Information and Behavioral Responses during a Pandemic: Evidence from Delays in Covid-19 Death Reports*

Emilio Gutierrez† Adrian Rubli‡ Tiago Tavares§

May 2021

Information is thought to be an important policy tool for managing epidemics. In particular, providing the public with data that tracks the severity of an outbreak – such as case and death counts – may allow individuals to assess risks and modify behaviors. However, issues with data collection and quality may hinder these efforts. Exploiting publicly available administrative data and conducting an online survey within the context of Covid-19 in Mexico, we provide evidence that behavior, and consequently the evolution of the pandemic, are considerably different when death counts are presented by date reported or by date occurred, due to non-negligible reporting delays. We then use an equilibrium model incorporating an endogenous behavioral response to illustrate how reporting delays lead to slower responses by individuals, and consequently, worse epidemic outcomes.

JEL codes: I12; I18; D83; H12

Key words: information; reporting delays; behavior; social distancing; Covid-19

*The authors acknowledge support from the Asociación Mexicana de Cultura and the ITAM-COVID center. We thank Jose Maria Barrero, Andrew Foster, Miguel Messmacher, Charles Wyplosz, and participants at the ITAM Brown Bag seminar for their helpful comments. Gerardo Sánchez-Izquierdo provided outstanding research assistance. This project received approval from the ITAM Institutional Review Board. None of the authors have any interests to declare. All errors are our own.

†Instituto Tecnológico Autónomo de México (ITAM), Department of Economics. Camino a Santa Teresa 930. Colonia Heroes de Padierna, Magdalena Contreras CDMX 10700, Mexico. Email: emilio.gutierrez@itam.mx

‡ITAM, Department of Business Administration. Email: adrian.rubli@itam.mx

§*Corresponding author.* ITAM, Department of Economics and CIE. Email: tiago.gomes@itam.mx

1 Introduction

The swift emergence of the Covid-19 global epidemic forced governments to adopt new policies and communication strategies (WHO, 2013, 2020). Notably among these policies were lengthy, albeit sometimes lumpy, efforts to inform the population of how the virus spreads and basic (non-pharmaceutical) actions to minimize contagion. An important novelty of this epidemic, relative to past outbreaks, was the dissemination of vast amounts of high-frequency (oftentimes daily) information about the prevalence of Covid-19 cases and deaths. Although the implicit assumption is that more informed agents are more likely to take actions to mitigate the spread of the virus, little empirical and theoretical work has focused on understanding the role of government information in the context of a pandemic.[1] If these reports indeed matter for behavioral responses,[2] then reliable real-time surveillance systems are paramount not just for tracking the epidemic, but also for managing it. This may be particularly challenging in low- and middle-income countries, where diminished state capacity may impede the collection of accurate instantaneous information.

In this paper, we shed light on the issues of government-provided information and its quality by asking whether the delays with which deaths are reported affect the evolution of the epidemic through their potential impact on behavior. We focus on the Covid-19 outbreak in Mexico, where reporting delays – that is, the time difference between when a death occurs and when it is publicly reported – are measurable and large. Hence, in this setting, daily death reports are not an accurate representation of the state of the epidemic.

Mexico provides an ideal setting for analyzing this issue for at least three reasons. First, the delays in Covid-19 death reports are not only measurable and large, but vary greatly across Mexican states.[3] Gutierrez et al. (2020) documents that these delays are correlated with local measures of the capacity of the public healthcare system. Figure 1 depicts these delays at the national level by showing cumulative deaths as reported versus as occurred as well as the distribution of

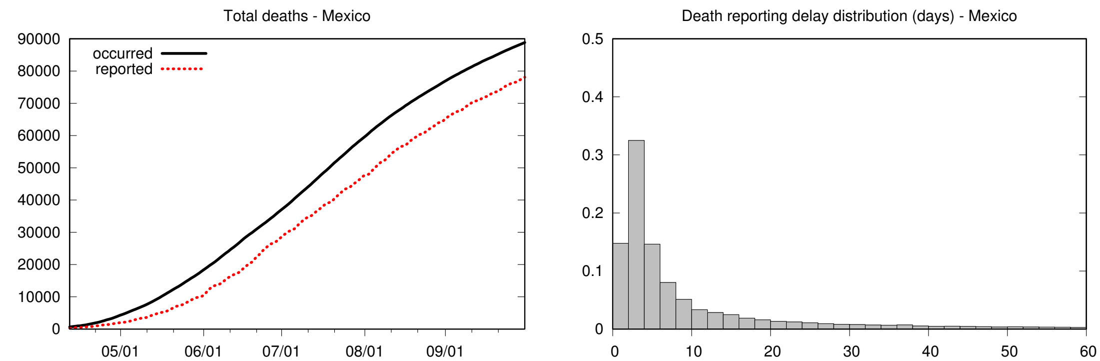
[1]In the context of Covid-19, it has been shown that information-focused policies in the US had the largest impact on limiting mobility (Gupta et al., 2020), and that in Italy information and communication were key tools for managing behavior and expectations (Briscese et al., 2020).

[2]For other contexts, the evidence regarding the impact of information provision on the adoption of mitigating behaviors is mixed. For instance, for HIV in Africa, studies have found large effects (Dupas, 2011; Dupas et al., 2018), while for vaccination in the US, it has been shown to be mostly ineffective (Nyhan and Reifler, 2015; Sadaf et al., 2013). Additional factors seem to mediate individuals' responses to information. For example, Oster (2012) shows that behavioral responses to reduce HIV risk in Africa are lower in places where non-HIV mortality is high.

[3]Delays in reporting deaths are a well-known problem, documented across a variety of settings (AbouZahr et al., 2015; Bird, 2015).

reporting delays in days. Second, Mexican officials routinely present information on confirmed Covid-19 deaths over time, giving particular salience to the number of *reported* deaths on each date.[4] Lastly, the Mexican government chose a relatively lenient strategy that consisted of mostly optional lockdowns and stay-at-home recommendations, implying that determinants of individual behavior have been crucial for the evolution of the pandemic in this setting.[5]

Figure 1:
Delays in Death Reports in Mexico: Occurred vs Reported Deaths



Notes: These plots depict reporting delays in Mexico. The plot on the left shows total deaths in Mexico up to September 30, 2020. The solid line corresponds to cumulative death counts based on the date of occurrence, while the dotted line uses the date on which deaths were reported. The plot on the right shows the distribution of delays in death reports measured in days (difference between when a death occurred and when it was reported). These graphs use information provided up to February 11, 2021.

Exploiting detailed daily data that allow us to separately count reported versus occurred cases and deaths, we begin by showing descriptive correlations. We document that the number of *reported* deaths is a better predictor of the growth in the number of Covid-19 cases than the number of *occurred* deaths. This suggests that individuals may incorrectly make inferences about the risk of contagion at any given moment assuming that deaths reported are a good approximation of deaths occurred on each date.

We then complement these empirical results by fielding an online survey, where we randomized information about the epidemic. Specifically, we compare respondents' beliefs regarding the severity of the epidemic and their reported intentions of complying with the government's shelter in place

[4]Information on deaths is presented by date occurred on this government website: https://coronavirus.gob.mx/datos/. During the daily Covid-19 press conference, the main statistic presented is the number of reported deaths, although trends by date of occurrence are also shown. See, for example, the first slide of the press conference presentation: https://presidente.gob.mx/conferencias-de-prensa-informe-diario-sobre-coronavirus-covid-19-ssa/.

[5]See, for example, https://www.informador.mx/mexico/No-habra-represion-para-detener-propagacion-del-COVID-19-reafirma-Lopez-Obrador-20200428-0039.html and https://piedepagina.mx/no-tenemos-camas-de-hospital-en-los-parques/, last accessed June 30, 2020.

recommendations between groups that were shown the evolution of Covid-19 deaths by the date on which they were reported versus the date on which they occurred. We find evidence consistent with individuals not fully accounting for reporting delays when adapting their behavior to the perceived risk prevalence.

Lastly, informed by these findings, we develop a simple equilibrium model that allows us to illustrate the impact that reporting information with delays has on the evolution of the pandemic through its impact on individuals' behavior. In the model, agents split their time at or away from home but risk getting infected when outside the home. Thus, the higher the prevalence of the virus among the population, the higher the incentive to stay at home. Assuming that agents are uninformed about the true prevalence of the virus, they rely on death reports provided by the government – which they take to be accurate – to form expectations about the current and future incidence of the disease. With each new report, agents discard their current expectations and form new ones.

We calibrate this model such that the dynamics of the epidemic resemble the first wave observed in Mexico. We compare model outcomes in a scenario where deaths are reported as occurred relative to a situation where delays in reporting follow the empirical distribution of delays for Mexico. Inaccurate information due to reporting delays leads to individuals being slower to adopt protective behaviors and to more severe epidemic outcomes in terms of cases and deaths, such that the peak of daily deaths is 25% larger than in the model without delays. Moreover, the faster speed of the epidemic induced by slower reactions will tend to generate excessive responses later on, which may exacerbate the negative economic impacts of the pandemic.

Overall, our analysis suggests that improving the capacity to collect accurate information and make it public matters for how well agents can react to the epidemic. More accurate information can also change the full dynamics of the epidemic itself with lower total deaths, smaller activity fluctuations, and less intense infection outbreaks, which may also imply a better management of hospital capacity constraints.[6] These results seem particularly relevant for many developing and emerging market economies with weak state capacity, as information can provide a relatively costless instrument for mitigating the spread of Covid-19 and other epidemic outbreaks.

[6] For example, Gutierrez and Rubli (2020) shows a strong relationship between hospital capacity and increases in in-hospital mortality during the 2009 H1N1 epidemic in Mexico.

We contribute to three strands of the growing literature on the economics of Covid-19.[7] First, our paper relates to those that have explored how messages and information affect various outcomes.[8] Akesson et al. (2020) provides different information about the infectiousness of Covid-19, finding that individuals who received the larger estimate of contagion risk were actually less likely to report complying with mitigating behaviors. Binder (2020) randomizes information about the Fed cutting interest rates, increasing consumers' optimism regarding unemployment and inflation. Coibion et al. (2020) randomizes information about different US government policies, finding a null impact on beliefs and spending plans of consumers, likely due to households' priors about the effectiveness of macroeconomic policies. While these studies focus on the effect of receiving information, our paper emphasizes the role of the accuracy of information received.

Second, we contribute to the recent literature that attempts to incorporate changes in behavior over the course of the pandemic into dynamic models that are aimed at predicting the evolution of the epidemic over time (Fernández-Villaverde and Jones, 2020; Brotherhood et al., 2020). The novelty in the model we propose consists in explicitly incorporating frictions in behavior that may emerge from misinformed agents.

Lastly, we add to the set of papers focusing on identifying the additional restrictions and challenges that low and middle-income countries face in managing the pandemic and subsequent economic recovery. Various studies have focused on features such as the capacity of the healthcare system, poverty, inequality, and corruption (Gallego et al., 2020; Gottlieb et al., 2020; Loayza, 2020; Monroy-Gómez-Franco, 2020; Ribeiro and Leist, 2020; Walker et al., 2020). We contribute to this line of work by focusing on the potential consequences of issues in collecting reliable real-time information during the pandemic. Given the relationship between reporting delays and state capacity (Gutierrez et al., 2020), this is likely to be an issue for many other low and middle-income countries.

[7]See Brodeur et al. (2020) for an overview of the Covid-19 literature in economics.

[8]Various papers in different settings have generally found that social distancing measures have a positive impact on containing the epidemic. See, for example, Hsiang et al. (2020); Dave et al. (2020); Alexander and Karger (2020); Juranek and Zoutman (2020); Jinjarak et al. (2020). Some mediating factors that the literature has analyzed include the roles of sociodemographic characteristics (Papageorge et al., 2020; Knittel and Ozaltun, 2020), political beliefs (Allcott et al., 2020; Baccini and Brodeur, 2020; Barrios and Hochberg, 2020), social capital (Bargain and Aminjonov, 2020; Brodeur et al., 2020; Ding et al., 2020; Durante et al., 2020), the media (Simonov et al., 2020; Bursztyn et al., 2020).

The rest of the paper is organized as follows. Section 2 presents the motivating correlations observed in the data. Section 3 presents descriptive statistics and results from our online survey. Section 4 outlines the equilibrium model and discusses the results. Section 5 concludes.

2 Motivating Correlations

2.1 Data

The Mexican government provides very detailed patient-level information for all recorded Covid-19 cases. This dataset is updated on a daily basis with new information. For each patient, we observe the patients' state of residence, the date on which they first sought medical attention for Covid-19 symptoms, the self-reported date for the onset of symptoms (all reported cases are symptomatic), the result of a Covid-19 laboratory test, and, if applicable, the date of death. With each daily update of the data, the information on the result of the Covid-19 lab test and the date of death may change.[9] No other variable changes across reports.

We illustrate the data with an example for clarity in Table 1. Suppose a patient first began feeling sick on March 30, 2020. They then waited until April 4 to go to the hospital. Because of delays in reporting, the centralized data system only started including this patient in the database until April 6. Their Covid-19 lab test did not confirm the infection until April 8, so that prior to that, the government reported it as a suspected case. This patient then died on April 9, but due to delays, did not appear in the dataset until April 13.

From the dataset published on February 11, 2021, we compute the number of Covid-19 cases and deaths per state-week according to the date on which they occurred. We restrict our attention to all weeks between March 14 and September 30, 2020, allowing up to four months for all occurred cases and deaths during this period to be reported in the system. We then recover the number of weekly reported cases and deaths in each state from the changes in the updated database from one week to the next. This effectively allows us to track the number of reported and occurred cases and deaths over time from March to September.

[9] To be clear, these changes in the date of death variable are only a transition from a missing value for patients that were alive to a date of death for patients that are now deceased.

Table 1:
Example of the Government Database of Covid-19 Patients

Date observed	Patient id	State resid.	Date medical attn.	Date symptoms	Covid test	Date of death
6-Apr	284859	CDMX	4-Apr	30-Mar	Susp.	–
7-Apr	284859	CDMX	4-Apr	30-Mar	Susp.	–
8-Apr	284859	CDMX	4-Apr	30-Mar	Pos.	–
9-Apr	284859	CDMX	4-Apr	30-Mar	Pos.	–
10-Apr	284859	CDMX	4-Apr	30-Mar	Pos.	–
11-Apr	284859	CDMX	4-Apr	30-Mar	Pos.	–
12-Apr	284859	CDMX	4-Apr	30-Mar	Pos.	–
13-Apr	284859	CDMX	4-Apr	30-Mar	Pos.	9-Apr
14-Apr	284859	CDMX	4-Apr	30-Mar	Pos.	9-Apr

Notes: This table shows a hypothetical example of the government data. The first column shows the date on which we observe the data, corresponding to each daily update. The third column is the patient's state of residence, which is Mexico City here. The fourth column is the date on which the patient first sought medical attention. The Covid test result can be positive (Pos.) or it can be a suspected case (Susp.) before it is lab-confirmed.

2.2 Empirical Correlations in the Data

We compute two similar measures of the growth rate of Covid-19 cases for each state in each period. We take the percentage change in the number of patients that self-reported having first shown symptoms during week t and during week $t+1$, and the change from week t to week $t+2$. In order to explore the relationship between the growth of the epidemic and deaths as reported or occurred, we estimate regressions of the following form:

$$y_{s,t} = \beta_1 \times \ln(\text{Occurred Deaths})_{s,t-1} + \beta_2 \times \ln(\text{Reported Deaths})_{s,t-1} + \lambda_s + \gamma_t + \Pi\mathbf{X}_{s,t} + \varepsilon_{s,t} \quad (1)$$

where $y_{s,t}$ is the percentage change in the number of patients having first shown symptoms during week t and week $t+1$ (or week $t+2$), λ_s are state fixed effects, γ_t are week fixed effects, $\mathbf{X}_{s,t}$ is a vector of observable characteristics which may correlate with the growth rate of Covid-19 cases, namely, the number of Covid-19 cases in state s in periods t and $t-1$, their growth rate in period $t-1$, and the the growth rate of reported and occurred cases in week $t-2$. The error component is denoted by $\varepsilon_{s,t}$.

The main regressors of interest are $\ln(\text{Occurred Deaths})_{s,t-1}$ and $\ln(\text{Reported Deaths})_{s,t-1}$, which measure the log of newly occurred and reported deaths in state s in week $t-1$. If individuals change their behavior – for instance, by reducing their exposure when the perceived risk of death is higher – we expect the coefficient associated with these two variables to be negative. In other words, the epidemic should grow at a lower rate as individuals take more precautions. If, however, the risk prevalence is inferred from deaths reported instead of deaths occurred, β_2 should be larger in magnitude than β_1.

The results from this exercise are presented in Table 2. The dependent variable in the first three columns is the percentage change in the number of patients declaring having first shown symptoms during week t and week $t+1$. For the last three columns, the dependent variable is the percentage change in the number of patients declaring having first shown symptoms during week t and week $t+2$. Columns 1 and 4 only include state and week fixed effects and the log of the number of Covid-19 cases in week t as controls. Columns 2 and 5 additionally include the percent change in new cases between week $t-2$ and $t-1$. Columns 3 and 6 also control for the percent change in occurred deaths and reported deaths between week $t-2$ and week $t-1$.

Table 2:
Correlates of the Growth in Covid-19 Cases

	Growth in Cases from t to $t+1$			Growth in Cases from t to $t+2$		
	(1)	(2)	(3)	(4)	(5)	(6)
ln(Reported Deaths)$_{s,t-1}$	-0.0852***	-0.0821***	-0.0841**	-0.144**	-0.149**	-0.172**
	[0.0244]	[0.0248]	[0.0337]	[0.0571]	[0.0565]	[0.0817]
ln(Occurred Deaths)$_{s,t-1}$	0.0103	0.0133	0.0179	-0.0254	-0.0300	0.0384
	[0.0303]	[0.0296]	[0.0331]	[0.0713]	[0.0723]	[0.0875]
Cases in Period t	Yes	Yes	Yes	Yes	Yes	Yes
Growth in Cases $t-1$	No	Yes	Yes	No	Yes	Yes
Growth in Deaths $t-1$	No	No	Yes	No	No	Yes
Observations	734	734	700	734	734	700
R-squared	0.495	0.497	0.470	0.548	0.549	0.565

Notes: This table shows how reported and occurred deaths correlate with the growth in Covid-19 cases as reported in the data. Observations are at the state-week level and we show estimates of equation 1. Columns 1-3 show the growth rate from week t to $t+1$, while columns 4-6 consider the rate from t to $t+2$. All regressions include state and week fixed effects. Different columns include different additional controls as indicated. Robust standard errors are presented in brackets.
*** p<0.01, ** p<0.05, * p<0.1

Across specifications, the coefficient associated with the number of reported deaths is negative and significantly different from zero at a high confidence level. In contrast, the coefficient associated with the number of occurred deaths is of a much smaller magnitude and not significantly different from zero. Results are very stable across specifications.

We interpret these correlations as motivating evidence that the growth rate in Covid-19 cases is more responsive to reported than to occurred deaths. We conjecture that this relationship may be driven by individuals incorrectly inferring the prevalence of Covid-19 in the population at any given time from the number of reported, rather than occurred, total deaths in each period. The observations derived from Table 2 motivate us to further explore whether reporting delays could impact both individuals' perceptions and actions.

3 Online Survey

3.1 Survey description and respondents' characteristics

In order to shed further light on whether delays affect individuals' perceptions, we conducted an online survey with a randomized informational treatment that presented the evolution of total deaths either by date reported or by actual date of death. The full survey consisted of 48 closed-response multiple choice questions and ran from May 28 to June 8, 2020.[10] We recruited participants via email and social media (namely, Twitter), and respondents were not compensated for participating.[11] Our final sample consists of 1,022 completed surveys.

The first set of questions were aimed at recovering socioeconomic characteristics of respondents, as well as their pre-intervention perceptions about the state of the Covid-19 pandemic in Mexico. We asked questions related to age, gender, state of residence, household composition, income, employment status, and a self-reported estimation of the number of Covid-19 cases and deaths up to May 20, 2020.

After these initial questions, respondents were taken to a new screen showing (randomly) one of the two graphs depicted in Figure 2. A total of 508 participants were shown Figure 2a, which plots cumulative deaths in Mexico by date on which they were reported. The remaining 514 participants were shown Figure 2b, which instead plots cumulative deaths by actual date of occurrence. Both figures show data from March 22 to May 15, using data up to May 27.[12] Additionally, we included the cumulative number of deaths by date reported in Sweden as a reference.[13] Both figures shown

[10]See online appendix B for a full translation of the survey questions and response options.

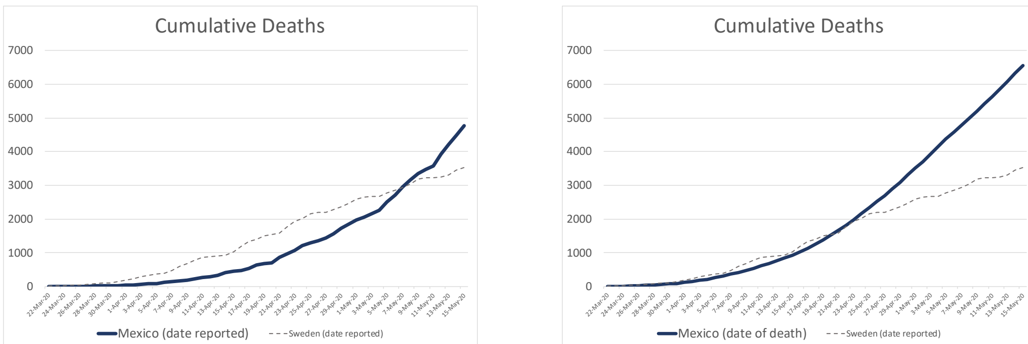
[11]Our mailing list was obtained from ITAM, and consisted of all faculty, administrative staff, and students.

[12]This means that we allow for deaths to be reported with a lag of at most 12 days.

[13]The data for Sweden were obtained from https://ourworldindata.org/coronavirus. Sweden followed a similar strategy to Mexico, imposing relatively light restrictions (Juranek and Zoutman, 2020). The trajectory of the epidemic

to respondents contain truthful information as presented by government authorities themselves. Note that, due to reporting delays, total deaths by date reported in Figure 2a understate total deaths by date occurred (Figure 2b) by 41% on average, with a difference of up to 2,055 deaths on May 11.

Figure 2:
Information Treatments in the Survey



(a) Cumulative deaths by date reported (b) Cumulative deaths by date of occurrence

Notes: These graphs show the information treatments that we randomized in the survey. Respondents were shown these exact figures with captions translated into Spanish. Each plot shows data from March 22 to May 15, using information reported up to May 27, 2020. We include the cumulative number of deaths by date reported in Sweden as a reference. The plot on the left shows the cumulative deaths in Mexico based on the date they were reported. The plot on the right shows cumulative deaths by date on which they actually occurred.

Afterwards, participants answered additional questions regarding whether they believed the epidemic in Mexico was evolving faster or slower than in Sweden, the expected number of total Covid-19 cases and deaths over the whole epidemic outbreak, and the number of times they expected to leave their home in the following weeks.

Table 3 tests for balance in observable characteristics between our treatment groups. Columns 1 and 2 present means for the sample that was shown the graph with cumulative deaths by date reported and by date occurred, respectively. Column 3 shows the corresponding difference in means. It is worth highlighting that, due to the nature of the survey conducted, the characteristics of participants suggest they belong to a relatively young, educated, and high-income group in Mexico. More than 78 percent of them live in Mexico City, and more than half of them live in a house with a yard. Evidently, this implies that none of our results can be used to infer the

in Mexico had been compared to that in Sweden by Mexican authorities a few days before the survey was implemented. See, for example, https://twitter.com/HLGatell/status/1257694745322819586?s=20 and https://www.milenio.com/politica/ya-aplanamos-la-curva-lopez-gatell, last accessed June 29, 2020.

distribution of beliefs and behavior in the general Mexican population.[14] However, given the very small differences in observables between our two treatment groups, we can confidently interpret the results below as the impact of the information provided on the different outcomes.

3.2 Empirical Strategy

We explore the impact of the information provided in the survey on four main outcomes. Specifically, we focus on participants' responses to questions regarding whether the epidemic was evolving faster or slower than in Sweden, the predicted number of total Covid-19 cases and deaths over the course of the current outbreak, and the number of times respondents reported they expected to leave their home four weeks after the survey was conducted. Figures A1 and A2 in the online appendix show histograms of the responses to the relevant survey questions.

We estimate the following equation to measure differences in perceptions and behavior between our treatment groups:

$$y_i = \alpha_0 + \alpha_1 \times [\text{Info By Date Occurred}]_i + \Psi \mathbf{X}_i + \nu_i \tag{2}$$

where y_i is an outcome variable for respondent i, α_0 is a constant, $[\text{Info By Date Occurred}]_i$ is a zero-one indicator for having received the informational treatment that displayed cumulative deaths by actual date of death, $\mathbf{X}_i$ is a vector of observable characteristics as listed in Table 3, and ν_i is the error term. Our estimate of interest corresponds to α_1, which measures the average difference in the outcome variable for survey respondents that were shown the cumulative death toll by date of occurrence with respect to those who were shown the information by date reported.

For simplicity, we construct two alternative binary measures of the answers to the question of whether the epidemic is evolving faster in Mexico than in Sweden. Specifically, the first outcome takes a value of one if respondents considered the speed of the epidemic's evolution to be faster or much faster than in Sweden, while the second variable is only equal to one if respondents' considered the evolution to be much faster. For constructing the variables measuring respondents' beliefs about the toll of the pandemic, we assign the total number of expected cases and deaths to be equal to the mid-value for the interval chosen by respondents. For respondents who chose

[14] Note also that our sample does not have enough variation to weight it so that it is representative of the entire population in Mexico.

Table 3:
Balance Table for Survey Covariates

	Informational treatments		
	Deaths by date reported	Deaths by date occurred	Difference in means
Female	0.496	0.490	-0.006
	(0.500)	(0.500)	(0.031)
Ages 18-22	0.321	0.383	0.062**
	(0.467)	(0.487)	(0.030)
Ages 23-29	0.274	0.253	-0.021
	(0.446)	(0.435)	(0.028)
Ages 30-49	0.230	0.216	-0.014
	(0.421)	(0.412)	(0.026)
Ages 50+	0.175	0.148	-0.027
	(0.381)	(0.355)	(0.023)
Works	0.409	0.329	-0.081***
	(0.492)	(0.470)	(0.030)
Attends school	0.368	0.416	0.048
	(0.483)	(0.493)	(0.031)
Works and attends school	0.157	0.158	0.000
	(0.365)	(0.365)	(0.023)
Other occupation/employment status	0.065	0.097	0.032*
	(0.247)	(0.297)	(0.017)
Lives in Mexico City	0.776	0.753	-0.023
	(0.418)	(0.432)	(0.027)
Lives in apartment	0.343	0.385	0.043
	(0.475)	(0.487)	(0.030)
Lives in house, no yard	0.124	0.117	-0.007
	(0.330)	(0.321)	(0.020)
Lives in house with yard	0.533	0.498	-0.035
	(0.499)	(0.500)	(0.031)
Household size: 1-2	0.232	0.251	0.019
	(0.423)	(0.434)	(0.027)
Household size: 3	0.207	0.245	0.038
	(0.405)	(0.431)	(0.026)
Household size: 4	0.252	0.226	-0.026
	(0.435)	(0.418)	(0.027)
Household size: 5+	0.561	0.504	-0.057*
	(0.497)	(0.500)	(0.031)
Has HH members over 70 years old	0.159	0.080	-0.080***
	(0.366)	(0.271)	(0.020)
Has HH members 60-70 years old	0.215	0.202	-0.012
	(0.411)	(0.402)	(0.025)
Has HH members 50-60 years old	0.461	0.471	0.010
	(0.499)	(0.500)	(0.031)
Does not seek healthcare when sick	0.140	0.154	0.014
	(0.347)	(0.361)	(0.022)
Self-medicates when sick	0.386	0.381	-0.005
	(0.487)	(0.486)	(0.030)
Observations	508	514	1,022

Notes: This table shows means and standard deviations for a series of covariates asked in the survey before the informational treatment was presented. We show statistics separately for each informational treatment, as well as the difference in the means. Stars denote significance from a difference in means test.
*** p<0.01, ** p<0.05, * p<0.1

the option "more than 2,000,000 cases", we assigned a value of 3,000,000, and for those who chose "more than 200,000 deaths", we assigned 300,000. For the social distancing outcome, we use both the number of days respondents expect to leave their homes in four weeks time (0, 1, 2, 3.5 or 5, given the options provided), and a binary variable taking value of one if respondents expected to leave their house three or more times.[15]

3.3 Results

The direction in which the treatment may affect beliefs about the evolution of the epidemic potentially depends on respondents' priors. Before the treatment, survey participants were asked to report their knowledge about the total number of Covid-19 cases recorded in Mexico by May 20, a full week before the launch of the survey. We use the responses to this question to stratify the sample into a low and high prior group.[16] The low prior subsample consists of those that reported that the total number of Covid-19 cases was lower than 50,000 (47.7 percent of the full sample), while the high prior group are those that reported over 50,000 cases.[17]

We present all our results for the full sample and separately for these two subgroups. If our informational treatment is shifting beliefs about the epidemiological curve, then we would expect to see stronger and larger effects among the low prior group, as they are the ones that would update their priors upward.

Table 4 shows our main results. Panel A corresponds to their assessment of whether the epidemic in Mexico was evolving faster than in Sweden (faster or much faster in columns 1-3, and much faster in columns 4-6). Panel B corresponds to the log of the expected number of total cases (columns 1-3) and total deaths (columns 4-6). And Panel C corresponds to our measures of self-reported compliance with social distancing (the continuous measure in columns 1-3, and the indicator variable for three or more times in columns 4-6). Throughout Table 4, columns 1 and 4 present results for

[15]For completeness, we show results in online appendix Figures A3 - A5 using indicators for each of the possible responses to each of the questions that make up our outcome variables.

[16]Tables A1 and A2 in the online appendix show balance tables separately for the low and high priors subsamples.

[17]The true number reported in the nightly press conference on May 20 was 56,594 cumulative cases in the country (see https://twitter.com/HLGatell/status/1263264663283908609?s=20, last accessed June 29, 2020). A histogram with the distribution of the responses to this question is presented in online appendix Figure A6. Stratifying the sample based on individuals' prior regarding total reported deaths by May 20 yields similar results (available upon request).

Table 4:
Estimates of Informational Treatments on Perceptions and Behavior

	(1) Full sample	(2) Low prior	(3) High prior	(4) Full sample	(5) Low prior	(6) High prior
Panel A: Pandemic's evolution	Compared to Sweden					
	Faster or much faster			Much faster		
Information by date occurred	0.195***	0.169***	0.198***	0.258***	0.220***	0.266***
	(0.024)	(0.035)	(0.034)	(0.030)	(0.044)	(0.041)
Observations	1,022	488	534	1,022	488	534
R-squared	0.077	0.100	0.097	0.082	0.105	0.101
Mean dependent variable	0.81	0.82	0.80	0.37	0.39	0.35
Panel B: Expected toll	Beliefs on full impact of epidemic outbreak					
	Log expected total cases			Log expected total deaths		
Information by date occurred	0.144***	0.203**	0.108	0.108**	0.173**	0.0414
	(0.055)	(0.085)	(0.072)	(0.055)	(0.082)	(0.074)
Observations	1,022	488	534	1,022	488	534
R-squared	0.020	0.044	0.028	0.014	0.035	0.022
Mean dependent variable	12.97	12.88	13.05	10.81	10.76	10.85
Panel C: Social distancing	Number of times expected to leave the house in 4 weeks					
	Number of times			Three or more times		
Information by date occurred	-0.0442	-0.286**	0.154	-0.0368	-0.105**	0.0161
	(0.093)	(0.136)	(0.130)	(0.029)	(0.042)	(0.041)
Observations	1,022	488	534	1,022	488	534
R-squared	0.119	0.120	0.150	0.089	0.096	0.115
Mean dependent variable	2.20	2.14	2.25	0.35	0.33	0.37

Notes: This table presents the results from estimating equation 2. Each panel corresponds to two different outcome variables constructed from survey responses (see text for details). Columns 1 and 4 show results for the full sample. Columns 2, 5, 3 and 6 stratify the sample by respondents' prior on their knowledge of the number of Covid-19 cases in Mexico up to May 20 into low and high reported cases, respectively. The estimates are the average difference between the responses in the treatment group that received information based on the actual date of death relative to information based on date of reports. Robust standard errors are reported in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

the full sample, columns 2 and 5 restrict to the low prior subsample, and columns 3 and 6 focus on the high prior subsample.[18]

Presenting cumulative deaths by actual date of occurrence seems to shift beliefs towards a perception of a more rapidly spreading epidemic. For example, for individuals that were shown cumulative deaths by date of death, the fraction of respondents considering that the epidemic was progressing much faster than in Sweden is 26 percentage points higher. Also, respondents predict a higher number of Covid-19 cases and deaths when shown the evolution of deaths by date of occurrence than by reported date.

While the outcomes in Panel C measure self-reported intentions of complying with stay-at-home recommendations – which may not necessarily map perfectly to actual behavior – the results are consistent with information presented by actual date of death having an effect on expectations to comply with social distancing measures. Having been shown the graph by date of occurrence is associated with a significant decrease in the number of times people expect to leave their homes. Once again, the effect is concentrated in the low prior sample.

Notwithstanding the limited statistical power due to the small sample size of the survey and the relatively small differences in the information provided to each group, we interpret the results presented in Table 4 as evidence that the delays with which deaths are reported are very likely to affect perceptions about the state of the epidemic and, consequently, compliance with social distancing measures. These findings also suggest that the individuals surveyed, despite being a selected sample of higher-income respondents, do not fully incorporate reporting delays when forming beliefs about the epidemic. We proceed then by taking these insights and incorporating them into an equilibrium model.

4 Model of Equilibrium Behavior

The previous sections documented that reporting delays in the information presented to individuals may affect their beliefs and behavior during an epidemic. We now present an equilibrium model

[18] Additional results, available upon request, show that the treatment also induced the low-prior sample to be more likely to believe the risk of contagion at a 100-person gathering (in one or four weeks) to be very high, and less likely to expect to leave their home three or more times during the week after the survey was conducted.

to illustrate the impact of reporting delays on the epidemic dynamics through the endogenous behavioral response of agents.

4.1 Baseline model

The equilibrium model we introduce follows Greenwood et al. (2019) and Brotherhood et al. (2020), with Covid-19-specific compartments as in Fernández-Villaverde and Jones (2020). The structure is purposely simple and allows for standard extensions.[19] The main difference introduced in this model is related with how agents update expectations about the prevalence of Covid-19. Given knowledge about an epidemic and a prior associated with prevalence, agents, who understand the disease dynamics, form plans about consumption and leisure over their life-cycle.

However, as the epidemic progresses, the government publishes data about deaths which is assumed by agents to be accurate. Given a deviation of deaths reported by the government and initial forecasts for a particular date, agents discard their current belief about prevalence and update their prior in order to rationalize the number of deaths as reported by the government. This updated belief about the prevalence also changes their planning over economic decisions in the remainder of their life-cycle, thus affecting the dynamics of the epidemic. We fully describe the model in what follows.

States. We set up the model in discrete time, with each period corresponding to one day. The economy is populated with a continuum of ex-ante identical agents. Agents can spend time outside the home (work or leisure), or at home. Given an outbreak of Covid-19, let j be an agent's health status. The initial state is never infected and corresponds to $j = s$ (**susceptible**). Time spent outside the home implies a risk of contracting the disease that leads to $j = i$ (**infected**). Agents in an infected state can contaminate susceptible ones (assuming a uniform mixing contact rate), but with probability γ contagiousness ceases and a **recovering** process follows denoted by $j = c$. Exiting this state occurs with probability θ, with a share $1 - \delta$ of these becoming fully **recovered** with $j = r$ and the remaining share δ **die** with status $j = d$.

[19] Extensions include macroeconomic implications, savings, non-pharmacy initiatives, testing, vaccines, optimal lockdowns, age heterogeneity, and asset heterogeneity. Some of these extensions have been studied, for example, in Eichenbaum et al. (2020a), Eichenbaum et al. (2020b), Acemoglu et al. (2020), Alvarez et al. (2020), and Kaplan et al. (2020).

As an assumption, recovered individuals become permanently immune to the virus. We also assume that hours at or outside home while an individual is infected or recovering is constrained below the normal limit. To summarize, agents can be in each of the following states $j = \{s, i, c, r, d\}$. Without the pandemic, agents discount the future at rate β.

Utility and hours. Each agent is also endowed with a single unit of labor every period, divided into work/leisure hours n outside home and hours at home h:

$$1 = n + h.$$

Flow utility is derived from hours outside and at home according to $u(n, h) = \log n + \lambda_h \log h + b$, where b captures the benefit of remaining alive over being dead, which delivers a normalized utility of zero. Combined with the endowment of hours, flow utility of being alive is given by:

$$u(n) = \log n + \lambda_h \log(1 - n) + b.$$

Infections. Susceptible agents are at risk of infection when spending hours outside the home. The probability of getting infected π is assumed to be proportional to the time spent outside the home n and a belief about the aggregate transmission risk $\tilde{\Pi}_t$ that is allowed to be different from the real transmission risk Π_t:

$$\pi\left(n, \tilde{\Pi}_t\right) = n\tilde{\Pi}_t. \tag{3}$$

Value functions. For susceptible agents, the value function is given by:

$$V(s, t) = \max_{n \in (0,1)} \left\{ u(n) + \beta \left(\left[1 - \pi\left(n, \widetilde{\Pi}_t\right) \right] V(s, t+1) + \pi\left(n, \widetilde{\Pi}_t\right) V(i) \right) \right\}. \tag{value susceptible}$$

Infected agents generate a value of:

$$V(i) = \max_{n \in (0,\bar{n})} \left\{ u(n) + \beta \left[\gamma V(c) + (1 - \gamma) V(i) \right] \right\}. \tag{value infected}$$

where $\bar{n} < 1$ is the maximum amount of hours an infected individual can spend outside. Recovering individuals, who face the same restrictions in hours supplied as infected agents, have a value of:

$$V(c) = \max_{n \in (0,\bar{n})} \{u(n) + \beta((1-\theta)V(c) + \theta[(1-\delta)V(r) + \delta V(d)])\}. \qquad \text{(value recovering)}$$

Once individuals fully recover, the supply of hours becomes unrestricted and the value is characterized by:

$$V(r) = \max_{n \in (0,1)} \{u(n) + \beta V(r)\}. \qquad \text{(value recoverd)}$$

Lastly, the value for individuals that die is given by $V(d) = 0$.

Laws of motion. Letting $n(j,t)$ be the optimal outside home choice of hours for states $j = s, i, c, r$, we then have the following laws of motion for the mass of agents in the different health states:

$$M_{t+1}(s) = M_t(s) - \pi(n(h,t), \Pi_t) M_t(s) \qquad \text{(mass susceptible)}$$

$$M_{t+1}(i) = M_t(i) - \gamma M_t(i) + \pi(n(h,t), \Pi_t) M_t(s) \qquad \text{(mass infected)}$$

$$M_{t+1}(c) = M_t(c) - \theta M_t(c) + \gamma M_t(i) \qquad \text{(mass infected)}$$

$$M_{t+1}(r) = M_t(r) + (1-\delta)\theta M_t(c) \qquad \text{(mass resolving)}$$

$$M_{t+1}(d) = M_t(d) + \delta\theta M_t(c). \qquad \text{(mass death)}$$

The total population is normalized to 1 implying that, in every period t:

$$1 = M_t(s) + M_t(i) + M_t(c) + M_t(r) + M_t(d).$$

Aggregate probability of infection. We let the (instantaneous) Poisson rate of infection to be given by:

$$\hat{\Pi}_t = \Lambda n(i,t) M_t(i), \tag{4}$$

where Λ is the biological transmissibility of the disease. It follows that the probability of being infected before the end of the period becomes:[20]

$$\Pi_t = 1 - \exp\left(-\hat{\Pi}_t\right) = 1 - \exp\left[-\Lambda n\left(i,t\right) M_t\left(i\right)\right]. \tag{5}$$

Information and priors. Agents accurately know all the parameters of the model, but are unaware of the initial mass of infectious individuals $M_0\left(i\right)$. Thus, they form a prior about this mass $\tilde{M}_0\left(i\right)$. However, their belief over the prevalence risk may differ from the correct one due to an inaccurate prior, that is, $\tilde{M}_0\left(i\right) \neq M_0\left(i\right)$. If this is the case, forecasts made by agents over the epidemic dynamics are biased. In particular, given a history of labor supply $\left\{n\left(s,t\right), n\left(i,t\right)\right\}_{t=0}^{t'} \equiv \left\{n\left(s\right)^{t'}, n\left(i\right)^{t'}\right\}$ for any $t' \geq 0$, the following probabilities emerge as potentially different:

$$\Pi_{t'} = 1 - \exp\left[-\Lambda n\left(i,t'\right) M_{t'}\left(i; n\left(s\right)^{t'}, n\left(i\right)^{t'}\right)\right] \tag{6}$$

$$\tilde{\Pi}_{t'} = 1 - \exp\left[-\Lambda n\left(i,t'\right) \tilde{M}_{t'}\left(i; n\left(s\right)^{t'}, n\left(i\right)^{t'}\right)\right], \tag{7}$$

where $M_{t'}$ and $\tilde{M}_{t'}$ are obtained from substituting $\left\{n\left(s\right)^{t'}, n\left(i\right)^{t'}\right\}$ in the laws of motions (mass susceptible) and (mass infected) using, respectively, M_0 and $\tilde{M}_0$, from $t = 0..., t'$.

4.2 Definition of an equilibrium

A belief-biased equilibrium in this economy with a mass of agents at time $t' \geq 0$ of $M_{t'}\left(j\right), j = s, i, c, r, d$ consists in a sequence of infection probabilities $\left\{\Pi_t\right\}_{t=t'}^{\infty}$ and $\left\{\tilde{\Pi}_t\right\}_{t=t'}^{\infty}$, initial beliefs $\tilde{M}_{t'}\left(j\right)$, and hour allocations $\left\{n\left(j,t\right)\right\}_{t=t'}^{\infty}$ for each $j \in \left\{s, i, c, r\right\}$, such that:

1. given $\tilde{M}_{t'}\left(j\right)$ and $\left\{\tilde{\Pi}_t\right\}_{t=t'}^{\infty}$, $n\left(j,t\right)$ solves the values in (value susceptible)-(value recoverd);

2. given $\left\{n\left(j,t\right)\right\}_{t=t'}^{\infty}$, the resulting laws of motion from (mass susceptible)-(mass death) using $M_{t'}\left(j\right)$ are consistent with $\left\{\Pi_t\right\}_{t=t'}^{\infty}$;

3. given $\left\{n\left(j,t\right)\right\}_{t=t'}^{\infty}$, the resulting laws of motion from (mass susceptible)-(mass death) using $\tilde{M}_{t'}\left(j\right)$ are consistent with $\left\{\tilde{\Pi}_t\right\}_{t=t'}^{\infty}$.

[20]Given an instantaneous Poisson rate of infection $\hat{\Pi}$, the probability of infection within $\bar{t}$ time is given by an exponential distribution with $Prob\left(t < \bar{t}\right) = 1 - \exp\left(-\hat{\Pi}\bar{t}\right)$. For a single period, $\bar{t} = 1$, we then have that $Prob\left(t < 1\right) = 1 - \exp\left(-\hat{\Pi}\right)$.

4.3 Model analysis

Note that the static solution of hours spent outside the home is given by the following optimization problem:

$$n^* = \arg\max_{n\in(0,1)} \{u(n)\} = \frac{1}{1+\lambda_h}.$$

These are also the hours spent by fully recovered individuals outside the home:

$$n(r,t) = n^*.$$

Further assuming that the maximum amount of hours infected individuals may spend away from home is capped by the static optimal, $\bar{n} < n^*$, infected and recovering individuals supply the following amount of hours:

$$\begin{aligned} n(i,t) &= \bar{n} < n^* \\ n(c,t) &= \bar{n} < n^*. \end{aligned}$$

With these hours supplied, the following values have a closed-form solution:

$$\begin{aligned} V(r) &= \frac{u(n^*)}{1-\beta} \\ V(c) &= \frac{u(\bar{n}) + \beta\theta(1-\delta)V(r)}{1-\beta(1-\theta)} \\ V(i) &= \frac{u(\bar{n}) + \beta\gamma V(c)}{1-\beta(1-\gamma)}. \end{aligned}$$

As for the problem faced by healthy agents, first order conditions imply:

$$\begin{aligned} \frac{\partial u(n)}{\partial n} &= \beta \frac{\partial \pi\left(n, \widetilde{\Pi}_t\right)}{\partial n} \left(V(h,t+1) - V(i)\right) \\ \Rightarrow \quad \frac{1}{n} - \frac{\lambda_h}{1-n} &= \beta\widetilde{\Pi}_t \left(V(h,t+1) - V(i)\right), \end{aligned}$$

That is, the marginal benefit of spending hours outside is equated with the discounted expected marginal cost of being infected in utility units. It is also easy to see that in this environment, $V(h,t+1) > V(i)$ for any $\pi\left(n, \tilde{\Pi}_t\right) > 0$. This implies that hours supplied by susceptible individuals are $n(h,t) < n^*$ for any $\tilde{\Pi}_t > 0$, and moreover, the larger is the perceived risk of transmissibility $\tilde{\Pi}_t$, the lower is the supply of hours outside the home. In other words, susceptible agents have an incentive to reduce the amount of hours spent outside to avoid a perceived risk of infection.

4.4 Update of beliefs based on deaths reported by the government

If agents do not have any information during the course of the epidemic, the equilibrium outcome from the definition in subsection 4.2 follows and the epidemic runs its course based on agents' behavior and potentially misspecified beliefs. Now suppose the government provides some information on current deaths D_t at the beginning of period t. Can agents update their prior about the prevalence of the disease?

If indeed $D_t \neq \tilde{M}_t^{prior}(d)$, then agents update their information about prevalence in order to make their forecast of current deaths match what the government is announcing. In particular, suppose the government provides the following information on deaths $\{D_t, \Delta D_t, \Delta D_{t-1}, \Delta D_{t-2}\}$, that is, an initial death count followed by counts of new deaths. Then using the realization of labor supplies up to $t-1$, $\left\{n(s)^{t-1}, n(i)^{t-1}\right\}$, and the dynamics of the epidemic summarized in (mass susceptible)-(mass death), the prevalence measured as the number of infected at $t-3$ can be back-traced.[21] Note that the number of recovering individuals, C, equals:

$$C_{t-1} = \Delta D_t / \theta\delta$$

$$C_{t-2} = \Delta D_{t-1} / \theta\delta$$

$$C_{t-3} = \Delta D_{t-2} / \theta\delta$$

With these, one can determine the infected individuals, I, as:

$$I_{t-2} = \left(C_{t-1} - C_{t-2} - \theta C_{-2}\right)/\gamma$$

$$I_{t-3} = \left(C_{t-1} - C_{t-2} - \theta C_{-2}\right)/\gamma$$

[21] Given the structure of the model and the state transitions from (mass susceptible)-(mass death), given information of the number of deaths up until t, the best agents can do is infer the prevalence in the number of infected in period $t-3$.

Hence, the probability of an infection at $t-3$ is given by (4) and (5):

$$\Pi_{t-3} = 1 - \exp\left(-\Lambda n\left(i, t-3\right) I_{t-3}\right)$$

Finally, with this probability, one can determine the number of susceptible individuals, S, at $t-3$ as:

$$S_{t-3} = \left(I_{t-2} - I_{t-3} - \gamma I_{t-3}\right) / \left(n\left(s, t-3\right) \Pi_{t-3}\right)$$

With this information, agents can forward on the system (mass susceptible)-(mass death) using $\left\{n\left(s\right)^{t-1}, n\left(i\right)^{t-1}\right\}$ to get an updated set of beliefs for the masses at t given by $\tilde{M}_t^{posterior}\left(j\right)$ for each $j = s, i, c, r, d$, where now $D_t = \tilde{M}_t^{posterior}\left(d\right)$. This new information consists in an unexpected ("MIT") shock to update beliefs that will change behavior plans until the end of the epidemic according to the definition in subsection 4.2: behavior schedules $\left\{n\left(j, t'\right)\right\}_{t'=t}^{\infty}$ are updated in accordance with the new belief $\tilde{M}_t^{posterior}\left(j\right)$ for each state $j = s, i, c, r, d$.

4.5 Delays in collection of deaths

Due to physical constraints in data collection, the government may actually provide biased information about the current level of deaths $D_t \neq M_t\left(d\right)$, such that:

$$D_t = D_{t-1} + f\left(\Delta M_t\left(d\right), \Delta M_{t-2}\left(d\right), ..., \Delta M_1\left(d\right)\right) < M_t\left(d\right)$$

where the function f captures that the government identifies the previous periods' new deaths with delays. Under these conditions, agents update their prior into a wrong posterior, that is, $D_t = \tilde{M}_t^{posterior}\left(d\right) \Rightarrow \tilde{M}_t\left(j\right) \neq M_t\left(j\right)$ for $j = s, i, c, r, d$.

4.6 Simulation of an epidemic with delays in deaths reported

In order to simulate this model, we set the parameters as summarized in Table 5. The discount factor $\beta = 0.98^{1/365}$ is set to capture a 2% annual interest rate. Parameters associated with infectiousness, resolving, and death rates are calibrated in order to target standard findings from the medical literature as documented in Bar-On et al. (2020). The remaining parameters are targeted to closely follow features of the Mexican economy. We assume that the initial population is 120 million and the time zero number of infected are 120 individuals (0.0001% of the total population). We use Mexican time use surveys to calibrate the parameter λ_h by targeting an expenditure of 36% of available hours in activities outside the home before the epidemic outbreak. The utility function parameter b captures a drop in total hours outside the home during the epidemic of 50% as suggested by evidence from Google Mobility data. Lastly, the baseline contagion rate parameter Λ is set to generate a basic reproduction number of two as documented by Marioli et al. (2020).

Table 5:
Baseline Calibration of the Behavioral Contagion Model

Parameter in the model		Value	Target
Discount factor	β	$0.98^{1/365}$	Standard 2% yearly interest rate
Probability of infection	γ	0.166	6 days while infectious (Bar-On et al., 2020)
Resolving probability	θ	0.10	16 days to clear Covid-19 (Bar-On et al., 2020)
Death rate	δ	0.008	From medical literature (Bar-On et al., 2020)
Initial mass of infected	$M_0(i)$	0.0001%	120 individuals in Mexico
Preference for staying home	λ_h	1.77	36% of hours spent outside home (ENUT)
Hours for sick individuals	$\bar{n}$	$0.5n^*$	Sick individuals can spend at most half the time outside
Preference for staying alive	b	6.5	50% drop in outside home activity (Google Mobility Data)
Baseline contagion rate	Λ	5.11	Basic reproduction number $R_0 = 2$ (Marioli et al., 2020)

Notes: This table shows the values for the parameters used to calibrate the model. ENUT refers to the Mexican Time Use Survey for 2014.

To model delays, we capture the distribution of the difference in days between deaths as reported and deaths as occurred provided by the Mexican data describen in Section 2. Specifically, we use the following formula:

$$D_t = D_{t-1} + p_t \Delta M_t(d) + ... + p_{t-60} \Delta M_{t-60}(d)$$

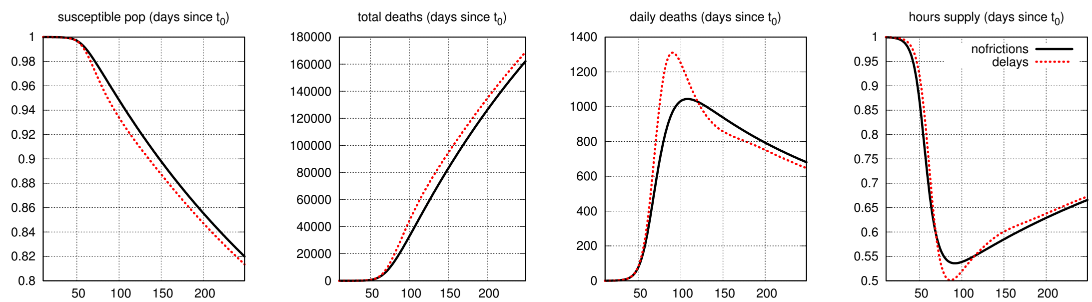
where the coefficients $\{p_t, ..., p_{t-60}\}$ capture the same density as what we observe in a histogram of the data. Taking the data for the entire country, the distribution of delays shows an average of 8.6 days and a standard deviation of 11.5 days (see Figure 1).

In the simulation, we also allow for the introduction of a vaccine, which is expected ex-ante. After it becomes available at time t^{vacc}, the transmissibility immediately disappears $\Lambda = 0$. We set $t^{vacc} = 350$ days. Given the parameterization and the function for reporting delays, we numerically simulate the model following the definition of the equilibrium in subsection 4.2 using value function iteration.

4.7 Results

Figure 3 shows the comparison of the epidemic dynamics in a model without any delays (solid line) relative to one with delays as described above (dashed line).

Figure 3:
Simulation Results of Behavioral Model with and without Delays in Death Reports



Notes: These graphs show the simulation results from the model by computing the equilibrium as defined in subsection 4.2. We show results for a situation without reporting delays (solid line) and with delays calibrated to the Mexican data (dashed line). We show the mass of susceptible individuals, total deaths, daily number of deaths, and hours supplied outside the home.

It is clear that the presence of delays contributes significantly to a faster progression of the epidemic, implying 25% higher daily deaths at the peak, as summarized in Table 6. The slow adjustment of hours when the disease starts spreading contributes to a higher prevalence of Covid-19 at the peak when delays are present, jumping from 0.67% to 0.90% of the whole population. Once the government starts updating the information on the number of deaths, agents realize that the spread of the disease is serious and thus adjust their behavior more abruptly. This causes a collapse of hours outside that is stronger than what would be the case without delays. Relative to the pre-pandemic levels, hours away from home dip up to 46% in the case with no delays, but 50% with delays.

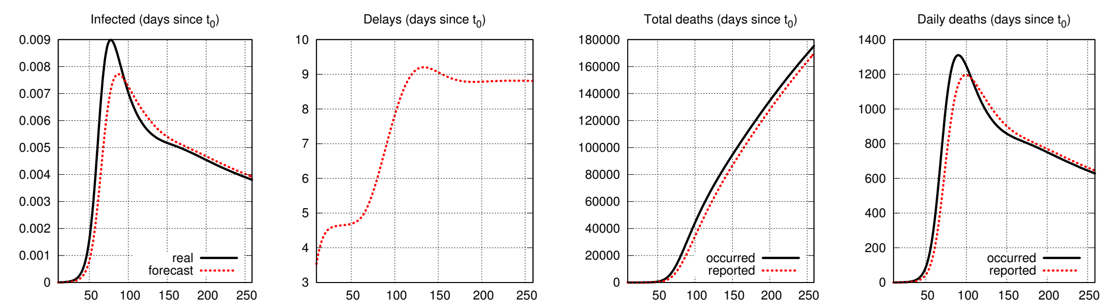
To better grasp the impact of delays on the beliefs about the disease prevalence, Figure 4 plots how deaths are reported and how they occurred, as well as the difference between the true prevalence of infected individuals and the corresponding biased belief. The first graph shows how beliefs about the mass of the infected population are consistently different from the truth. In particular, when the epidemic starts exploding, this leads to an underestimation of the severity of the disease, and therefore a lower behavioral response from agents.

Table 6:
Epidemic Statistics from Model Simulations with and without Delays in Death Reports

	Peak infections (% of pop)	Peak deaths (days)	Maximum deaths (daily)	Total deaths (at day 120)	Total deaths (at day 500)	Hours outside (at trough)
No delays	0.668	107	1,044	53,602	229,910	19.34
With delays	0.899	89	1,310	66,891	233,616	18.09
Reported deaths	-	98	1,196	57,263	233,616	-

Notes: This table shows statistics on the epidemic generated in the model. We show results for a situation without reporting delays (first row) and with delays calibrated to the Mexican data (second row). The last row presents statistics on deaths as reported in the model with delays.

Figure 4:
Beliefs and Truth of Epidemic Dynamics in Simulated Model with Delays in Death Reports

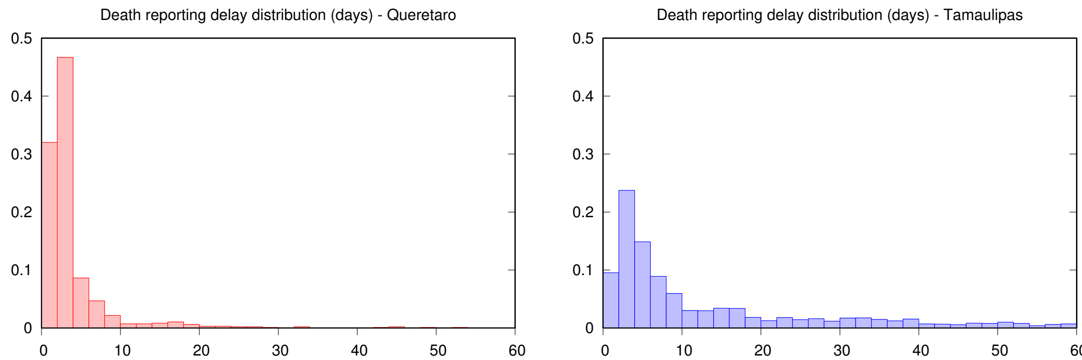


Notes: These graphs show the simulation results from the model. We show results for a situation without reporting delays (solid line) and with delays calibrated to the Mexican data (dashed line). The plot on the left shows the beliefs agents have about the mass of infected individuals over time from the onset of the epidemic. We also show the evolution of delays, and both total and daily deaths as occurred and as reported.

Comparison between state with highest and lowest delays. Aggregate data on delays for Mexico mask large variations at finer geographical levels (Gutierrez et al., 2020). Figure 5 shows the distribution of delays in death reportes in the states of Queretaro and Tamaulipas, representing extreme cases of low and high delays, respectively. Reporting delays in Tamaulipas are on average three times larger than in Queretaro.

To highlight how reporting delays can affect the overall dynamics of the epidemic, we simulate the model under the parameters shown in Table 5, but using a delay distribution as observed in each of these states. Figure 6 summarizes the results of this exercise. The epidemic progresses much faster when delays in reporting are larger, as in Tamaulipas. Agents are slower to react to the progression of the disease, which contributes to a larger peak for prevalence, with a share of 1.06% of population infected relative to 0.73% if delays occur as in Querataro (Table 7). Due to the

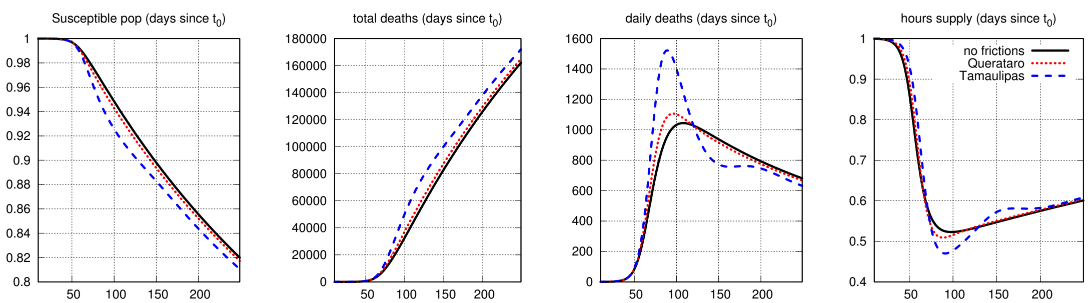
Figure 5:
Distribution of Reporting Delays for Queretaro and Tamaulipas



Notes: These graphs show the distribution of delays in death reports for two states representing extreme cases. Delays are very short in Queretaro (mean=3.4, sd=5.04), while delays are large in Tamaulipas (mean=12.6, sd=14.32).

acceleration in contagion, peak deaths increase to 1,522 in the case of large delays (Tamaulipas), which is 38% more than in the case of small delays (Queretaro).

Figure 6:
Simulation Results of Behavioral Model with and without Delays in
Death Reports: Small vs Large Delays



Notes: These graphs show the simulation results from the model by computing the equilibrium as defined in subsection 4.2. We show results for a situation without reporting delays (solid line), with short delays as in Queretaro (short-dashed line), and long delays as in Tamaulipas (long-dashed line). We show the mass of susceptible individuals, total deaths, daily number of deaths, and hours supplied outside the home.

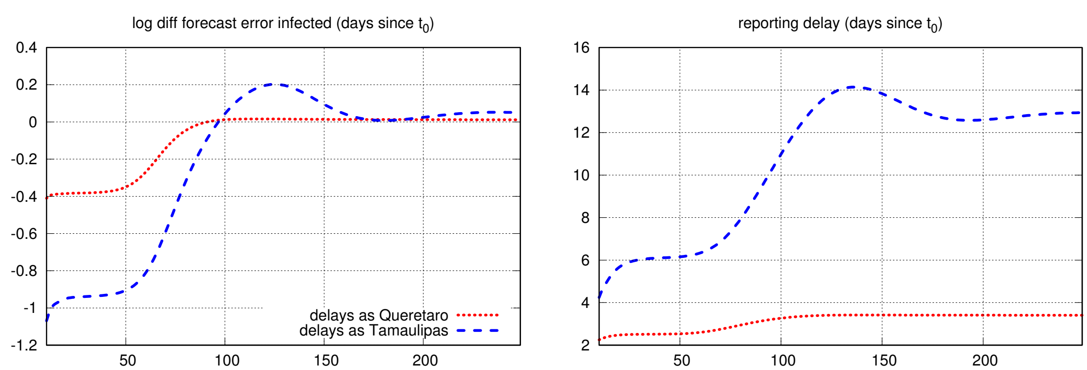
Figure 7 helps explain the difference in epidemic dynamics between both extreme cases. The right panel shows how delays evolve in the two cases, with Queretaro stabilizing at around 3 days while Tamaulipas stabilizes at about 14 days. In the left panel, it is shown that these types of delays imply that, at the beginning of the epidemic, agents would underestimate the prevalence of the infected population by almost a full log point (-63%) in the case of the large delays in Tamaulipas relative to a 0.4 log point (-33%) in the case of shorter delays in Queretaro.

Table 7:
Epidemic Statistics from Model Simulations with and without Delays in Death Reports: Small vs Large Delays

	Peak infections (% of pop)	Peak deaths (days)	Maximum deaths (daily)	Total deaths (at day 120)	Total deaths (at day 500)	Hours outside (at trough)
No delays	0.668	107	1,044	53,602	229,910	19.34
Short delays (Queretaro)						
Delays	0.731	95	1,106	58,760	231,510	18.80
Reported deaths	-	99	1,094	55,256	231,510	-
Long delays (Tamaulipas)						
Delays	1.06	88	1,522	74,407	235,350	17.68
Reported deaths	-	100	1,285	59,053	235,350	-

Notes: This table shows statistics on the epidemic generated in the model. We show results for a situation without reporting delays (first row) and with either short delays (modeled after Queretaro) or long delays (Tamaulipas). The "reported deaths" row presents statistics on deaths as reported in the models with delays.

Figure 7:
Forecast Error of Infectiousness and Reporting Delays in Simulated Model: Small vs Large Delays



Notes: These graphs show the simulation results from the model, computing the forecast error of the mass of infected individuals and the reporting delays in models with delays. We show results for a situation with short delays as in Queretaro (short-dashed line) and long delays as in Tamaulipas (long-dashed line).

5 Conclusion

Providing individuals with information about the extent of an epidemic may be an effective and low-cost intervention that induces compliance with non-pharmaceutical mitigating behaviors. However, the quality of the available information may hinder authorities' ability to effectively induce changes in behavior through information provision.

To shed light on how the accuracy of the available information may impose challenges for the containment of epidemics, we analyze how individual beliefs and behavior are affected by differing information due to lags in reporting of Covid-19 deaths in Mexico. We first showed that the growth in Covid-19 cases is more responsive to reported – rather than occurred – deaths. We then presented results from an online survey where participants that were shown total deaths over time by date reported – that is, a measure that understates the true death toll – were more likely to perceive a lower risk of contagion and to report lower intentions of complying with stay-at-home recommendations. Finally, we developed an equilibrium behavioral model to show that, if individuals receive lagged information because of reporting delays, they are slower to modify their risky behavior, which in turn leads to more severe epidemic outcomes.

Delays in death reports are a common feature across settings, but are likely to be exacerbated by the low state capacity in low- and middle-income countries. Hence, our results suggest that data collection issues in these contexts may magnify the extent of the epidemic, adding to the particular challenges facing these countries. Furthermore, other issues with information accuracy linked to differential counting of tests, cases, and deaths across and within countries may also affect individual behavior via their effect on perceptions and beliefs, which in turn may limit effective management of the Covid-19 pandemic.

From a policy perspective, our results highlight the importance of collecting and disseminating reliable real-time information on the state of the epidemic, or at least being upfront and clear about the shortcomings of the available data. While government resources may be better spent on other palliative measures, our findings suggest that simple, low-cost interventions – such as using statistical techniques to correct for delays – may improve outcomes during a pandemic.

References

AbouZahr, C., D. De Savigny, L. Mikkelsen, P. W. Setel, R. Lozano, and A. D. Lopez (2015). Towards universal civil registration and vital statistics systems: the time is now. *The Lancet 386*(10001), 1407–1418.

Acemoglu, D., V. Chernozhukov, I. Werning, and M. D. Whinston (2020). A multi-risk SIR model with optimally targeted lockdown. Technical report, National Bureau of Economic Research.

Akesson, J., S. Ashworth-Hayes, R. Hahn, R. D. Metcalfe, and I. Rasooly (2020). Fatalism, beliefs, and behaviors during the COVID-19 pandemic. Technical report, National Bureau of Economic Research.

Alexander, D. and E. Karger (2020). Do stay-at-home orders cause people to stay at home? Effects of stay-at-home orders on consumer behavior.

Allcott, H., L. Boxell, J. Conway, M. Gentzkow, M. Thaler, and D. Y. Yang (2020). Polarization and public health: Partisan differences in social distancing during the Coronavirus pandemic. *NBER Working Paper* (w26946).

Alvarez, F. E., D. Argente, and F. Lippi (2020). A simple planning problem for covid-19 lockdown. *Covid Economics 1*(14), 1–32.

Baccini, L. and A. Brodeur (2020). Explaining governors' response to the COVID-19 pandemic in the United States.

Bar-On, Y. M., A. Flamholz, R. Phillips, and R. Milo (2020). Science forum: SARS-CoV-2 (COVID-19) by the numbers. *Elife 9*, e57309.

Bargain, O. and U. Aminjonov (2020). Trust and compliance to public health policies in times of COVID-19.

Barrios, J. M. and Y. Hochberg (2020). Risk perception through the lens of politics in the time of the covid-19 pandemic. Technical report, National Bureau of Economic Research.

Binder, C. (2020). Coronavirus fears and macroeconomic expectations. *Review of Economics and Statistics*, 1–27.

Bird, S. M. (2015). End late registration of fact-of-death in England and Wales. *The Lancet 385*(9980), 1830–1831.

Briscese, G., N. Lacetera, M. Macis, and M. Tonin (2020). Compliance with covid-19 social-distancing measures in Italy: the role of expectations and duration. Technical report, National Bureau of Economic Research.

Brodeur, A., D. M. Gray, A. Islam, S. J. Bhuiyan, et al. (2020). A literature review of the economics of COVID-19. Technical report, Institute of Labor Economics (IZA).

Brodeur, A., I. Grigoryeva, and L. Kattan (2020). Stay-at-home orders, social distancing and trust.

Brotherhood, L., P. Kircher, C. Santos, and M. Tertilt (2020). An economic model of the Covid-19 epidemic: The importance of testing and age-specific policies.

Bursztyn, L., A. Rao, C. Roth, and D. Yanagizawa-Drott (2020). Misinformation during a pandemic. *University of Chicago, Becker Friedman Institute for Economics Working Paper* (2020-44).

Coibion, O., Y. Gorodnichenko, and M. Weber (2020). Does policy communication during COVID work? *Covid Economics 1*(29), 1–49.

Dave, D. M., A. I. Friedson, K. Matsuzawa, J. J. Sabia, and S. Safford (2020). Were urban cowboys enough to control COVID-19? Local shelter-in-place orders and coronavirus case growth. Technical report, National Bureau of Economic Research.

Ding, W., R. Levine, C. Lin, and W. Xie (2020). Social distancing and social capital: Why US counties respond differently to COVID-19. *Available at SSRN 3624495*.

Dupas, P. (2011). Do teenagers respond to HIV risk information? Evidence from a field experiment in Kenya. *American Economic Journal: Applied Economics 3*(1), 1–34.

Dupas, P., E. Huillery, and J. Seban (2018). Risk information, risk salience, and adolescent sexual behavior: Experimental evidence from Cameroon. *Journal of Economic Behavior & Organization 145*, 151–175.

Durante, R., G. Gulino, et al. (2020). Asocial capital: Civic culture and social distancing during COVID-19.

Eichenbaum, M. S., S. Rebelo, and M. Trabandt (2020a). The macroeconomics of epidemics. Technical report, National Bureau of Economic Research.

Eichenbaum, M. S., S. Rebelo, and M. Trabandt (2020b). The macroeconomics of testing and quarantining. Technical report, National Bureau of Economic Research.

Fernández-Villaverde, J. and C. I. Jones (2020). Estimating and simulating a SIRD model of COVID-19 for many countries, states, and cities. Technical report, National Bureau of Economic Research.

Gallego, J. A., M. Prem, and J. F. Vargas (2020). Corruption in the times of pandemia. *Available at SSRN 3600572*.

Gottlieb, C., J. Grobovšek, and M. Poschke (2020). Working from home across countries. *Covid Economics 1*(8), 71–91.

Greenwood, J., P. Kircher, C. Santos, and M. Tertilt (2019). An equilibrium model of the African HIV/AIDS epidemic. *Econometrica 87*(4), 1081–1113.

Gupta, S., T. D. Nguyen, F. L. Rojas, S. Raman, B. Lee, A. Bento, K. I. Simon, and C. Wing (2020). Tracking public and private response to the COVID-19 epidemic: Evidence from state and local government actions. Technical report, National Bureau of Economic Research.

Gutierrez, E. and A. Rubli (2020). Shocks to hospital occupancy and mortality: Evidence from the 2009 H1N1 pandemic. Technical report, Working paper.

Gutierrez, E., A. Rubli, and T. Tavares (2020). Delays in death reports and their implications for tracking the evolution of COVID-19. *Covid Economics 1*(34), 116–144.

Hsiang, S., D. Allen, S. Annan-Phan, K. Bell, I. Bolliger, T. Chong, H. Druckenmiller, A. Hultgren, L. Y. Huang, E. Krasovich, et al. (2020). The effect of large-scale anti-contagion policies on the coronavirus (covid-19) pandemic. *MedRxiv*.

Jinjarak, Y., R. Ahmed, S. Nair-Desai, W. Xin, and J. Aizenman (2020). Accounting for global COVID-19 diffusion patterns, January-April 2020. Technical report, National Bureau of Economic Research.

Juranek, S. and F. Zoutman (2020). The effect of social distancing measures on the demand for intensive care: Evidence on covid-19 in Scandinavia.

Kaplan, G., B. Moll, and G. Violante (2020). Pandemics according to HANK. *Powerpoint presentation, LSE 31*.

Knittel, C. R. and B. Ozaltun (2020). What does and does not correlate with COVID-19 death rates. *medRxiv*.

Loayza, N. V. (2020). Costs and trade-offs in the fight against the Covid-19 pandemic: A developing country perspective.

Marioli, F. A., F. Bullano, S. Kučinskas, and C. Rondón-Moreno (2020). Tracking R of COVID-19: A new real-time estimation using the Kalman filter. *medRxiv*.

Monroy-Gómez-Franco, L. (2020). ¿Quién puede trabajar desde casa? Evidencia desde México.

Nyhan, B. and J. Reifler (2015). Does correcting myths about the flu vaccine work? An experimental evaluation of the effects of corrective information. *Vaccine 33*(3), 459–464.

Oster, E. (2012). HIV and sexual behavior change: Why not Africa? *Journal of health economics 31*(1), 35–49.

Papageorge, N. W., M. V. Zahn, M. Belot, E. van den Broek-Altenburg, S. Choi, J. C. Jamison, E. Tripodi, et al. (2020). Socio-demographic factors associated with self-protecting behavior during the COVID-19 pandemic. Technical report, Institute of Labor Economics (IZA).

Ribeiro, F. and A. Leist (2020). Who is going to pay the price of Covid-19? Reflections about an unequal Brazil. *International Journal for Equity in Health 19*, 1–3.

Sadaf, A., J. L. Richards, J. Glanz, D. A. Salmon, and S. B. Omer (2013). A systematic review of interventions for reducing parental vaccine refusal and vaccine hesitancy. *Vaccine 31*(40), 4293–4304.

Simonov, A., S. K. Sacher, J.-P. H. Dubé, and S. Biswas (2020). The persuasive effect of fox news: non-compliance with social distancing during the covid-19 pandemic. Technical report, National Bureau of Economic Research.

Walker, P. G., C. Whittaker, O. J. Watson, M. Baguelin, P. Winskill, A. Hamlet, B. A. Djafaara, Z. Cucunubá, D. O. Mesa, W. Green, et al. (2020). The impact of Covid-19 and strategies for mitigation and suppression in low-and middle-income countries. *Science*.

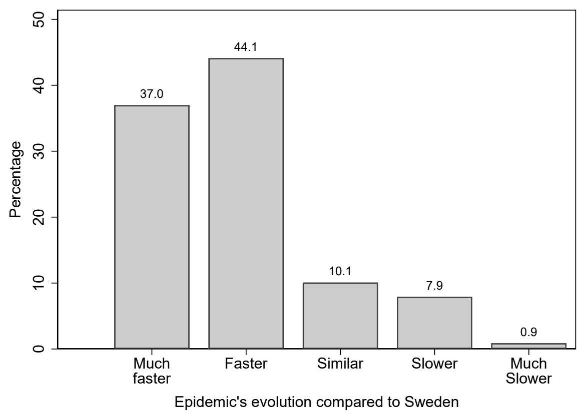
WHO (2013). Pandemic influenza risk management: WHO interim guidance.

WHO (2020). Coronavirus disease 2019 (Covid-19): situation report, 88.

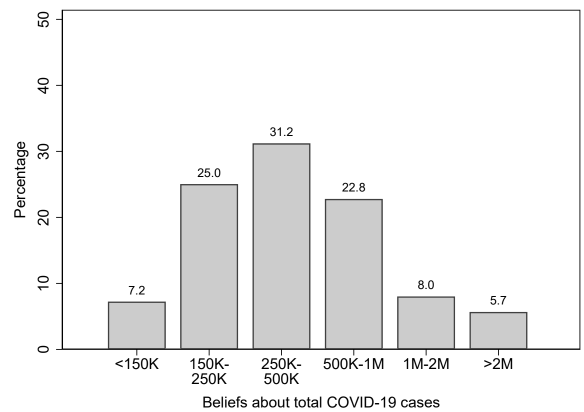
Appendix for Online Publication

A Supplementary Figures and Tables for the Survey

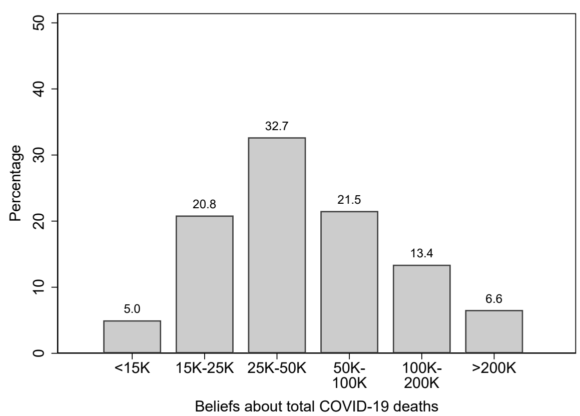
Figure A1:
Histograms of Risk Perceptions and Behavior



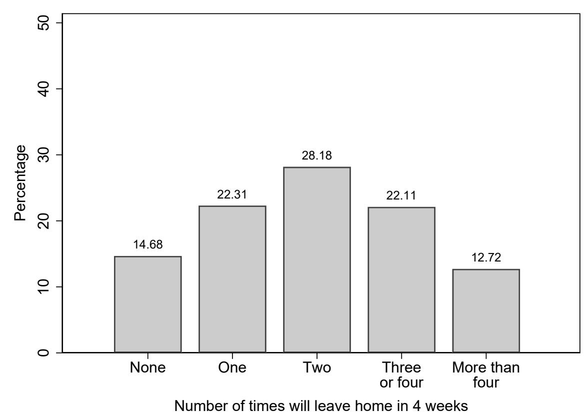
(a) Comparison with Sweden



(b) Expected total cases



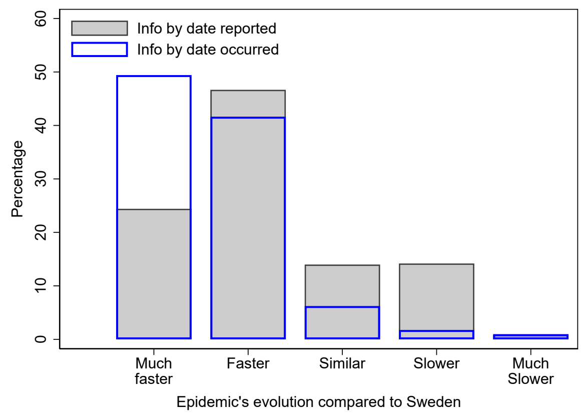
(c) Expected total deaths



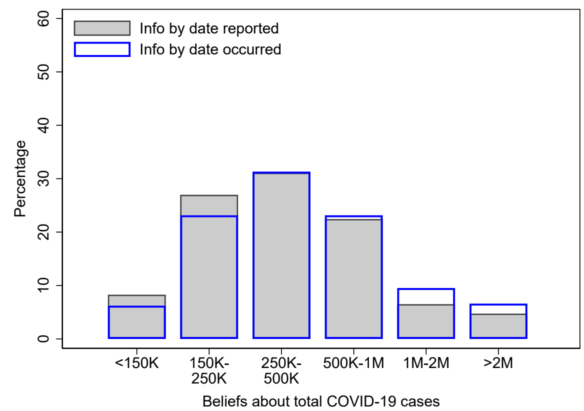
(d) Times leaving the house in 4 weeks

Notes: These graphs show histograms for the raw questions that make up our six main outcome variables related to perceptions and expected behavior elicited in the survey for our full sample of participants. Each plot shows the percentage of total respondents that chose each of the answers.

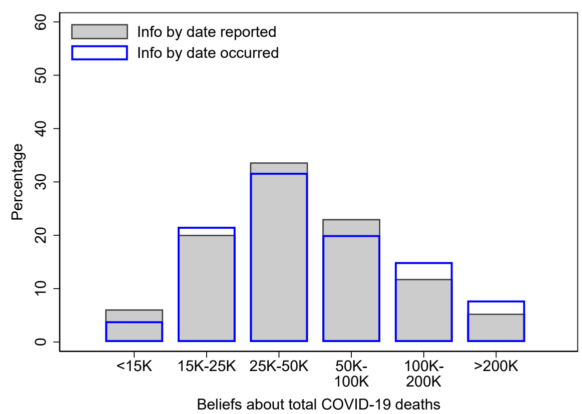
Figure A2:
Histograms of Risk Perceptions and Behavior by Informational Treatments



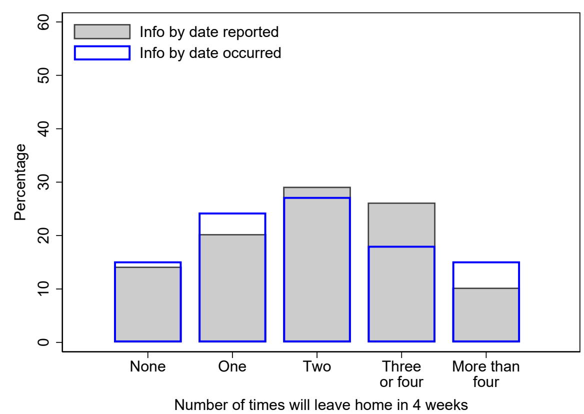
(a) Comparison with Sweden



(b) Expected total cases



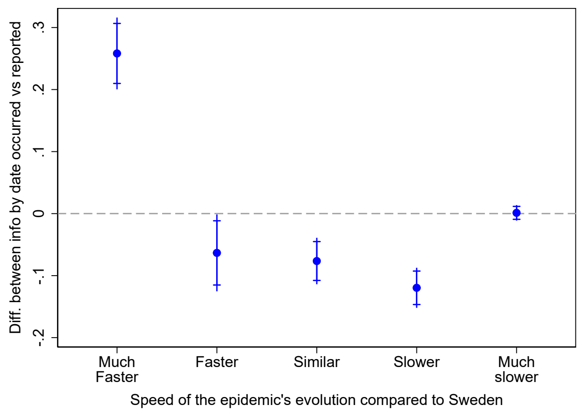
(c) Expected total deaths



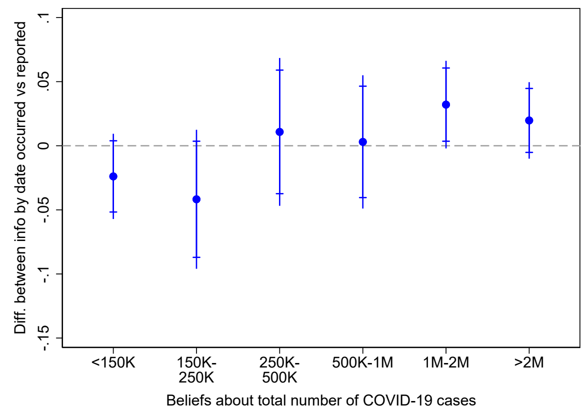
(d) Times leaving the house in 4 weeks

Notes: These graphs show histograms for the raw questions that make up our six main outcome variables related to perceptions and expected behavior elicited in the survey for our full sample of participants. We distinguish between the two informational treatments. Each plot shows the percentage of total respondents that chose each of the answers.

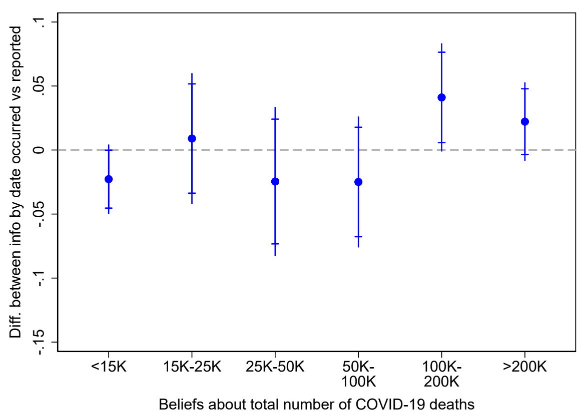
Figure A3:
Estimates of Informational Treatments for Full Set of Responses



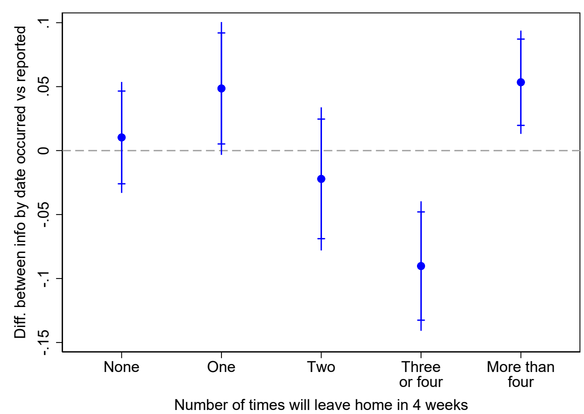
(a) Comparison with Sweden



(b) Expected total cases



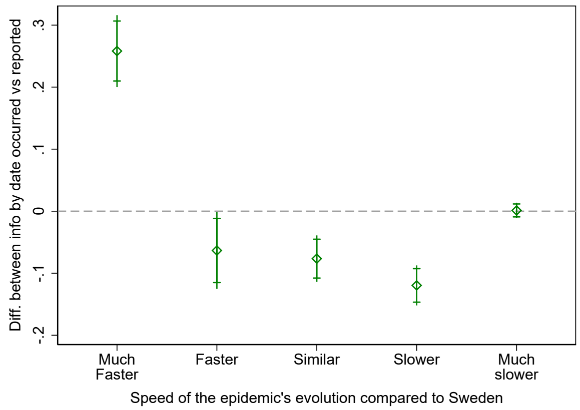
(c) Expected total deaths



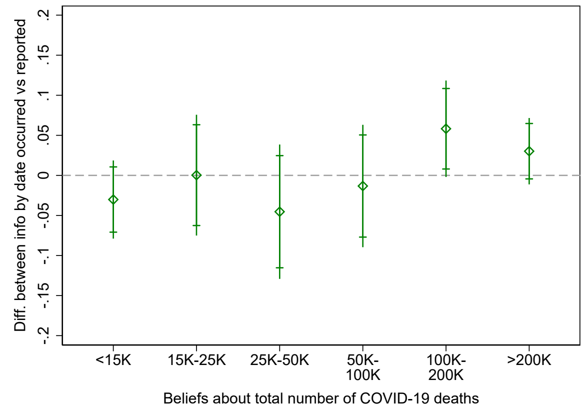
(d) Times leaving the house in 4 weeks

Notes: These graphs show estimates of the difference by informational treatment on the raw questions that make up our six main outcome variables related to perceptions and expected behavior elicited in the survey for our full sample of participants. Each plot shows coefficients from multiple regressions with indicators for each response as the outcome variable. Coefficients correspond to the average difference between respondents that received information based on the actual date of death relative to those that received information based on date of reports. Vertical bars show 95 and 90% confidence intervals.

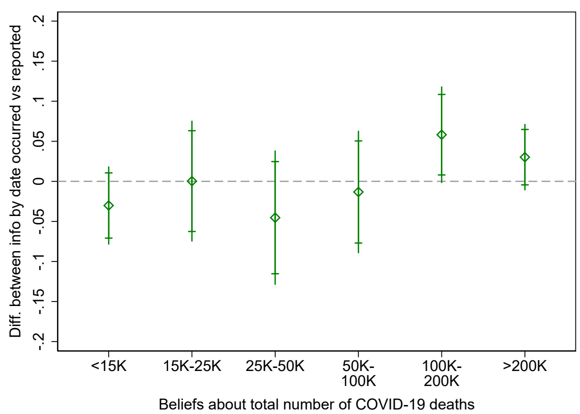
Figure A4:
Estimates of Informational Treatments for Full Set of Responses: Low Prior Sample



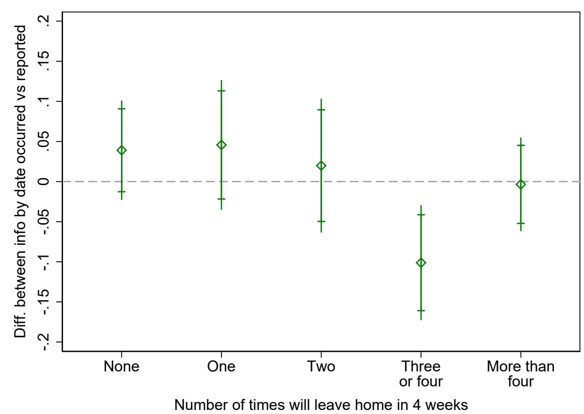
(a) Comparison with Sweden



(b) Expected total cases



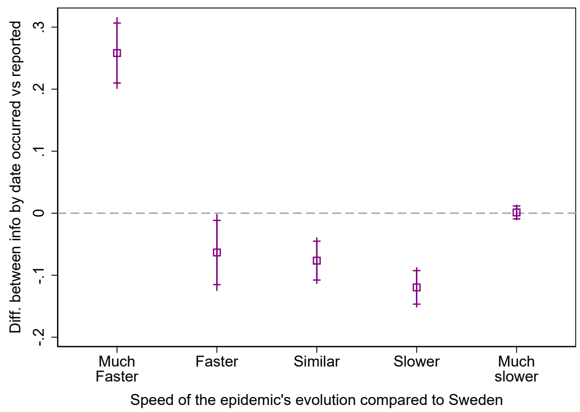
(c) Expected total deaths



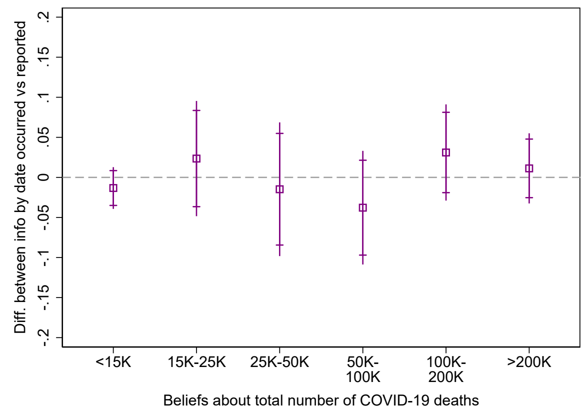
(d) Times leaving the house in 4 weeks

Notes: These graphs show estimates of the difference by informational treatment on the raw questions that make up our six main outcome variables related to perceptions and expected behavior elicited in the survey for the sample of participants with a low prior of total Covid-19 cases as of May 20. Each plot shows coefficients from multiple regressions with indicators for each response as the outcome variable. Coefficients correspond to the average difference between respondents that received information based on the actual date of death relative to those that received information based on date of reports. Vertical bars show 95 and 90% confidence intervals.

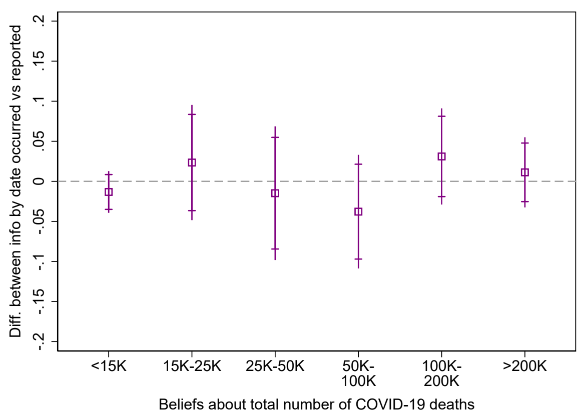
Figure A5:
Estimates of Informational Treatments for Full Set of Responses: High Prior Sample



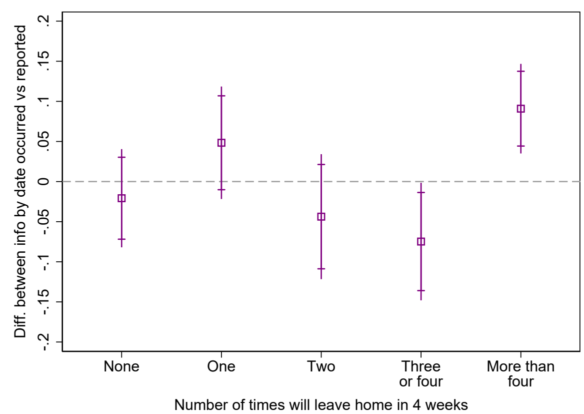
(a) Comparison with Sweden



(b) Expected total cases



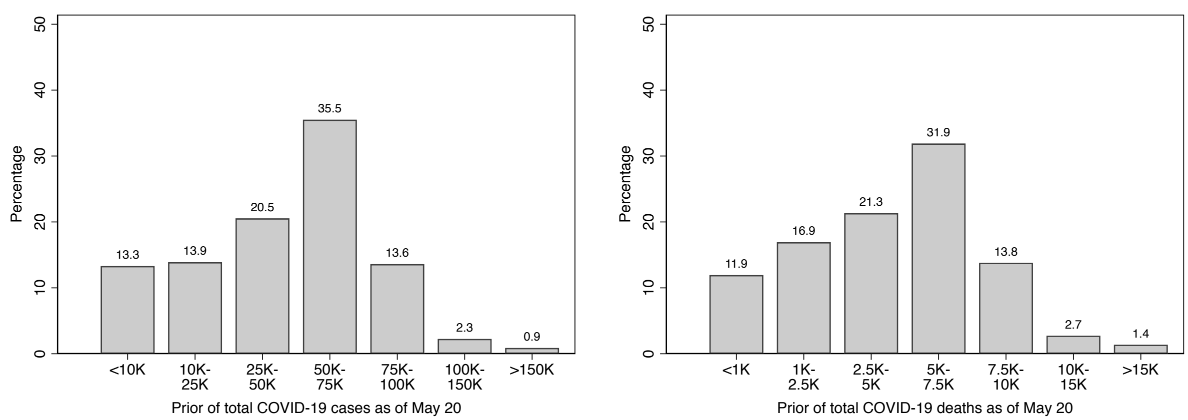
(c) Expected total deaths



(d) Times leaving the house in 4 weeks

Notes: These graphs show estimates of the difference by informational treatment on the raw questions that make up our six main outcome variables related to perceptions and expected behavior elicited in the survey for the sample of participants with a high prior of total Covid-19 cases as of May 20. Each plot shows coefficients from multiple regressions with indicators for each response as the outcome variable. Coefficients correspond to the average difference between respondents that received information based on the actual date of death relative to those that received information based on date of reports. Vertical bars show 95 and 90% confidence intervals.

Figure A6:
Histograms of Prior Beliefs on Total Cases and Deaths



(a) Beliefs on total cases

(b) Beliefs on total deaths

Notes: These graphs show histograms for the questions eliciting beliefs about total cases and total deaths up to May 20 (one week prior to when the survey was launched) for our full sample of participants. Each plot shows the percentage of total respondents that chose each of the answers. The actual number of cumulative cases reported by the government on May 20 was 56,594, and the cumulative deaths reported were 6,090 (see https://twitter.com/HLGatell/status/1263264663283908609?s=20, last accessed June 29, 2020).

Table A1:
Balance Table for Survey Covariates: Low Prior Sample

	Informational treatments		
	Deaths by date reported	Deaths by date occurred	Difference in means
Female	0.475	0.500	0.025
	(0.500)	(0.501)	(0.045)
Ages 18-22	0.324	0.412	0.088**
	(0.469)	(0.493)	(0.044)
Ages 23-29	0.252	0.232	-0.020
	(0.435)	(0.423)	(0.039)
Ages 30-49	0.248	0.192	-0.056
	(0.433)	(0.395)	(0.037)
Ages 50+	0.176	0.164	-0.012
	(0.382)	(0.371)	(0.034)
Works	0.420	0.344	-0.076*
	(0.495)	(0.476)	(0.044)
Attends school	0.353	0.432	0.079*
	(0.479)	(0.496)	(0.044)
Works and attends school	0.155	0.152	-0.003
	(0.363)	(0.360)	(0.033)
Other occupation/employment status	0.071	0.072	0.001
	(0.258)	(0.259)	(0.023)
Lives in Mexico City	0.773	0.764	-0.009
	(0.420)	(0.425)	(0.038)
Lives in apartment	0.340	0.396	0.056
	(0.475)	(0.490)	(0.044)
Lives in house, no yard	0.118	0.108	-0.010
	(0.323)	(0.311)	(0.029)
Lives in house with yard	0.542	0.496	-0.046
	(0.499)	(0.501)	(0.045)
Household size: 1-2	0.223	0.240	0.017
	(0.417)	(0.428)	(0.038)
Household size: 3	0.231	0.236	0.005
	(0.422)	(0.425)	(0.038)
Household size: 4	0.214	0.224	0.010
	(0.411)	(0.418)	(0.038)
Household size: 5+	0.546	0.524	-0.022
	(0.499)	(0.500)	(0.045)
Has HH members over 70 years old	0.181	0.076	-0.105***
	(0.386)	(0.266)	(0.030)
Has HH members 60-70 years old	0.206	0.228	0.022
	(0.405)	(0.420)	(0.037)
Has HH members 50-60 years old	0.496	0.432	-0.064
	(0.501)	(0.496)	(0.045)
Does not seek healthcare when sick	0.130	0.108	-0.022
	(0.337)	(0.311)	(0.029)
Self-medicates when sick	0.357	0.396	0.039
	(0.480)	(0.490)	(0.044)
Observations	238	250	488

Notes: This table shows means and standard deviations for a series of covariates asked in the survey before the informational treatment was presented for the sample of participants with a low prior of total Covid-19 cases as of May 20. We show statistics separately for each informational treatment, as well as the difference in the means. Stars denote significance from a difference in means test.
*** p<0.01, ** p<0.05, * p<0.1

Table A2:
Balance Table for Survey Covariates: High Prior Sample

	Informational treatments		
	Deaths by date reported	Deaths by date occurred	Difference in means
Female	0.515	0.481	-0.034
	(0.501)	(0.501)	(0.043)
Ages 18-22	0.319	0.356	0.038
	(0.467)	(0.480)	(0.041)
Ages 23-29	0.293	0.273	-0.020
	(0.456)	(0.446)	(0.039)
Ages 30-49	0.215	0.239	0.024
	(0.411)	(0.427)	(0.036)
Ages 50+	0.174	0.133	-0.041
	(0.380)	(0.340)	(0.031)
Works	0.400	0.314	-0.086**
	(0.491)	(0.465)	(0.041)
Attends school	0.381	0.402	0.020
	(0.487)	(0.491)	(0.042)
Works and attends school	0.159	0.163	0.004
	(0.367)	(0.370)	(0.032)
Other occupation/employment status	0.059	0.121	0.062**
	(0.237)	(0.327)	(0.025)
Lives in Mexico City	0.778	0.742	-0.035
	(0.417)	(0.438)	(0.037)
Lives in apartment	0.344	0.375	0.031
	(0.476)	(0.485)	(0.042)
Lives in house, no yard	0.130	0.125	-0.005
	(0.337)	(0.331)	(0.029)
Lives in house with yard	0.526	0.500	-0.026
	(0.500)	(0.501)	(0.043)
Household size: 1-2	0.241	0.261	0.021
	(0.428)	(0.440)	(0.038)
Household size: 3	0.185	0.254	0.069*
	(0.389)	(0.436)	(0.036)
Household size: 4	0.285	0.227	-0.058
	(0.452)	(0.420)	(0.038)
Household size: 5+	0.574	0.485	-0.089**
	(0.495)	(0.501)	(0.043)
Has HH members over 70 years old	0.141	0.083	-0.057**
	(0.348)	(0.277)	(0.027)
Has HH members 60-70 years old	0.222	0.178	-0.044
	(0.417)	(0.383)	(0.035)
Has HH members 50-60 years old	0.430	0.508	0.078*
	(0.496)	(0.501)	(0.043)
Does not seek healthcare when sick	0.148	0.197	0.049
	(0.356)	(0.398)	(0.033)
Self-medicates when sick	0.411	0.367	-0.044
	(0.493)	(0.483)	(0.042)
Observations	270	264	534

Notes: This table shows means and standard deviations for a series of covariates asked in the survey before the informational treatment was presented for the sample of participants with a high prior of total Covid-19 cases as of May 20. We show statistics separately for each informational treatment, as well as the difference in the means. Stars denote significance from a difference in means test.
*** p<0.01, ** p<0.05, * p<0.1

B Survey Text in English

This is an anonymous online survey that is being conducted for an academic project aimed at better understanding the public's perceptions about the evolution of the Covid-19 pandemic in Mexico. Responding to the survey takes approximately 10 minutes. We ask you to please answer to all the questions if you choose to participate. Despite the fact that you received an invitation to participate in this survey via email or social media, the dataset where the information you provide will be stored does not collect any type of personal information (such as your name, phone number or IP address). We take all the relevant measures to safeguard your identity. Clicking on the "accept" button below you certify that you are over 18 years of age, and that you agree to respond to all the questions asked. The information you provide will only be used for academic purposes and statistical analyses, never revealing individual-level responses.

Sociodemographic Questions

Sex: Male / Female / Other or Prefer not to say

What is your age?: 18-22 / 23-29 / 30-39 / 40-49 / 50-59 / 60-69 / 70-79 / 80 or older

The highest schooling degree you have obtained is: Elementary school / Secondary school / Highschool / Undergraduate degree / Graduate degree

Occupation: Works / Attends school / Works and attends school / Unemployed / House work / Retired

Where do you live?: CDMX or its suburbs / Aguascalientes / Baja California / Baja California Sur / Campeche / Coahuila / Colima / Chiapas / Chihuahua / Durango / Guanajuato / Guerrero / Hidalgo / Jalisco / EdoMex outside CDMX metro area / Michoacan / Morelos / Nayarit / Nuevo Leon / Oaxaca / Puebla / Queretaro / Quintana Roo / San Luis Potosí / Sinaloa / Sonora / Tabasco / Tamaulipas / Tlaxcala / Veracruz / Yucatan / Zacatecas

How would you describe the house you live in: Apartment / House with yard / House without yard

Do you have internet access at home (Wi-Fi)?: Yes / No

Do you have access to a computer at home?: Yes, but I share it with others / Yes, and I am the only user / No

Apart from you, how many people live in your home?: 1 / 2 / 3 / 4 / 5 or more

Is anyone in your household aged more than 70?: Yes / No

Is anyone in your household aged between 60 and 70?: Yes / No

Is anyone in your household aged between 50 and 60?: Yes / No

What is your household's approximate monthly income?: 0-4,000 pesos / 4000-10,000 pesos / 10,000-20,000 pesos / 20,000-30,000 pesos / 30,000-40,000 pesos / 40,000-50,000 pesos / 50,000-75,000 pesos / 75,000-100,000 pesos / more than 100,000 pesos

Do you have access to private health insurance?: Yes / No

Do you have access to health services from IMSS, ISSSTE, PEMEX, SEDENA or SEMAR?: Yes / No

Do you have access to health services from INSABI or Seguro Popular?: Yes / No

When you fall sick, what do you usually do?: Nothing / Take OTCs / Go to a pharmacy-adjacent doctor's office / Go to a doctor's appointment in the private sector / Go to a doctor's appointment in the public sector / Use the medical services at my office or university

Who did you vote for in the last presidential election?: Andres Manuel Lopez Obrador / Ricardo Anaya / Jose Antonio Meade / Other candidate / No vote

What is your opinion about Andres Manuel Lopez Obrador's government's performance?: Completely approve / Approve / Disapprove / Completely disapprove

Covid-19 related questions

How often do you watch the press conference that Dr. Hugo Lopez-Gatell holds daily at 7pm?: Every day / Several times a week / Once a week / Sporadically / Never

How trustworthy do you think is the information about the evolution of Covid-19 shared by Mexican authorities during the daily 7pm press conference?: Very trustworthy / Somewhat trustworthy / Somewhat untrustworthy / Very untrustworthy

Have you received information regarding the evolution of Covid-19 through Facebook?: Yes / No

How trustworthy do you think is the information about the evolution of Covid-19 shared through Facebook?: Very trustworthy / Somewhat trustworthy / Somewhat untrustworthy / Very untrustworthy

Have you received information regarding the evolution of Covid-19 through Twitter?: Yes / No

How trustworthy do you think is the information about the evolution of Covid-19 shared through Twitter?: Very trustworthy / Somewhat trustworthy / Somewhat untrustworthy / Very untrustworthy

Have you received information regarding the evolution of Covid-19 through Whatsapp?: Yes / No

How trustworthy do you think is the information about the evolution of Covid-19 shared through Whatsapp?: Very trustworthy / Somewhat trustworthy / Somewhat untrustworthy / Very untrustworthy

Think of may 20th. According to you, approximately how many Covid-19 cases had be reported by that date?: Less than 10,000 / Between 10,000 and 25,000 / Between 25,000 and 50,000 / Between 50,000 and 75,000 / Between 75,000 and 100,000 / Between 100,000 and 150,000 / More than 150,000

Think of may 20th. According to you, approximately how many Covid-19 deaths had be reported by that date?: Less than 1,000 / Between 1,000 and 2,500 / Between 2,500 and 5,000 / Between 5,000 and 7,500 / Between 7,500 and 10,000 / Between 10,000 and 15,000 / More than 15,000

What is your opinion about the president's actions in face of the Covid-19 pandemic?: Completely approve / Approve / Disapprove / Completely disapprove

How many times did you leave home last week?: You did not leave home / Once / Twice / Three or four times / More than four times

Information treatments

Cumulative deaths by date reported: The following graph compares the evolution of total Covid-19 related deaths in Mexico and Sweden, from march 22nd to may 15th. The information is presented according to the date on which deaths were reported.

Cumulative deaths by date occurred: The following graph compares the evolution of total Covid-19 related deaths in Mexico and Sweden, from march 22nd to may 15th. For Sweden, rhe information is presented according to the date on which deaths were reported. For Mexico, according to the date on which deaths occurred.

Post-treatment questions

Dr. Hugo Lopez-Gatell has said that the evolution of the pandemic in Mexico is similar to the one experienced by Sweden. In your opinion, the Covid-19 pandemic in Mexico is evolving: Much faster than in Sweden / Faster than in Sweden / Similar to Sweden / Slower than in Sweden / Much slower than in Sweden

What is your opinion about Dr. Hugo Lopez-Gatell and other Mexican health authorities' strategy in face of Covid-19?: Completely approve / Approve / Disapprove / Completely disapprove

When do you expect that Mexico will reach 150,000 total confirmed Covid-19 cases?: Early June / Mid June / Late June / Early July / Mid July / Late July (or later) / There will be less than 150,000 total cases

When do you expect we will reach the maximum number of daily Covid-19 cases in Mexico?: Early June / Mid June / Late June / Early July / Mid July / Late July (or later)

How many cases of Covid-19 do you think will have been confirmed in Mexico by the end of this epidemic outbreak?: Less than 100,000 cases / Between 100,000 and 150,000 cases / Between 150,000 and 250,000 cases / Between 250,000 and 500,000 cases / Between 500,000 and one million cases / Between one and two million cases / More than two million cases

When do you expect that Mexico will reach 15,000 total confirmed Covid-19 deaths?: Early June / Mid June / Late June / Early July / Mid July / Late July (or later) / There will be less than 15,000 deaths

When do you expect we will reach the maximum number of daily Covid-19 deaths in Mexico?: Early June / Mid June / Late June / Early July / Mid July / Late July (or later)

How many deaths due to Covid-19 do you think will have been confirmed in Mexico by the end of this epidemic outbreak?: Less than 10,000 deaths / Between 10,000 and 15,000 deaths / Between 15,000 and 25,000 deaths / Between 25,000 and 50,000 deaths / Between 50,000 and 100,000 deaths / Between 100,000 and 200,000 deaths / More than 200,000 deaths

Imagine an extremely optimistic scenario (which would only happen with a probability lower than 10 percent). In such scenario, the total number of Covid-19 deaths in Mexico would be : Less than 3,000 deaths / Between 3,000 and 6,000 deaths / Between 6,000 and 9,000 deaths / Between 9,000 and 12,000 deaths / Between 12,000 and 15,000 deaths / Between 15,000 and 18,000 deaths / Between 18,000 and 21,000 deaths / Between 21,000 and 30,000 deaths / Between 30,000 and 50,000 deaths / Between 50,000 and 80,000 deaths / Between 80,000 and 120,000 deaths / More than 120,000 deaths

Imagine an extremely pessimistic scenario (which would only happen with a probability lower than 10 percent). In such scenario, the total number of Covid-19 deaths in Mexico would be : Less than 12,000 deaths / Between 12,000 and 15,000 deaths / Between 15,000 and 18,000 deaths / Between 18,000 and 21,000 deaths / Between 21,000 and 30,000 deaths / Between 30,000 and 50,000 deaths / Between 50,000 and 80,000 deaths / Between 80,000 and 120,000 deaths / Between 120,000 and 180,000 deaths / Between 180,000 and 250,000 deaths / Between 250,000 and 500,000 deaths / More than 500,000 deaths

When do you think that Mexico City will stop being under the maximum alter level due to Covid-19?: Early June / Mid June / Late June / Early July / Mid July / Late July / Early August / Mid August / Late August / September or later

Next week, how many times do you expect to leave home?: Will not leave home / Once / Twice / Three or four times / More than four times

If next week you had to attend a social gathering with 100 people, how high do you think the risk of being infected with the virus would be?: Very high risk / High risk / Moderately high risk / Moderately low risk / Low risk / Very low risk

In four weeks, how many times do you expect to leave home?: Will not leave home / Once / Twice / Three or four times / More than four times

If in four weeks you had to attend a social gathering with 100 people, how high do you think the risk of being infected with the virus would be?: Very high risk / High risk / Moderately high risk / Moderately low risk / Low risk / Very low risk

Do you think that most private universities in Mexico will be back on campus in August?: Yes, everything will go back to normal / Yes, but some courses will still be online / No, all courses will be online next semester

If the 2018 presidential election were today (with the same candidates), who would you vote for?: Andres Manuel Lopez Obrador / Ricardo Anaya / Jose Antonio Meade / Other candidate / Would not vote

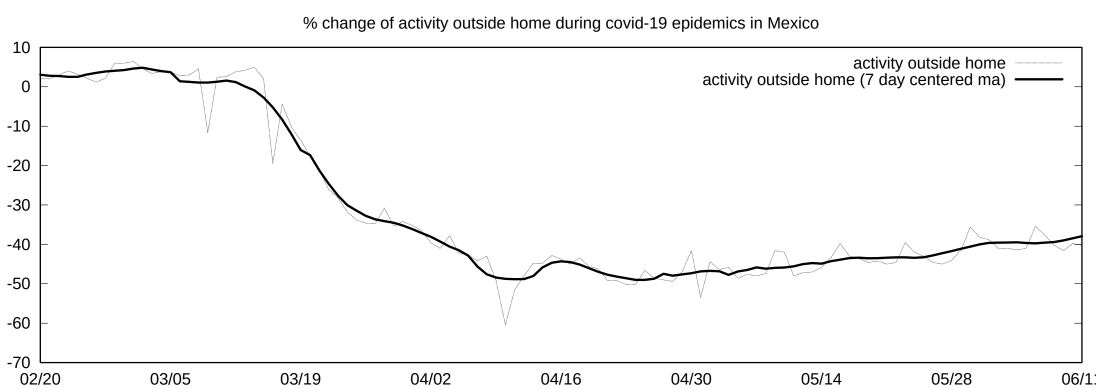
C Additional Details and Results on the Model

C.1 Details of model calibration for Mexico

The calibration regarding the parameter λ_p that captures hours spent outside the home in Mexico uses information from the 2014 household time use survey *Encuesta Nacional sobre Uso del Tiempo 2014* from the National Statistics Office (INEGI).[1] From the survey we consider time spent outside the home as the sum of aggregate hours in *market activities and consumption goods*, *entertainment and social activities*, and *study and related activities*. As for time spent at home, we aggregate all hours in *non-remunerated work at home*, and *personal activities* that include sleeping, eating, and personal hygiene. We conclude that on average a Mexican household spends 36% of total time in activities outside the home, which corresponds to a parameter of $\lambda_p = 1.77$.

As for the parameter that regulates the preferences for staying alive b, we use data from Google Community Mobility Reports for Mexico to determine the reduction in away-from-home activities during the Covid-19 epidemic.[2] We average all the non-home activities (retail and recreation, grocery and pharmacy visits, visit to parks, transit, and workplace activity) and measure a 7-day centered moving average, as shown in Figure A7. This analysis reveals that outside the home activity decreased by about 50% at the trough of the epidemic and we use this decline to calibrate the parameter b in the model simulations.

Figure A7:
Mobility as a Response to Covid-19 in Mexico



Notes: This graph shows the percentage change in activity outside the home using data from Google Community Mobility Reports for Mexico (available at https://www.google.com/covid19/mobility/). We show the actual daily data as well as a 7-day centered moving average.

[1]The dataset can be accessed at https://en.www.inegi.org.mx/programas/enut/2014/.
[2]Google Mobility data can be accessed at https://www.google.com/covid19/mobility/.

C.2 Robustness over the choice of parameters

Table A3 shows how the model results change when we either increase or decrease important parameters, while keeping all other constant to the baseline model (delays use the histogram for the whole country).

Table A3:
Robustness Checks on the Model

	Peak infections (% of pop)	Days to peak deaths	Maximum daily deaths	Total deaths (day 120)	Total deaths (day 500)	Hrs. susceptible to infection at trough
Baseline						
No delays	0.760	107	1017	51812	210585	19.65
Delays	0.988	90	1271	64534	213990	18.42
Higher death rate: $\delta = 0.016$						
No delays	0.302	111	955	52891	249282	18.76
Delays	0.401	85	1176	65660	256449	17.47
Lower death rate: $\delta = 0.004$						
No delays	1.504	103	1151	54553	186432	20.40
Delays	2.043	92	1466	67516	186285	19.36
Higher infection rate: $1/\gamma = 10$						
No delays	1.868	95	1741	103149	345529	12.93
Delays	2.647	78	2293	129788	353123	11.89
Lower infection rate: $1/\gamma = 5$						
No delays	0.424	120	798	33868	179085	22.76
Delays	0.541	102	965	42650	181792	21.73
Higher resolving probability: $\theta = 0.2$						
No delays	0.670	98	1066	58871	230210	19.43
Delays	0.901	83	1404	72356	233912	18.20
Lower resolving probability: $\theta = 0.05$						
No delays	0.666	125	984	44010	229306	19.22
Delays	0.896	100	1133	55868	233019	17.94

Notes: This table shows results from changing parameters of the model. We consider a higher and lower death rate, infection rate and resolving probability. For each case, we show estimates from model with and without delays in death reporting. We present the estimates for the peak number of infections (expressed as a percentage of the total population), the number of days it takes from the onset of the epidemic to reach the peak for deaths, the maximum number of daily deaths, the total number of deaths accrued up to the 120th and 500th day, and the hours in a day susceptible to infection at the trough of the curve.

C.3 Model computation

In order to solve the model we use the follow algorithm:

1. Choose a sequence for a large T and some sequence $\left\{\Pi_t^0\right\}_{t=0}^T$; make sure $\Pi_T = 0$;

2. Solve for the values using backward induction and get policies on $n(j,t)$;

3. Compute the path of $M_t(j)$ for each j;

4. Update probabilities Π_t^1;

5. Iterate until $\left|\Pi^1 - \Pi^0\right| < \epsilon$ for small ϵ, otherwise set $\Pi^0 = \Pi^1$ and go back to (2).

C.4 Alternative model specifications

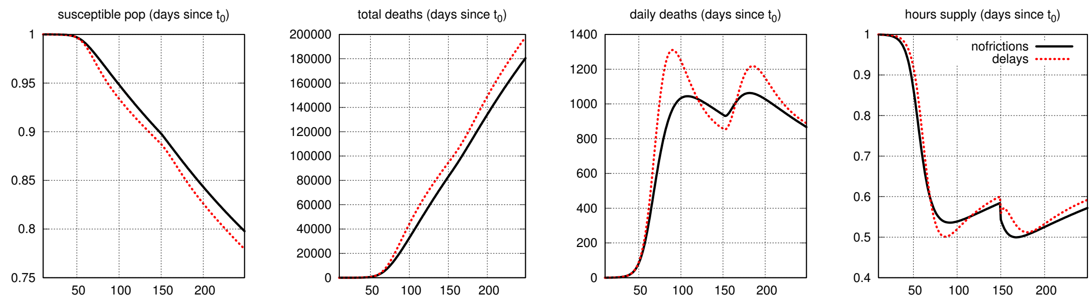
We present simulations for three modifications to the model. First, we introduce a second wave of infections, for example, due to seasonal variations in transmissibility or new virus mutations. Second, we incorporate additional days between when an agent becomes infected and when that agent becomes infectious (i.e., when the individual is a carrier of the virus but cannot transmit it). Lastly, we consider a death rate that declines over time due to, for instance, improvement in medical technology or treatments.

C.4.1 Second wave of infections

Notice that including a second wave, modeled for example as an unexpected shock to the Covid-19 transmissibility risk Λ would aggravate the problem of delays. Figure A8 and Table A4 illustrate this exercise when, at $t = 150$ of the epidemic, we introduce an unexpected increase in the transmissibility risk with $\Lambda' = 1.3\Lambda$, all else equal. This alternative specification is meant to capture in a very stylized fashion an unexpected change in transmissibility risk that can emerge, for example, from seasonal climate differences or from virus variants emerging from natural mutations. Such unexpected environmental changes imply a larger scope for the information frictions to play a role in aggravating the impact of the epidemic due to behavioral responses.

Results are qualitatively similar to the ones obtained in the baseline simulation with the main difference associated with a much larger increase in the total number of deaths relative to a frictionless information world by the end of the epidemic: a difference of 20,000 deaths, accounting for an increment of 7.6% more relative to a no frictions world, which compares to an increment of 1.6% in the baseline simulation.

Figure A8:
Simulation Results of Behavioral Model with and without Delays in Death Reports: Multiple Waves



Notes: These graphs show the simulation results from the model by computing the equilibrium as defined in subsection 4.2. We show results for a situation without reporting delays (solid line) and with delays calibrated to the Mexican data (dashed line) when introducing multiple waves. We show the mass of susceptible individuals, total deaths, daily number of deaths, and hours supplied outside the home.

Table A4:
Epidemic Statistics from Model Simulations with and without Delays in Death Reports: Multiple Waves

	Peak infections (% of pop)	Peak deaths (days)	Maximum deaths (daily)	Total deaths (at day 120)	Total deaths (at day 500)	Hours outside (at trough)
No delays	0.668	181	1,063	53,602	264,804	18.04
With delays	0.899	89	1,310	66,891	284,821	18.09
Reported deaths	-	98	1,196	57,263	284,821	-

Notes: This table shows statistics on the epidemic generated in the model when introducing multiple waves. We show results for a situation without reporting delays (first row) and with delays calibrated to the Mexican data (second row). The last row presents statistics on deaths as reported in the model with delays.

C.4.2 Time between infection and infectiousness

An alternative model specification can allow for a few periods where, despite contracting the virus, agents are not infectious. The medical literature seems to identify four days.[3] Hence, we just add four more state variables between susceptibility, $j = s$ and infection $j = i$. Value functions are now

[3] See for example: https://medical.mit.edu/covid-19-updates/2020/10/exposed-to-covid-19-how-soon-contagious.

given by:

$$
\begin{aligned}
V(s,t) &= \max_{n\in(0,1)} \left\{u(n) + \beta\left(\left[1-\pi\left(n,\widetilde{\Pi}_t\right)\right]V(s,t+1) + \pi\left(n,\widetilde{\Pi}_t\right)V(b,1)\right)\right\} \\
V(b,1) &= \max_{n\in(0,1)} \{u(n) + \beta V(b,2)\} \\
V(b,2) &= \max_{n\in(0,1)} \{u(n) + \beta V(b,3)\} \\
V(b,3) &= \max_{n\in(0,1)} \{u(n) + \beta V(b,4)\} \\
V(b,4) &= \max_{n\in(0,1)} \{u(n) + \beta V(i)\} \\
V(i) &= \max_{n\in(0,\bar{n})} \{u(n) + \beta[\gamma V(c) + (1-\gamma)V(i)]\} \\
V(c) &= \max_{n\in(0,\bar{n})} \{u(n) + \beta((1-\theta)V(c) + \theta[(1-\delta)V(r) + \delta V(d)])\} \\
V(r) &= \max_{n\in(0,1)} \{u(n) + \beta V(r)\} \\
V(d) &= 0
\end{aligned}
$$

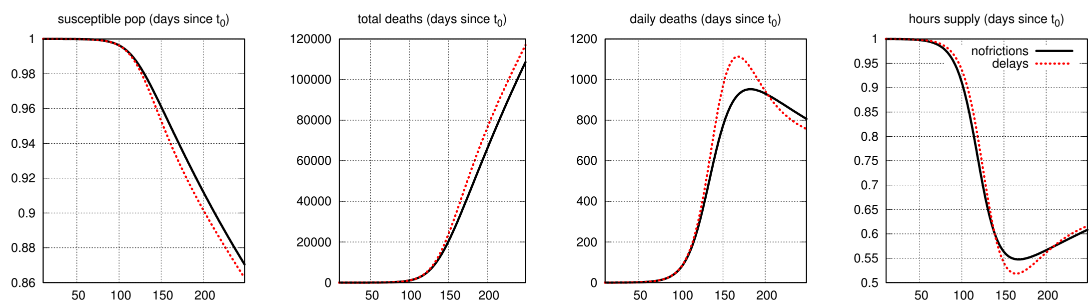
This structure assumes that agents know when they get infected and when they become infectious. We also assume that while infected but not infectious, agents can supply an unrestricted amount of hours outside the home. The system of laws of motion becomes:

$$
\begin{aligned}
M_{t+1}(s) &= M_t(s) - \pi(n(h,t), \Pi_t)M_t(s) \\
M_{t+1}(b,1) &= \pi(n(h,t), \Pi_t)M_t(s) \\
M_{t+1}(b,2) &= M_t(b,1) \\
M_{t+1}(b,3) &= M_t(b,2) \\
M_{t+1}(b,4) &= M_t(b,3) \\
M_{t+1}(i) &= M_t(i) - \gamma M_t(i) + M_t(b,4) \\
M_{t+1}(c) &= M_t(c) - \theta M_t(c) + \gamma M_t(i) \\
M_{t+1}(r) &= M_t(r) + (1-\delta)\theta M_t(c) \\
M_{t+1}(d) &= M_t(d) + \delta\theta M_t(c)
\end{aligned}
$$

And finally, when faced with a situation that requires belief updating, agents take into account the existence of these intermediate states.

Figure A9 and Table A5 summarize the results while keeping the calibration and delays as in the baseline simulation shown in the main text. Notice how the role of the periods between infection and infectiousness slows down the progression of the virus. Apart from this, the results are kept more or less unchanged relative to the baseline. Also interesting is the fact that the delays in providing information have a relatively less impactful effect on the peak of daily deaths. This happens since the time buffer before an agent becomes infectious mitigates the forecast errors made by agents due to the lagged government information, as this is updated on a daily basis.

Figure A9:
Simulation Results of Behavioral Model with and without Delays in Death Reports: Increased Time between Infection and Infectiousness



Notes: These graphs show the simulation results from the model by computing the equilibrium as defined in subsection 4.2. We show results for a situation without reporting delays (solid line) and with delays calibrated to the Mexican data (dashed line), in a model where we allow for additional periods between infection and becoming infectious. We show the mass of susceptible individuals, total deaths, daily number of deaths, and hours supplied outside the home.

Table A5:
Epidemic Statistics from Model Simulations with and without Delays in Death Reports: Increased Time between Infection and Infectiousness

	Peak infections (% of pop)	Peak deaths (days)	Maximum deaths (daily)	Total deaths (at day 120)	Total deaths (at day 500)	Hours outside (at trough)
No delays	0.604	181	952	4,267	190,463	19.77
With delays	0.723	167	1,113	4,460	194,978	18.72
Reported deaths	-	177	1,061	2,963	194,978	-

Notes: This table shows statistics on the epidemic generated in the model where we allow for additional periods between infection and becoming infectious. We show results for a situation without reporting delays (first row) and with delays calibrated to the Mexican data (second row). The last row presents statistics on deaths as reported in the model with delays.

C.4.3 Better technology of preventing deaths

Another alternative considers a situation where, over time, there is a decrease in the death rate due to better medical treatment or the development of better drugs. We implement this extension by assuming that the death rate of the virus decays linearly during the course of the epidemic such that, by day 365, the death rate is half of the initial one. Allowing for agents to have perfect foresight of this change in medical technology means also a change in the value functions that incorporate such improvement:

$$V(s,t) = \max_{n\in(0,1)} \left\{ u(n) + \beta \left(\left[1 - \pi\left(n, \widetilde{\Pi}_t\right) \right] V(s,t+1) + \pi\left(n, \widetilde{\Pi}_t\right) V(i,t+1) \right) \right\}$$

$$V(i,t) = \max_{n\in(0,\bar{n})} \left\{ u(n) + \beta \left[\gamma V(c,t+1) + (1-\gamma) V(i,t+1) \right] \right\}$$

$$V(c,t) = \max_{n\in(0,\bar{n})} \left\{ u(n) + \beta \left((1-\theta) V(c,t+1) + \theta \left[(1-\delta(t)) V(r) + \delta V(d) \right] \right) \right\}$$

with $\delta(t)$ being the specific period t death rate, and $V(r)$ and $V(d)$ are defined as in the baseline model. Similarly, the laws of motion for the system and how agents update their beliefs about the current prevalence take into consideration the time-varying death rate $\delta(t)$.

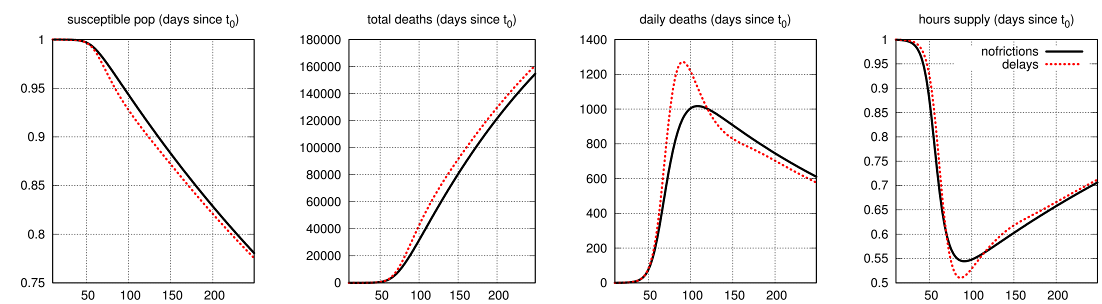
The results, summarized in Figure A10 and Table A6, are very similar to the ones obtained in the baseline model with a slight mitigation of the impact of delays due to a progressively less dangerous risk of contracting the virus. The main difference in this version of the model is on the faster path of recovery for hours, also consistent with the fact that agents adjust their behavior taking into account the lower risk of dying once infected.

Table A6:
Epidemic Statistics from Model Simulations with and without Delays in Death Reports: Declining Death Rate

	Peak infections (% of pop)	Peak deaths (days)	Maximum deaths (daily)	Total deaths (at day 120)	Total deaths (at day 500)	Hours outside (at trough)
No delays	0.760	107	1,017	51,812	210,585	19.65
With delays	0.988	90	1,271	64,534	213,990	18.42
Reported deaths	-	99	1,161	55,174	213,990	-

Notes: This table shows statistics on the epidemic generated in the model. We show results for a situation without reporting delays (first row) and with delays calibrated to the Mexican data (second row) for a model with a death rate that declines over time. The last row presents statistics on deaths as reported in the model with delays.

Figure A10:
Simulation Results of Behavioral Model with and without Delays in Death Reports: Declining Death Rate



Notes: These graphs show the simulation results from the model by computing the equilibrium as defined in subsection 4.2. We show results for a situation without reporting delays (solid line) and with delays calibrated to the Mexican data (dashed line) for a model with a death rate that declines over time. We show the mass of susceptible individuals, total deaths, daily number of deaths, and hours supplied outside the home.