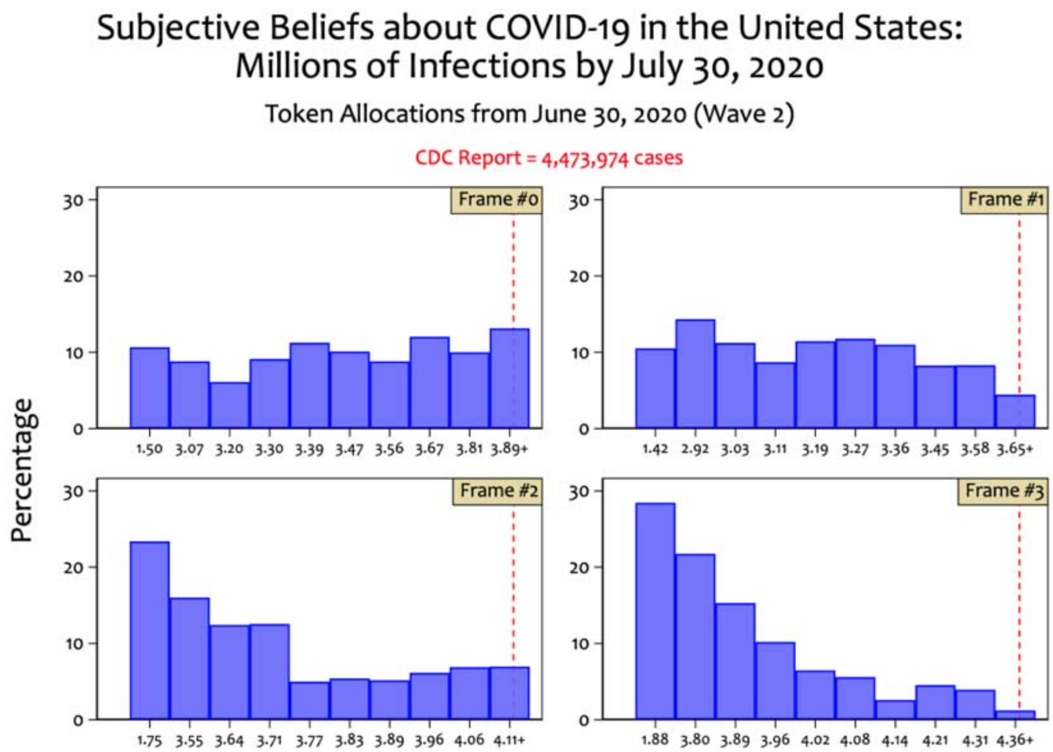

Figure B1: Beliefs about COVID-19 Infections in the US by July 30, 2020

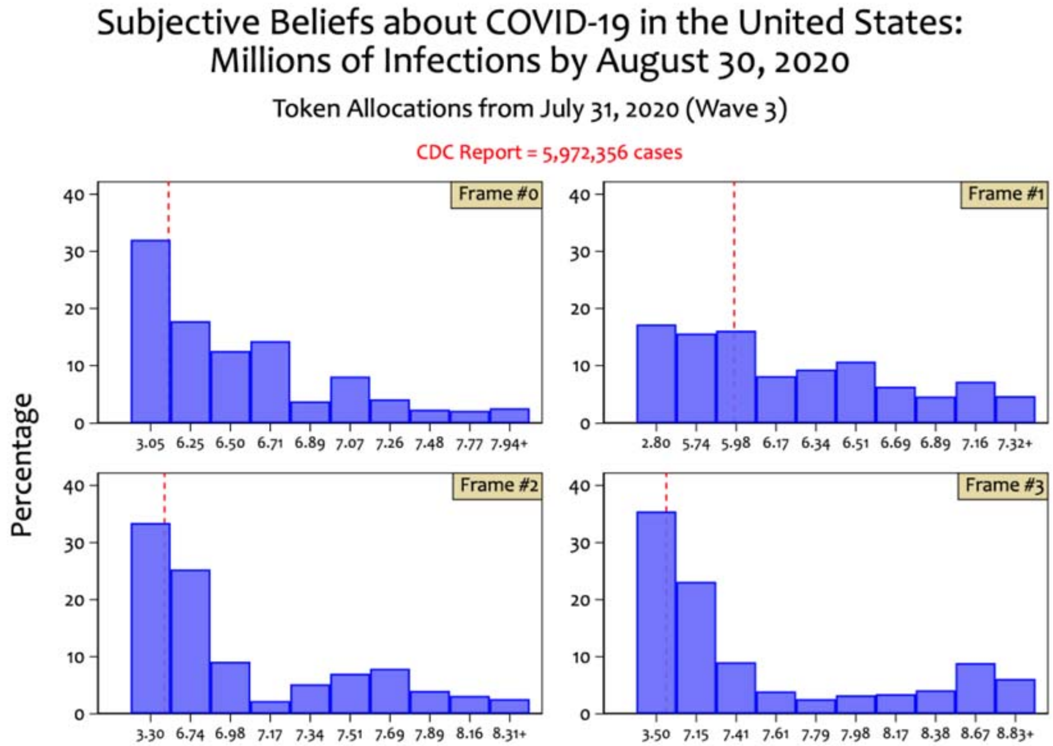

Figure B2: Beliefs about COVID-19 Infections in the US by August 30, 2020

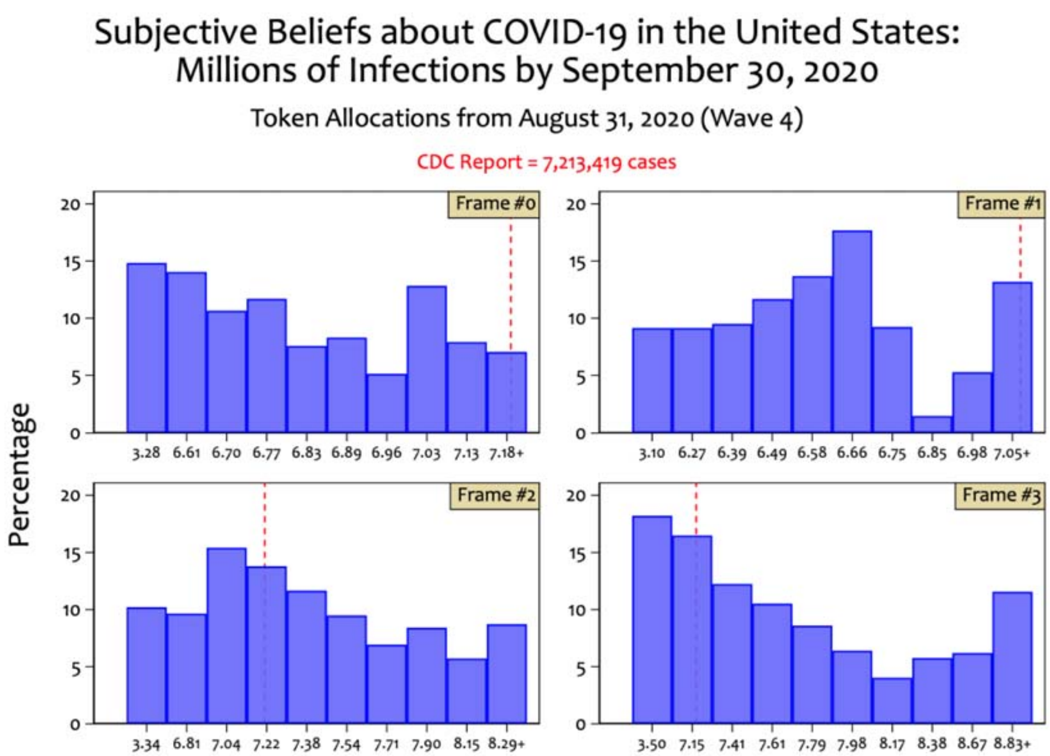

Figure B3: Beliefs about COVID-19 Infections in the US by September 30, 2020

Appendix C
[Online Working Paper]

This appendix shows the distributions of subject reports about COVID-19 mortality from waves 2, 3, and 4 of our study. The main text provides a general discussion of the results from these waves, and how they follow a similar pattern to that which we observed in wave 1.

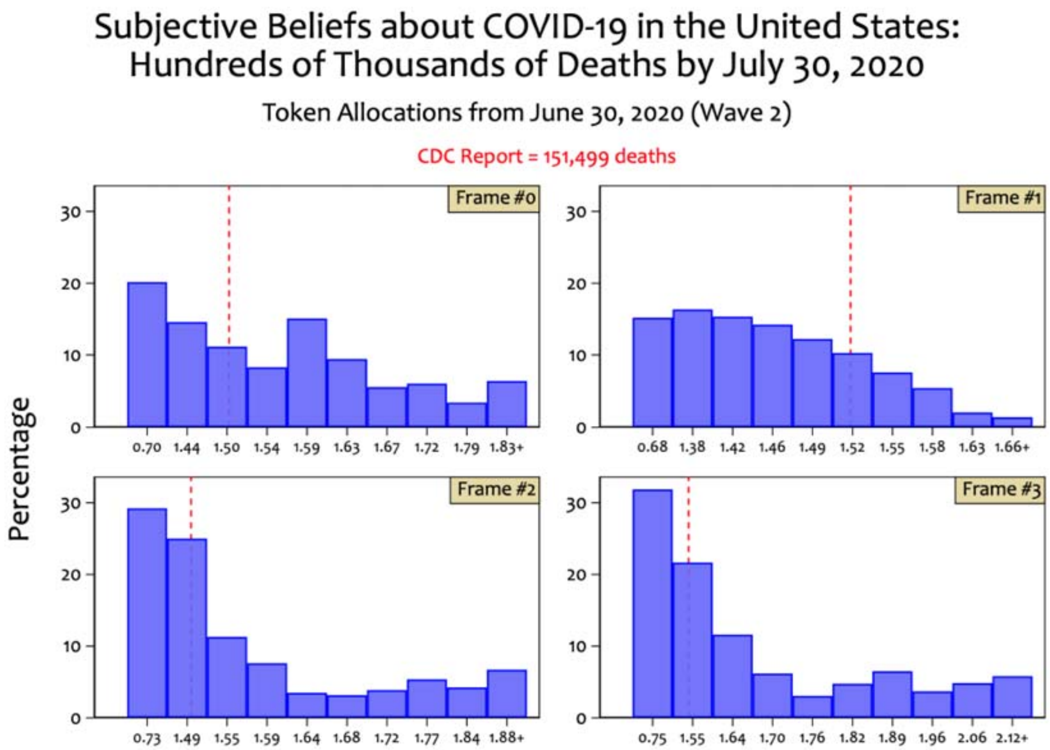

Figure C1: Beliefs about COVID-19 Deaths in the US by July 30, 2020

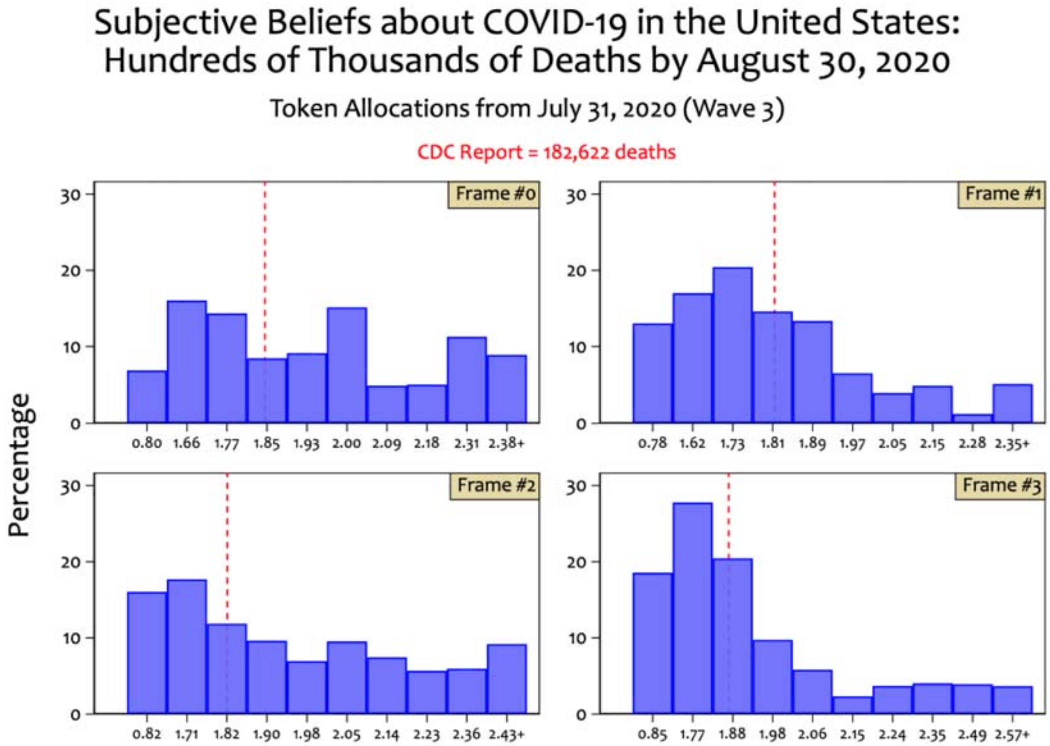

Figure C2: Beliefs about COVID-19 Deaths in the US by August 30, 2020

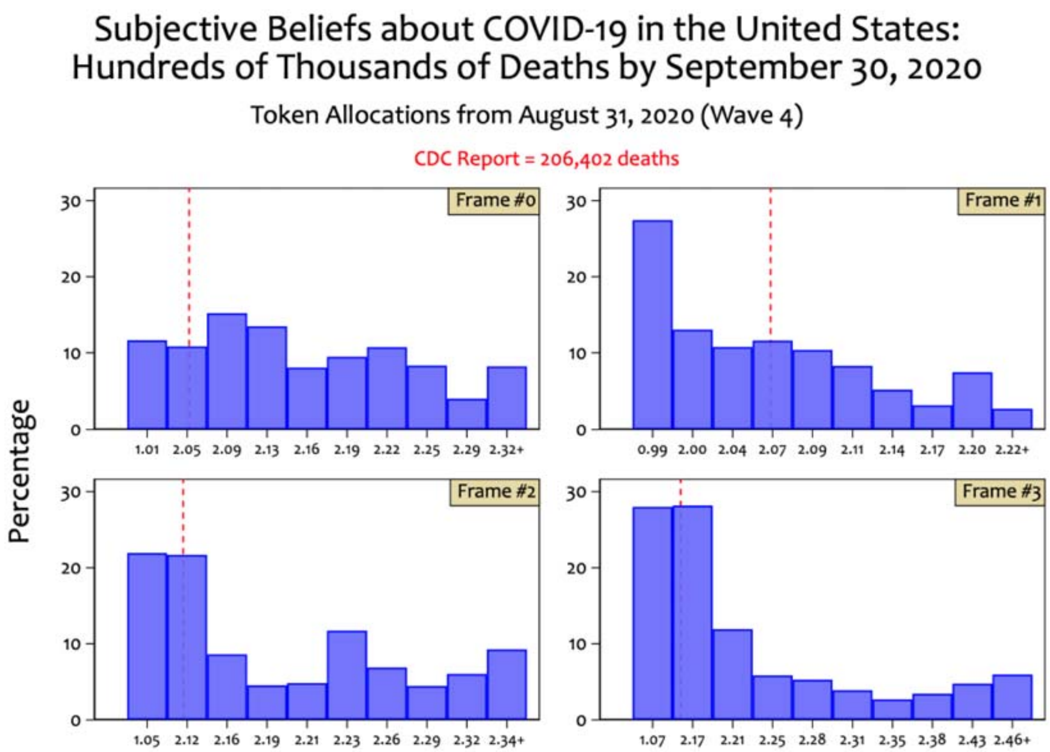

Figure C3: Beliefs about COVID-19 Deaths in the US by September 30, 2020