communications medicine

A Nature Portfolio journal

Article

https://doi.org/10.1038/s43856-025-00751-8

COVID-19 death toll predictions show that triggering counterfactual thinking deteriorates judgmental performance

Check for updates

Matthias Seifert & Jeeva Somasundaram

Abstract

Background Effective communication during a health crisis is critical as it directly influences psychological and behavioral responses that will shape the further progression of the crisis. Past research has suggested that one type of cognitive mechanism that is likely to be affected by the framing of public health messages relates to counterfactual thinking.
Methods Based on 6731 incentivized daily forecasts collected over 377 days (from April 2020–May 2021), we investigate the role of triggering counterfactual thinking when interpreting public information regarding the daily US death toll from COVID-19.
Results Here we show that individuals who engaged in thinking about "interventions that could have led to an alternative evolution of the death toll" prior to making forecasts exhibit greater judgmental bias in their predictions compared to the control group. Specifically, subjects in the treatment group tend to generate upward counterfactuals and underestimate the death toll, potentially due to anchoring on more favorable scenarios and insensitivity to trend changes. Interestingly, this behavior is also observed among individuals who had recovered from COVID-19 (or someone in their close social circle).
Conclusions Our findings underscore the importance of using debiasing strategies and neutral communication during health crises to mitigate the generation of upward counterfactuals, thus reducing the likelihood of systematic misperceptions and flawed decision-making.

Plain language summary

How information is provided to the public about health can impact the interpretation by the public and how they behave. Counterfactual thinking occurs when people construct mental representations of alternatives to the past. Over the period of one year, we collected daily predictions from a sample of 6731 US citizens to study how triggering counterfactual thinking influences judgmental predictions regarding the US death toll caused by COVID-19. We find that counterfactual thinking substantially worsens predictive accuracy and propose strategies for improving the effectiveness of crisis communication to avoid counterfactual thinking occurring.

Effective communication during a health crisis is critical as it directly influences behavioral responses that shape the progression of the crisis[1]. A recent report commissioned by the European Parliament[2] concluded that ineffective crisis communication strategies in some countries during the COVID-19 pandemic led to lower acceptance of health and safety measures by public citizens. Moreover, previous studies have shown that fear-based messaging of COVID-19 related information resulted in heightened states of anxiety and psychological stress, yet without enhancing consumers' intentions to adopt mitigation measures[3]. Finally, presenting information related to COVID-19 in terms of absolute versus relative changes[4] as well as logarithmic versus linear graphs[5] systematically affect judgmental accuracy and subsequent choice behavior.

Examples like these show that it is important to understand how the framing of public health messages impacts judgment. One cognitive mechanism that is likely to be affected by the framing of messages relates to counterfactual thinking, defined as the process of constructing mental representations of alternatives to the past[6]. Individuals frequently speculate how things could have been if they had turned out differently in an attempt to make sense of past events and prepare themselves for the future[7]. Importantly, counterfactual thinking produces psychological consequences that can be positive or negative. For instance, Bertolotti and Catellani (2023)[8] found that counterfactual thinking served as a prebunking strategy to the spread of misinformation on COVID-19. Similarly, García Ferrés & DePalma (2023)[9] reported that counterfactual thinking may attenuate polarized attitudes with respect to COVID-19 prevention behaviors. Finally, Ahn et al. [10] suggested that repeated exposure to "what-if plans" during COVID-19 can lead to an increase in reported social distancing behavior over time.

In the present paper, we study how counterfactual thinking influences information processing and predictive performance when estimating the daily US death toll caused by COVID-19. Specifically, media updates on the daily COVID-19 death toll—which served as the primary source of

IE Business School, IE University, Madrid, Spain. e-mail: matthias.seifert@ie.edu

information about the state of the crisis—not only shaped individuals' risk perceptions but also influenced subsequent decisions regarding public transportation use, vaccination, social distancing, and other related behaviors[11]. Understanding the effect of counterfactual thinking on judgmental performance is key because the framing of public health messages released during the pandemic may have unintentionally induced various degrees of counterfactual thinking among recipients[12,13]. Furthermore, the development of effective communication strategies during future health crises requires an understanding of the cognitive mechanisms at work that drive behavioral responses.

While some studies suggest that counterfactual thinking can improve information processing[14], extant research has yet to examine the implications of counterfactual thinking in judgmental prediction tasks and rarely studied it outside controlled laboratory environments. Moreover, it remains unclear how counterfactual thinking interacts with behavioral biases that forecasters commonly fall prey to in prediction tasks[15]. For example, recent studies comparing the judgmental performance of experts and consumers in health crises found a general tendency to underestimate infection fatality rates. However, this underestimation was less pronounced among experts compared to the general public[16]. In addition, Harvey and Reimers (2013)[17] found that forecasters are likely to dampen trends in their judgmental predictions when the presented series is steeper than what is thought to be representative for the underlying data environment.

The COVID-19 pandemic represents a naturalistic, emotionally charged, high impact setting in which the effectiveness of health policy communication was critical due to the substantial personal risks involved. Given the unprecedented nature and severe personal consequences of the health crisis, we expected forecasters engaging in counterfactual thinking to generate upward counterfactuals[7,18] that consider an alternative, more positive evolution of the death toll. Moreover, these counterfactuals are likely to focus on uncontrollable factors tied to *others'* actions or inactions rather than one's own[19]. While previous research has explored the impact of counterfactual thinking on various motivational, intentional, and behavioral outcomes[19], its influence on judgmental performance remains underexplored. One exception is Hoch's (1985)[20] study, which found that generating predictions about personal events, where unrealistic optimism was expected, led to improved predictive accuracy when 'con reasons' were considered. However, this study only examined events with *positive, controllable* outcomes (e.g., job search efforts). The effect of counterfactual thinking on judgmental performance in situations involving *negative, uncontrollable* outcomes—such as in the case of health crises—remains unexplored.

Counterfactual thinking in prediction tasks requires forecasters not only to extrapolate the target event from factual data, but also to engage in mental simulations that consider hypothetical trajectories of the historical data. In the context of COVID-19 pandemic, we expect both these processes to be influenced by a trend damping bias[17]. Given that forecasters are likely to generate upward counterfactuals in the context of COVID-19, they may anchor their judgment on the positive impact of the hypothetical alternative, resulting in greater judgmental underestimation of the actual death toll compared to those who do not engage in counterfactual thinking. This rationale also applies to the behavioral pattern observed in Hoch (1985)[20] for prediction tasks with positive personal outcomes. In a similar vein, we expect people who contracted COVID-19 in the past but escaped the crisis unscathed to anchor judgments on their own personal experience and therefore underestimate the true risk of dying akin to the base-rate fallacy in probability estimation tasks. However, despite the plausibility of these arguments, it is possible that upward counterfactuals may result in overestimation of the death toll due to a contrast effect when comparing the current situation with the counterfactual reality[7,21].

Drawing on an online field study involving US citizens predicting the US daily death toll caused by COVID-19 over a period of 377 days, we test these competing predictions. Specifically, we study how individuals thinking about "interventions that could have led to an alternative evolution of the death toll" prior to generating predictions perform compared to forecasters in a control group. We find that invoking counterfactual thinking leads to substantially larger judgmental bias and underestimation relative to the control group.

Methods

Task design

We recruited subjects from Amazon Mechanical Turk between April 24, 2020 and May 5, 2021. Each day, we recorded responses from approximately 18 individuals ($N = 6731$), who were randomly assigned to one of two experimental conditions. Subjects began by reading through a detailed instruction screen, on which they were introduced to the forecasting task and received verbal as well as graphical stimuli of the key measures to be elicited as part of the experiment. In both treatments, subjects provided judgmental forecasts regarding tomorrow's US death toll caused by COVID-19. To facilitate this elicitation, we included an interactive graph on the top half of the task screen, which displayed the daily confirmed cases of deaths caused by COVID-19 in the US since the day of the first officially recorded fatality.

We elicited single point estimates about the most likely death toll for the next day as well as judgments regarding the lower and upper boundaries of a 90% confidence interval. In addition, we manipulated counterfactual reasoning (as in Grossman et al. 2023)[22] in the context of COVID-19 by asking subjects to think and write about "interventions that could have led to an alternative evolution of the death toll" prior to generating predictions. Importantly, treatment and control conditions in our experiment only differed with regards to whether subjects described the counterfactual event right after observing the time series graph and before completing the actual forecasting task (treatment), or at the end of the experiment (control). Subjects also provided an assessment of whether they believed that the general trend of the data series had changed ("Since last Monday, has the trend of the daily COVID-19 death toll in the US changed?") and answered several follow questions in relation to COVID-19 (e.g. "Have you or anyone in your close social circle of family and friends been infected with COVID-19?", "Have you been vaccinated?"), which were added to the survey four months after the start of the data collection. Finally, subjects provided general demographic information.

Subjects received a fixed payment of $0,40 USD in return for their participation and a lottery-based performance bonus that rewarded participants with the lowest forecast error an additional $1 USD.

Data collection

We limited participation to US subjects with a hit approval rate of above 90%. Although the study was based on a between-subject design, the length and size of the data collection resulted in some subjects participating more than once. Specifically, 876 subjects provided multiple forecasts.

The median completion time was 4,3 minutes. Our sample comprised $N = 6731$ forecasts from 3261 unique subjects (female: 2004, mean age: 39.37). There were 3361 forecasts in the control group, and 3370 forecasts in the treatment group. Ninety-five percent of the subjects provided fewer than 5 forecasts across the duration of the study. Mean age and gender proportion are similar in treatment and control conditions (t-test: age $p = 0.86$, and gender $p = 0.65$). All t-tests reported in this study are two-tailed.

Protocol

Subjects started by providing their informed consent. The experiment was coded in Qualtrics and approved by the IE University IRB. In the task we embedded a visualization of the daily death toll provided by Our World in Data, which was customized to display data for the United States only and can be accessed at https://ourworldindata.org/explorers/coronavirus-data-explorer?zoomToSelection=true&hideControls=true&Metric=Confirmed+deaths&Interval=New+per+day&Relative+to+Population=false&Align+outbreaks=false&country=~USA[23]. The graph shows the daily new confirmed COVID-19 deaths since the day of the first officially recorded fatality in the US (February 29, 2020) and sources data from the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. The full data set can be accessed at https://github.com/CSSEGISandData/COVID-19[24]. Since not all confirmed deaths may be known yet at any time period t + 1 (i.e., numbers could be adjusted retrospectively due to differences in the timing when hospitals and public authorities reported registered cases), our performance measures only consider the number of deaths known (and displayed) on the specific day the forecasting judgment was elicited.

The actual questions used in the survey, the data as well as all codes are available in OSF with the identifier https://doi.org/10.17605/OSF.IO/XPRWN (https://doi.org/10.17605/OSF.IO/XPRWN).

Statistics and reproducibility

To study the effects of our treatment condition on judgmental forecasting performance, we rely on a combination of non-parametric comparisons and regression analysis. Specifically, when the DV was based on the most-likely forecasts, to account for non-normality of residuals, we estimated median regression models in addition to OLS models, whereas whenever our DV was binary (i.e., such as in the case of the elicited confidence intervals), we relied on logit regressions. We report details of our main analyses directly, whereas additional analyses are fully described in the supplementary information and referred to in the main text as Supplementary Fig. and supplementary table.

We begin with a descriptive analysis of the heterogeneity of subjects' responses across control and treatment conditions and compare differences in the most-likely forecasts using non-parametric tests. Then, in our regression analysis, more rigorously, we examine the effect of counterfactual thinking on three key performance measures commonly used in forecasting theory and practice[15]. In the regressions, our results can be reproduced by controlling for subject-level variation using control variables and for repeated observations from the same subject (or non-independence of observations) by clustering the standard errors at the subject level as commonly done in panel data analyses[25,26]. Note that due to random assignment to the experimental conditions, one subject may have provided forecasts in both conditions. Therefore, we also re-analyze our data by only considering the first response of subjects. In addition, we also explore the rationale behind our findings by analyzing forecasts based on COVID-19 infection status and trend perceptions.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Results

Most-likely forecasts summary statistics

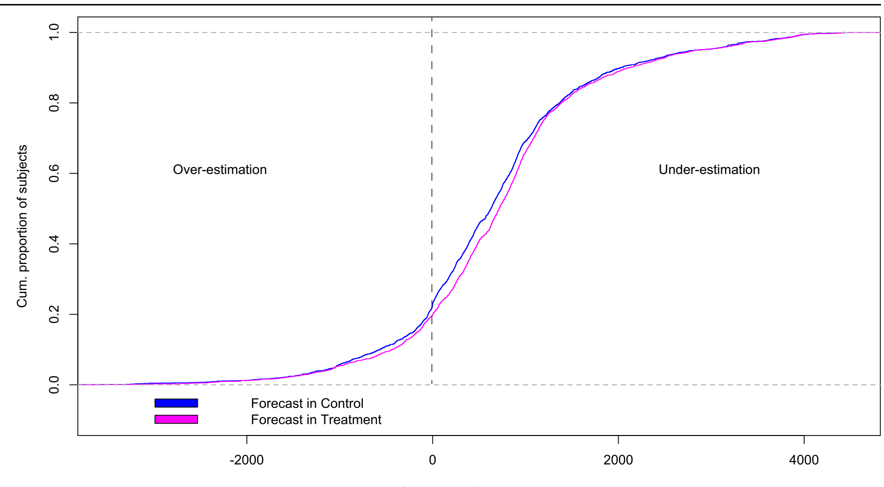
We analyze the average daily number of COVID-19 deaths in the United States during the specified period and the average most-likely forecasts in both the treatment and control groups. Most-likely forecasts that fell short of or exceeded the actual death count for the day by 200% were dropped from the analysis. This left us with a sample of 5920 forecasts entering our analyses (Control group: 2964 and Treatment group: 2928). Dropping these outliers did not affect age and gender proportion of participants between treatment and control conditions (t-test: age $p = 0.29$ and gender $p = 0.63$). Our results remain consistent even when considering more extreme outlier exclusions rules that include eliminating forecasts that are 300% or 400% higher than actual death toll (see supplementary information: sensitivity analysis, Supplementary Tables S17(a) and (b) to Table S22).

When conducting non-parametric Mann-Whitney U tests, most-likely forecasts in the treatment group (Mean forecast: 662.78) are lower than in the control group (Mean forecast: 735.15, Cohen's d: Mean = 0.07, 95% CI = [0.02,0.12], $p < 0.001$). This is illustrated in Figure S1 in the supplementary information. Moreover, forecasts in both control and treatment groups are lower than the actual death count (Mean of actual death = 1390, Control vs actual death - Cohen's d: Mean =0.63, 95% CI = [0.59,0.67], $p < 0.001$, treatment vs actual death toll - Cohen's d: Mean = 0.69, 95% CI = [0.65,0.73], $p < 0.001$). In other words, participants systematically underestimated the number of deaths, particularly in the treatment group, where they were prompted to consider a counterfactual event. We find similar results even when dropping repeated observations and considering only subjects' first forecasts (Mean: control group = 689.32, treatment group =553.83, Mann Whitney U test: $p = 0.006$, Cohen's d: Mean 0.15, 95% CI: [0.07,0.23]).

Heterogeneity in most-likely forecasts

In Fig. 1, we plot the differences between actual deaths and most-likely forecasting judgments for control and treatment conditions. A negative difference (i.e. when subtracting most-likely forecasts from actual deaths) implies that forecasts overestimate the actual death toll (i.e., the most-likely forecast is higher than the actual death toll). In contrast, a positive difference implies that forecasts underestimate the actual number of deaths (i.e., the most-likely forecast is less than the actual death toll). Figure 1 reveals overestimation for 21% of forecasts, while the remaining 79% underestimate deaths. Additionally, the cumulative distribution function (CDF) of the treatment condition consistently lies below the one of the control condition – as confirmed by the Kolmogorov-Smirnov (KS) test ($p < 0.001$) – suggesting that forecasts in the treatment condition consistently underestimate the true death toll in comparison to the control condition.

Fig. 1 | CDF showing the extent of under- (or over-) estimation of deaths.



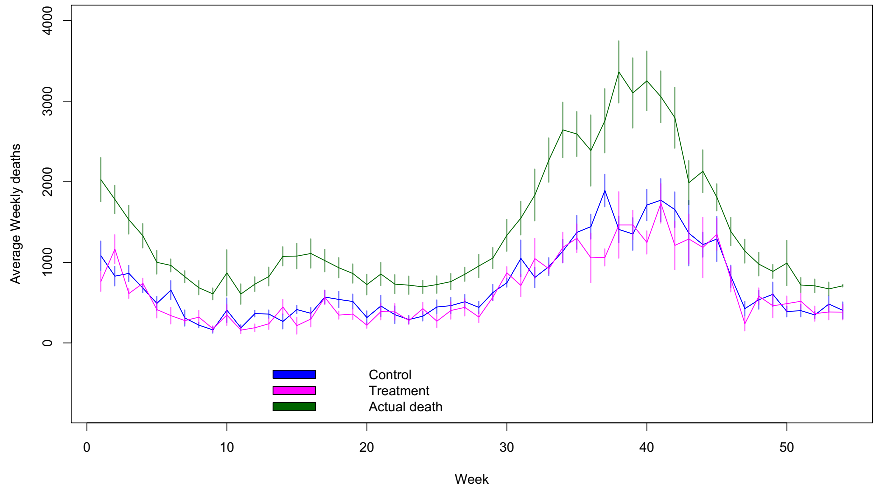


Fig. 2 | Actual death toll and most-likely forecasts in treatment and control conditions (standard error bars are displayed on the graph).

Figure 1 legend: The CDF is computed based on $n = 2964$ most-likely forecasts in the control group and $n = 2928$ most-likely forecasts in treatment group.

Time-series and regression results

We next examine how judgments change over time in response to the evolving daily death figures. Notably, in Fig. 2, we observe that forecasts in treatment and control groups consistently underestimate daily deaths. This is particularly pronounced during the peaks. Furthermore, forecasts in the treatment group consistently fall below those of the control group, indicating a greater degree of underestimation.

We analyze the extent of underestimation statistically using regression analysis. Since residual plots are not normal when running OLS regression (Shapiro test, $p < 0.001$, see Supplementary Fig. S2 for the residual plot), we estimated a quantile regression with tau=0.5 (i.e., which can be also referred to as a median regression). Quantile models are more robust to violations of normality in the data, less sensitive to outliers, and do not make assumptions about the distribution of the parameters[27]. We cluster standard errors in the median regression to account for the correlation between repeated forecasts of the same subjects and control for subjects' gender and age. The regression results reported in Table 1: Column 1 show that the coefficient for the treatment variable is positive and significant, suggesting that in the treatment condition, people underestimate the actual death toll by an additional 110 median deaths compared to the control condition ($p = 0.0002$).

Even without controlling for subject level characteristics (see Supplementary Table S1: Column 1), we find identical results with median regression: Forecasters in the control group underestimate the death on average by 538 ($p < 0.001$) and forecasters in the treatment condition underestimate the death toll on average by an additional 89 compared to the control condition ($p = 0.002$). We find qualitatively similar results when conducting OLS regression with clustered standard errors (see Supplementary Table S2: Column 1).

Figure 2 legend: The average weekly death rate (mean and standard error) was computed based on average of actual deaths that occurred during the 7 days of the particular week. While average weekly forecasts (for $n = 54$ weeks) for the control and treatment group were computed based on average of most-likely forecasts that occurred during the particular week of the study (most-likely forecasts over all weeks: n = 2964 in the control group and $n = 2928$ in treatment group).

We further examine deviations from actual deaths by segmenting the time series using Bai-Perron (1998)[28] multiple breakpoint tests. We identify three breakpoints that occurred on the 56th day (week 8), 221st day (week 32), and 309th day (week 45), which represented turning points for the underlying trend in the time series data. Notably, the peak on the 56th day appears to be an anomaly (see week 8 in Fig. 2), while the real peak in the number of deaths occurs between the 221st and 309th days. We therefore consider the interval between 221st and 309th days to be the peak period and the remaining days in the time-series to be the non-peak period. We run a median regression to compare forecast differences between treatment and control groups during peak and non-peak periods (see Supplementary Table S3) and observe more underestimation of 863 deaths in both groups during the peak period compared to non-peak periods. However, differences in underestimation between treatment and control groups during the peak period are not statistically meaningful ($p = 0.28$, see interaction treatment × period in Table S3). We find identical results with OLS regression (Supplementary Table S4).

Table 1 | Median regression on the effect of the treatment on forecasting performance

	Death - Forecast	sMAPE
Treatment	110.400*** (28.97) $p = 0.0002$	0.068*** (0.011) $p < 0.0001$
Female	17.400 (57.2) $p = 0.762$	−0.008 (0.025) $p = 0.751$
Age	5.200** (2.522) $p = 0.040$	0.004*** (0.001) $P = 0.0002$
Constant	260.200* (134.77) $p = 0.054$	0.518*** (0.060) $P < 0.0001$
Observations	5527	5527

*$p < 0.1$, **$p < 0.05$, ***$p < 0.01$; Standard errors (clustered at subject-level) are shown in parentheses, Reference category: Control, Male.

Forecast error

We next analyzed how much forecasting judgments deviated from actual deaths using alternative, commonly used forecasting performance measures. Specifically, we quantified forecast errors in terms of the symmetric mean absolute percentage error (sMAPE)[29] due to the non-stationary nature of the daily death time series, which varied greatly in magnitude across different segments of the series. Formally, sMAPE, is defined as follows:

$$sMAPE = \frac{1}{NT}\sum_{i=1,\ldots,N}\sum_{t=1,\ldots,T}\frac{|D_t - P_{it}|}{|D_{it}| + |P_{it}|}$$

where D_t is the actual death toll on day t and P_{it} refers to the death toll predicted by individual i on day t.

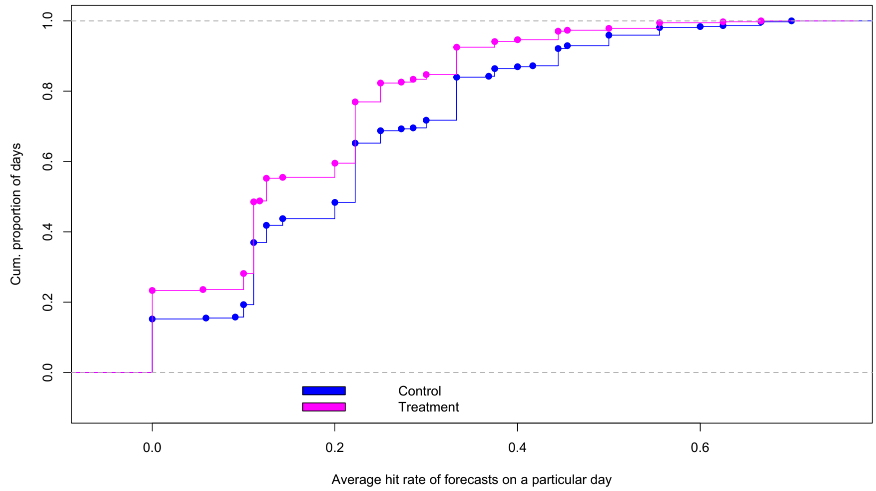


Fig. 3 | Proportion of forecasts (in treatment and control) whose CI captures the actual death on a particular day.

As the residual error plots are not normally distributed, we relied on median regression using sMAPE as the dependent variable and included a binary independent variable to indicate whether subjects had been assigned to the control or treatment conditions. Median regression results (controlling for age and gender) with clustered standard errors are reported in Table 1: Column 2. The coefficient of the binary treatment variable is significant and indicates that forecast errors are larger in the treatment group compared to the control group ($p < 0.001$). We find identical results without the controls (Supplementary Table S1: Column 2) and in a corresponding OLS regression (Supplementary Table S2: Column 2). Our results also remain the same when dropping repeated observations and focusing only on the first response of the subjects (see median and OLS regressions in Supplementary Table S15a, b).

Confidence intervals and hit rate

Subjects also provided judgments regarding the lower and upper boundaries of a 90% confidence interval (CI), such that the actual death toll would fall within the predicted interval 90% of the time. If both lower and upper boundaries were entered correctly, we considered it a valid entry. We constructed a binary hit rate measure to indicate whether the actual death toll fell into the elicited confidence intervals (defined as 1 when the death toll for a specific day lay within the CI and 0 otherwise).

Figure 3 illustrates the cumulative distribution function (CDF) of the average hit rate on a particular day. The hit rate of the treatment group lies to the left of the control group (KS test, $p < 0.001$). A Mann-Whitney U test shows a higher hit rate in the control group (21.3%) than in the treatment group (16.4%) (p < 0.001). Even when removing repeated observations, the control group (22.3%) is associated with a higher hit rate than the treatment group (15.9%) ($p < 0.001$). This suggests that the CIs provided by subjects in the control group capture actual deaths more accurately compared to the treatment group, indicating better predictive performance. Additionally, we estimated a logit regression model with the experimental treatment as independent variable and hit rate as the binary dependent variable. The logit regression results, controlling for age and gender, are reported in Table 2 (and without controls in Supplementary Table S6). The coefficient for the treatment variable confirms that the hit rate is lower for subjects in the treatment group compared to the control group subjects ($p < 0.001$). We also find similar results when estimating a linear probability model based on OLS regression (see Table S2: Column 3). Our results also remain the same when dropping repeated observations (see median and OLS regressions in Supplementary Table S16a, b.

Figure 3 legend: The CDF is based on average hit rate computed for a particular based on $n = 3348$ CIs in the control group and $n = 3362$ CIs in the treatment group.

Table 2 | Logit regression effect of experimental treatment on hit rate

	Hit rate (1/0)
Treatment	−0.332*** (0.062) $p < 0.0001$
Female	0.218** (0.106) $p = 0.040$
Age	−0.011** (0.005) $p = 0.019$
Constant	−0.767*** (0.280) $p = 0.007$
Observations	6312
Log likelihood	−3097.662
Akaike Inf. Crit.	6202.323

*$p < 0.1$, **$p < 0.05$, ***$p < 0.01$
Standard errors (clustered at subject-level) are shown in parentheses, Reference category: Control, Male.

Moreover, we examined the width of elicited confidence intervals between treatment and control groups using OLS regression and found no significant differences ($p = 0.28$) (see Supplementary Table S5). Although treatment and control groups exhibit similar levels of judgmental precision, predictions in the treatment group are systematically inferior to the control group.

Effect of COVID-19 infection status on forecasting

We further analyzed subjects' responses to questions related to whether they or someone in their immediate social circle had experienced a COVID-19 infection prior to the day of their participation. As the residual error plots for these models indicate a potential violation of OLS assumptions, we specify median regression models with clustered standard errors using a binary predictor variable to capture subjects' responses ('yes'/'no') in order to estimate the extent of underforecasting and forecast errors. The results are summarized in Supplementary Table S7. On average, respondents across the two groups who answered 'yes' to this question underestimated deaths by 424 on average compared to those who answered 'no' ($p < 0.001$). We also ran a median regression with an interaction term between treatment and COVID-19 infection status variables. However, the interaction term was not significantly different ($p = 0.350$), indicating that underestimation in the treatment group was not systematically higher than in the control group (see Supplementary Table S10). In addition to the effect on the extent of underforecasting, we observe that people who answered 'yes' had higher sMAPE and lower hit rate compared to people who answered 'no' (Supplementary Tables S8-S9). Therefore, our findings show that although all subjects were provided identical information, those who (or whose family members) had experienced COVID-19 and escaped the pandemic unscathed provided lower estimates and had higher forecasting error.

Trend judgment and predictive accuracy

For analyzing subjects' trend assessments (i.e. whether or not the trend of the death toll had changed during the week in which forecasters provided their judgments), we estimated a logit regression using a binary variable ('trend') that was associated with a value of 1 when subjects believed that the trend of the death toll had recently changed into the opposite direction and zero otherwise. When adding this predictor to the otherwise identical regression models and after excluding outliers, we find that the proportion of subjects who felt that the trend remained constant was larger in the treatment condition compared to the control group ($p = 0.01$) (see Supplementary Table S14: negative coefficient of the treatment variable in the logit regression indicates that subjects in the treatment condition were less likely to perceive changes in the direction of the death toll trend). Yet, trend judgments elicited in both conditions were associated with similar levels of judgmental confidence (t-test, $p = 0.11$). Moreover, those subjects who perceived the trend to be unchanged underforecasted the daily death toll to a greater extent and exhibited larger forecasting errors ($p < 0.001$) (median regression results with clustered standard errors in Supplementary Table S11, logit results in Supplementary Table S12, and OLS results in Supplementary Table S13). Therefore, counterfactual reasoning appears to desensitize laypeople's perceptions of changes in trend, ultimately resulting in larger forecast errors.

Sentiment analysis and character count

We finally analyzed subjects' verbal justifications in the treatment group, which they provided right before generating forecasts. First, we coded counterfactual statements as "upward" or "downward". Consistent with our predictions, 95% of the legible statements that the subject wrote were upward counterfactual statements, which may have anchored subjects on a more favorable evolution of the death toll that led them to underestimate the death toll.

In addition, we calculated character count, word count, and general sentiment of the text that they wrote. Subjects in the treatment group on average wrote 24 words and 136 characters, respectively. When estimating the sentiment of counterfactuals by counting the number of words that were associated with positive and negative sentiments in comparison to the "bing" dictionary, we obtained the following results: Subjects on average used 1.36 words (SD = 1.86) with net-negative sentiments, whereas the perceived importance of their counterfactual for the evolution of the pandemic was rated on average 2.78 (SD = 1.18) (on a 5 point scale). Finally, we ran median regression models to estimate the effects of sentiment and character count on our focal dependent variables (Supplementary Tables 21, 22). When controlling for gender and age, we find the slope coefficient for character count to be significant, showing a positive association with forecast errors (sMAPE) ($p = 0.031$) and leading to a lower hit rate ($p < 0.001$). However, sentiments generally did not have any effect on the extent of underestimation and forecasting errors.

Discussion

Our study shows that in forecasting environments with negative, uncontrollable outcomes, triggering counterfactual thinking when assessing crisis-related information can deteriorate consumers' subsequent judgmental performance. We found that individuals who thought and wrote about interventions that could have led to an alternative evolution of the COVID-19 crisis prior to observing publicly available information regarding the US death toll caused by the virus, exhibited inferior predictive performance that was associated with larger forecasting errors compared to the control group. Considering the potential mechanism, our results revealed that forecasters in the treatment group generated upward counterfactuals and consequently were less sensitive to trend changes compared to the control group. This could be potentially due to anchoring on the positive counterfactual scenario. Consistent with this view, we found forecasters who used more characters to describe their counterfactual reasoning in the treatment group (and consequently anchored more) underforecasted the death toll to a greater extent.

The present study contributes to the existing literature in the following ways. First, previous research has shown that external threats induced, for example, by the COVID-19 pandemic can have a severe impact on consumers' sense of security and therefore lead to adaptive responses in their decision-making behavior[30]. The present study adds to this research by investigating how such external shocks influence the effect of counterfactual reasoning on judgmental performance. While past studies have only considered the role of counterfactuals in contexts where the uncertain future was related to positive personal outcomes[20], our study focuses on domains with negative personal outcomes such as in the case of a health crisis. We find that when making predictions about the death toll caused by COVID-19, the consideration of upward counterfactuals, leads to systematically lower judgmental accuracy. This finding likely resulted from a judgmental bias when engaging in a mental simulation of a positive counterfactual scenario. Specifically, anchoring on this hypothetical positive scenario made subjects insensitive to changes in trends, ultimately leading them to underestimate the actual death toll. We find that counterfactual thinking can be problematic in these environments, because flaws in the simulated reality are likely to anchor judgmental extrapolations of the true progression of the pandemic. Similarly, we observed that individuals who reported to have previously recovered from COVID-19 (or someone in their close social circle) at the time of their forecast underforecasted the number of deaths. This behavior was also common among forecasters who used more characters to describe their counterfactual thoughts and, thus may have anchored more strongly on the counterfactual scenario.

Our study contributes to the judgmental forecasting literature. Previous studies have highlighted several biases that forecasters are likely to suffer from when confronted with time series forecasting tasks[15]. For example, the general finding that human forecasters are likely to underforecast true values is not new, because forecasters have frequently been found to dampen trends in their judgmental predictions when the presented series is steeper than what is thought to be representative for the underlying data environment[17]. However, past research has so far only studied trend damping behavior in linear, stationary environments, but not in non-stationary settings such as COVID-19. Furthermore, a systematic test of the role of counterfactual thinking on (under-) forecasting performance provides insights into the circumstances in which judgmental underforecasting is particularly prevalent. Considering this, we demonstrate that, contrary to conventional wisdom, careful task deliberation, such as induced by counterfactual reasoning, does not always improve judgmental accuracy. Moreover, the majority of past studies have investigated judgmental biases in controlled laboratory environments, not accounting for the fact that the task may be emotionally charged. Instead, we build on an incentivized, naturalistic forecasting experiment with real personal outcomes, which can shed further lights on the determinants of predictive performance.

Our study has several implications for risk communication practices by public policy makers in the context of health crises. In fact, past studies have demonstrated that linguistic nuances in the communication of COVID-19-related information can severely affect consumer perceptions about danger as well as precautionary behavioral intentions[31]. Since counterfactual thinking can be induced through verbal descriptions of an event[6], our research suggests that the framing of crisis-related information in announcements provided by healthcare providers and mass media to consumers matters as it can result in misperceptions regarding the evolution of the crisis. However, counterfactual thinking triggered in a more passive way (e.g. by reading pre-formulated counterfactuals) may lead to weaker effects than the ones observed in our active manipulation. In order to counteract such misperceptions among consumers, public policy makers, marketers and the mass media should use neutral language to reduce the likelihood of triggering upward counterfactual thinking. It is also important to understand what kinds of messages trigger counterfactual thinking, which can help policy makers to adopt more effective communication strategies by carefully formulating public announcements. Moreover, similar to debiasing strategies adopted in the context of base-rate neglect in probability judgments, it is important to train consumers to interpret crisis-related information to better understand the underlying causes[32].

In conclusion, understanding how consumers interpret critical health crisis information is crucial, as it may lead to flawed decision-making and hinder the effectiveness of public policy interventions such as vaccination strategies or social distancing. Future research could explore the effect of counterfactual thinking on broader behavioral outcomes, including risk perceptions and decisions related to vaccination or willingness-to-pay for insurance products and test the generalizability of our findings in other high-impact crisis situations involving, for example, natural disasters and counterterrorism.

Data availability

The CSV data that support the findings of this study are available in OSF with the identifier https://doi.org/10.17605/OSF.IO/XPRWN (https://doi.org/10.17605/OSF.IO/XPRWN)[33].

Code availability

The R code used for data analysis is also available in OSF with the identifier https://doi.org/10.17605/OSF.IO/XPRWN (https://doi.org/10.17605/OSF.IO/XPRWN)[33].

Received: 31 May 2024; Accepted: 24 January 2025;
Published online: 04 February 2025

References

1. Anwar, A., Malik, M., Raees, V. & Anwar, A. Role of mass media and public health communications in the COVID-19 pandemic. *Cureus* **12**, e10453 (2020).
2. Jacob, C., Hausemer, P., Zagoni-Bogsch, A., & Diers-Lawson, A. (2023). The effect of communication and disinformation during the Covid-19 pandemic. *Special Committee on the COVID-19 pandemic: lessons learned and recommendations for the future, European Parliament*.
3. Richmond, B., Sharpe, L. & Menzies, R. E. Are Fear Campaigns Effective for Increasing Adherence to COVID-Related Mitigation Measures? *Int. J. Behav. Med.* **30**, 714–730 (2023).
4. Tomaino, G., & Walters, D. J. Presenting time-series data as absolute versus relative changes impacts judgments and choices. *J. Consumer Psychol.* **34**, 510–518 (2024).
5. Romano, A., Sotis, C., Dominioni, G., & Guidi, S. (2020). The public do not understand logarithmic graphs used to portray COVID-19. Impact of Social Sciences Blog.
6. Roese, N. J. Counterfactual thinking. *Psychol. Bull.* **121**, 133 (1997).
7. Epstude, K. & Roese, N. J. The functional theory of counterfactual thinking. *Personal. Soc. Psychol. Rev.* **12**, 168–192 (2008).
8. Bertolotti, M. & Catellani, P. Counterfactual thinking as a prebunking strategy to contrast misinformation on COVID-19. *J. Exp. Soc. Psychol.* **104**, 104404 (2023).
9. García Ferrés, E. A., & DePalma, M. T. Counterfactual thinking may attenuate polarization of COVID-19 prevention behavior. *Soc. Personality Psychol. Compass*. **17**, e12891 (2023).
10. Ahn, J. N., Hu, D. & Vega, M. Changing pace: Using implementation intentions to enhance social distancing behavior during the COVID-19 pandemic. *J. Exp. Psychol.: Appl.* **27**, 762 (2021).
11. Schlager, T., & Whillans, A. V. People underestimate the probability of contracting the coronavirus from friends. *Humanit. Soc. Sci. Commun.* **9,** 1–11 (2022).
12. Baek, T. H., Shen, L. & Reid, L. N. Effects of message framing in anti–binge drinking PSAs: The moderating role of counterfactual thinking. *J. Health Commun.* **18**, 442–458 (2013).
13. Wang, S., Kirillova, K. & Lehto, X. Reconciling unsatisfying tourism experiences: Message type effectiveness and the role of counterfactual thinking. *Tour. Manag.* **60**, 233–243 (2017).
14. Krishnamurthy, P. & Sivaraman, A. Counterfactual thinking and advertising responses. *J. Consum. Res.* **28**, 650–658 (2002).
15. Lawrence, M., Goodwin, P., O'Connor, M. & Önkal, D. Judgmental forecasting: A review of progress over the last 25 years. *Int. J. Forecast.* **22**, 493–518 (2006).
16. Recchia, G., Freeman, A. L. & Spiegelhalter, D. How well did experts and laypeople forecast the size of the COVID-19 pandemic? *PloS one* **16**, e0250935 (2021).
17. Harvey, N. & Reimers, S. Trend damping: Under-adjustment, experimental artifact, or adaptation to features of the natural environment? *J. Exp. Psychol.: Learn., Mem., Cognition* **39**, 589 (2013).
18. Smallman, R. & Summerville, A. Counterfactual thought in reasoning and performance. *Soc. Personal. Psychol. Compass* **12**, e12376 (2018).
19. Roese, N. J., & Epstude, K. (2017). The functional theory of counterfactual thinking: New evidence, new challenges, new insights. *In Advances in experimental social psychology* (Vol. 56, pp. 1-79). Academic Press.
20. Hoch, S. J. Counterfactual reasoning and accuracy in predicting personal events. *J. Exp. Psychol.: Learn., Mem., Cognition* **11**, 719 (1985).
21. Trope, Y. & Liberman, N. Construal-level theory of psychological distance. *Psychological Rev.* **117**, 440 (2010).
22. Grossmann, I. et al. Insights into the accuracy of social scientists' forecasts of societal change. *Nat. Hum. Behav*. **7**, 484–501 (2023).
23. Roser, M. and Ortiz-Ospina, E. (2021) – "Daily new confirmed COVID-19 deaths". Published online at OurWorldInData.org. Retrieved from: 'https://ourworldindata.org/explorers/coronavirus-data-explorer?zoomToSelection=true&hideControls=true&Metric=Confirmed+deaths&Interval=New+per+day&Relative+to+Population=false&Align+outbreaks=false&country=~USA'. Deposited on June 1, 2021.
24. Dong, E., Du, H. & Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Inf. Dis.* **20**, 533–534, https://github.com/CSSEGISandData/COVID-19 (2021). Accessible atDeposited on May 31, 2021.
25. Abadie, A., Athey, S., Imbens, G. W. & Wooldridge, J. M. When should you adjust standard errors for clustering? *Q. J. Econ.* **138**, 1–35 (2023).
26. Bertrand, M., Duflo, E. & Mullainathan, S. How much should we trust differences-in-differences estimates? *Q. J. Econ.* **119**, 249–275 (2004).
27. Yu, K., Lu, Z. & Stander, J. Quantile Regression: Applications and Current Research Areas. *J. R. Stat. Soc. Ser. D. (Statistician)* **52**, 331–350 (2003).
28. Bai, J. & Perron, P. Estimating and Testing Linear Models With Multiple Structural Changes. *Econometrica* **66**, 47–78 (1998).
29. Armstrong, J Scott. 1985. Long-Range Forecasting: From Crystal Ball to Computer. Wiley.
30. Campbell, M. C., Inman, J. J., Kirmani, A. & Price, L. L. In times of trouble: A framework for understanding consumers' responses to threats. *J. Consum. Res.* **47**, 311–326 (2020).
31. Mecit, A., Shrum, L. J. & Lowrey, T. M. COVID-19 is feminine: Grammatical gender influences danger perceptions and precautionary behavioral intentions by activating gender stereotypes. *J. Consum. Psychol.* **32**, 316–325 (2022).
32. Larrick, R. P. (2004). Debiasing. *Blackwell handbook of judgment and decision making*, 316-338.
33. https://doi.org/10.17605/OSF.IO/XPRWN.

Acknowledgements
The study was supported by research funds awarded by the Ministerio de Ciencia e Innovación España [https://doi.org/10.13039/501100011033, Grants HACOM - PID2022-138667OB-I00 and HDL-HS-280218].

Author Contributions
These authors contributed equally. MS focused on survey design and data collection, JS worked on the empirical strategy and analysis. Both authors wrote the manuscript.

Competing interest
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43856-025-00751-8.

Correspondence and requests for materials should be addressed to Matthias Seifert.

Peer review information *Communications Medicine* thanks Noemi Mazzoni and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2025