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

Emilio Gutierrez† Adrian Rubli‡ Tiago Tavares§

October 2021

Providing information is important for managing epidemics, but issues with data accuracy may hinder its effectiveness. Focusing on Covid-19 in Mexico, we ask whether delays in death reports affect individuals' beliefs and behavior. Exploiting administrative data and an online survey, we provide evidence that behavior, and consequently the evolution of the pandemic, are considerably different when death counts are presented by date reported rather than 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 individual responses, 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 are very grateful to the editor and anonymous referees for providing comments that greatly improved this paper. We thank Jose Maria Barrero, Andrew Foster, Miguel Messmacher, Charles Wyplosz, and participants at the ITAM Brown Bag seminar for their helpful feedback. 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.

†*Corresponding author.* 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

§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). A particularly 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 importance of the *accuracy* of this information.[1] In particular, governments across the world have made efforts to collect and communicate information to their citizens. If these government 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.[3]

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 is ideal for studying these issues for at least three reasons. First, the delays in Covid-19 death reports are not only measurable and large, but vary greatly across states.[4] Gutierrez et al. (2020) documents that these delays are correlated with local measures of the capacity of the public healthcare system. The top panel in Figure 1 depicts these delays at the national level by showing cumulative deaths as reported versus as occurred as well as the distribution of reporting delays

[1] In the context of Covid-19, it has been shown that information-focused public policies have been important for managing behavior (Gupta et al., 2020; Briscese et al., 2020).

[2] For other contexts, the evidence is mixed. For instance, for HIV in Africa, studies have found large effects of information on behavior (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 non-HIV mortality matters for adopting protective behaviors.

[3] We abstract from instances where individuals may make simple adjustments in an attempt to account for data inaccuracies.

[4] Delays in reporting deaths have been documented across many settings (AbouZahr et al., 2015; Bird, 2015).

in days. Second, Mexican officials routinely present information on confirmed Covid-19 deaths over time, giving salience to the number of *reported* deaths.[5] Lastly, the Mexican government chose a relatively lenient strategy that consisted of mostly optional lockdowns and stay-at-home recommendations, as well as restricting testing to symptomatic individuals seeking medical care. Hence, deaths have been particularly salient as a more reliable measure of the state of the epidemic and behavioral responses have been crucial for its containment.[6]

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 by assuming that deaths reported are a good approximation of deaths occurred.

We complement these empirical results by fielding an online survey, where we randomized information about the epidemic. We compare respondents' beliefs regarding the severity of the epidemic and their reported intentions of complying with stay-at-home recommendations between groups that were shown the evolution of Covid-19 deaths by date reported versus the date on which they actually 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 delays have on the evolution of the pandemic through their 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, the higher the incentive to stay home. Agents rely on death reports provided by the government – which they take to be accurate – to form expectations about the current and future prevalence of the disease, updating expectations with each new report.

We calibrate this model to our setting and compare outcomes in a scenario where deaths are reported as occurred relative to a situation where reporting delays follow the empirical distribution we observe in Mexico. Inaccurate information due to reporting delays leads to individuals being

[5] See, for instance, this government website: https://coronavirus.gob.mx/datos/.

[6] See, for example, https://globalhealthsciences.ucsf.edu/sites/globalhealthsciences.ucsf.edu/files/la_respuesta_de_mexico_al_covid_esp.pdf for details on how information on cases and deaths in Mexico has been interpreted.

slower to adopt protective behaviors and to more severe epidemic outcomes in terms of cases and deaths, with a peak of daily deaths 25% larger (266 additional deaths) than in the model without delays, which implies that eliminating delays in the model would decrease total deaths by 13,289 by day 120 of the epidemic. Moreover, the faster speed of the epidemic induced by slower reactions will tend to generate excessive responses later on, which may exacerbate the associated negative economic impacts.

Overall, our analysis suggests that improving the capacity to convey accurate information matters for how agents respond. Communicating more accurate information can change the full dynamics of the epidemic itself with lower total deaths, smaller activity fluctuations, and less intense outbreaks, which may also imply a better management of hospital capacity.[7] These results seem particularly relevant for settings with weak state capacity, where the reliability of the data has often come into question.[8] In a broader sense, our findings suggest that the accuracy of information – of any kind and at any moment – influences behavior that may lead to substantially different epidemic outcomes.

We contribute to three strands of the growing literature on the economics of Covid-19. First, our paper relates to those that have explored how messages and information affect various outcomes.[9] Akesson et al. (2020) provides different information about Covid-19 infectiousness, finding that individuals who received the larger estimate of contagion risk were actually less likely to report complying with mitigating behaviors. Falco and Zaccagni (2020) shows that reminders matter for stay-at-home compliance, while Breza et al. (2021) finds that messages from healthcare workers reduced mobility. Binder (2020) and Coibion et al. (2020) randomize information on government policies in the US to measure its impact on consumer beliefs and spending. While these studies focus on the effect of receiving information, our paper emphasizes the role of its accuracy.[10] In a

[7] 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.

[8] See, for example, undercounting in India (https://www.nytimes.com/2021/04/24/world/asia/india-coronavirus-deaths.html?searchResultPosition=1) and Ecuador (https://www.nytimes.com/2020/04/23/world/americas/ecuador-deaths-coronavirus.html?searchResultPosition=3), last accessed May 13, 2021.

[9] Other mediating factors that the literature has analyzed include 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), and the media (Simonov et al., 2020).

[10] Studies have also analyzed the link between risk perceptions and prosocial behavior during Covid-19 (Campos-Mercade et al., 2021; Abel et al., 2021; Brañas Garza et al., 2020).

similar spirit, Bursztyn et al. (2021) analyzes the role of veracity of messages in the US, but centers on opinion programs in the media.

Second, our paper relates to studies incorporating changes in behavior into dynamic epidemiological models (Fernández-Villaverde and Jones, 2020; Brotherhood et al., 2020). While we favor a simple and parsimonious approach, 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 during Covid-19, including 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 conveying reliable real-time information. 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.

2 Motivating Correlations

2.1 Data

The Mexican government provides detailed patient-level information for all recorded Covid-19 cases, with daily updates. We observe the patients' state of residence, when they first sought medical attention for Covid-19 symptoms, the self-reported date for the onset of symptoms (all reports are symptomatic), the result of a Covid-19 laboratory test, and, if applicable, the date of death. With each daily update, patients may transition from unconfirmed to confirmed Covid-19, and from alive to dead.

The testing rate for Covid-19 in Mexico has been, by design, one of the lowest.[11] Only symptomatic individuals seeking medical attention are tested. Hence, deaths are arguably more precisely measured than cases. Furthermore, this feature has made deaths more salient in this context.

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 to cases from March 14 to September 30, 2020, allowing up to four months for all occurred cases and deaths

[11] See, for example, https://ourworldindata.org/coronavirus-testing.

during this period to be reported. 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, allowing us to track the number of reported and occurred cases and deaths over time.

2.2 Empirical correlations

In a standard model with exogenous behavior, the growth rate of cases is fast at first, but then slows down as more individuals stop being susceptible. Since deaths increase over time, one would expect a negative correlation between the growth rate of cases and the number of deaths, particularly before the epidemic peaks. If agents' behavior responds to the epidemic – with individuals reducing their exposure when prevalence is high – an increase in deaths could then predict a decrease in epidemic growth.

We consider two measures of deaths in our data: the true count (occurred) and a noisy signal (reported). We then estimate correlations conditional on state and time fixed effects (FE), controlling for other features of the epidemic curve, and ask whether the growth rate is more responsive to reported or occurred deaths. In an epidemic model without endogenous behavior and assuming we cannot fully control for the shape of the curve, then one might expect the correlation between growth in cases and occurred deaths to be stronger statistically, due to noise in reported deaths. Likewise, in a model with endogenous behavior, perfectly informed agents should respond to actual deaths and not to reported deaths. However, if agents are not perfectly informed, the growth rate of cases may have a stronger correlation with this imprecise measure instead.

We compute two similar measures of the growth rate of Covid-19 cases for each state-week in our data. We take the percentage change in the number of patients that self-reported having first shown symptoms from week t to $t+1$ and the change from week t to $t+2$, and estimate:

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

where $y_{s,t}$ is the growth rate of Covid-19, λ_s are state FE, γ_t are week FE, $\mathbf{X}_{s,t}$ is a vector of epidemic characteristics (such as the number of cases in time t), and $\varepsilon_{s,t}$ is the error term.

Table 1 shows the estimates. The dependent variable in the first three columns is the percentage change in the number of self-declared symptoms from week t to $t+1$, while the last three columns

consider the change from t to $t+2$. Columns 1 and 4 include state and week FE and log Covid-19 cases in week t as controls. Columns 2 and 5 add 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 and reported deaths between week $t-2$ and week $t-1$.

Across specifications, the coefficient associated with the number of reported deaths is negative and significantly different from zero at a high confidence level. Column 1 indicates that when reported deaths double, there is an associated decline of 0.085 points in the growth rate of cases. In contrast, the coefficient associated with the number of occurred deaths is smaller in magnitude and statistically insignificant.

We interpret these correlations as motivating evidence that the growth rate in Covid-19 cases is more responsive to the *noisy* measure of deaths rather than the *actual* number of deaths, even after we control for overall time trends and features of the epidemic. We conjecture that this relationship may be driven by individuals incorrectly inferring Covid-19 prevalence from the number of reported, instead of occurred, total deaths. These empirical correlations motivate the question of whether reporting delays could impact individuals' perceptions and actions.

3 Online Survey

3.1 Survey description and respondents' characteristics

To shed further light on whether delays matter, we conducted an online survey with a randomized informational treatment presenting the evolution of total deaths either by date reported or by actual date of death. The survey ran from May 28 to June 8, 2020, with participants recruited mainly from an ITAM mailing list of faculty, administrative staff, and students ($N = 1,022$).[12]

After the initial questions on socioeconomic characteristics and pre-intervention perceptions, respondents were taken to a new screen showing (randomly) one of the two graphs depicted in the bottom panel of Figure 1. Half ($N = 508$) were shown the plot on the left, which plots cumulative deaths in Mexico by date reported. The rest were instead shown the plot on the right with cumulative deaths by actual date of occurrence. Both figures show counts from March 22 to May 15, using data up to May 27. As a reference, we also included the cumulative number of

[12]For the full survey questions, see the online repository https://github.com/tgstavares/covidinfo.

deaths by date reported in Sweden.[13] Both plots contain truthful information, although the plot on the left understates total deaths as occurred by 41% on average, with a difference of up to 2,055 deaths on May 11.

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

Due to the composition of our mailing list, our participant characteristics suggest they belong to a relatively young, educated, and high-income group. Hence, we cannot infer the distribution of beliefs and behavior in the general population. Lastly, although observable differences between our treatment groups are small (Table S2 in the supplementary materials), we control for those characteristics below.

3.2 Empirical strategy

We estimate the following equation:

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

where y_i is an outcome for respondent i, α_0 is a constant, $[\text{Info By Date Occurred}]_i$ is an indicator for receiving the informational treatment that displayed cumulative deaths by actual date of death, $\mathbf{X}_i$ is a vector of observable characteristics (see Table S2), and ν_i is the error term. The coefficient α_1 measures the average difference in the outcome for respondents that were shown the cumulative death toll by date of occurrence relative to those with information by date reported.

We explore three outcomes. First, for simplicity, we construct two binary measures for whether the epidemic is evolving faster in Mexico than in Sweden, depending on whether respondents considered the epidemic evolving "faster" or "much faster" than in Sweden, or strictly considering it "much faster". Second, regarding beliefs about the epidemic's toll, we assign the total number of expected cases and deaths to be equal to the mid-value of the interval chosen by respondents.[14]

[13] These data were obtained from https://ourworldindata.org/coronavirus. Sweden followed a similar strategy of relatively light restrictions (Juranek and Zoutman, 2020). The epidemic in Mexico had been compared to Sweden's by government authorities. 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.

[14] We assigned a value of 3,000,000 for "more than 2,000,000 cases", and 300,000 for "more than 200,000 deaths".

Lastly, for social distancing, we use both the number of days respondents expect to leave their homes in four weeks, and an indicator for expecting to leave their house three or more times.[15]

3.3 Results

Table 2 shows the results. Each panel corresponds to a different pair of outcomes. Columns 1 and 4 consider the full sample. We then decompose results by respondents' priors based on self-reported knowledge about the total number of cases pre-intervention.[16] Columns 2 and 5 restrict to respondents that reported fewer cases than the truth (low prior) and columns 3 and 6 show those with a larger number (high prior).

Presenting cumulative deaths by actual date of occurrence shifts beliefs towards a perception of a faster spreading epidemic: the fraction of respondents considering that the epidemic was progressing much faster than in Sweden increased by 26 percentage points relative to those shown the plot by date reported. We take this as a first stage result showing that our informational intervention had the expected effect.

Respondents also predict a higher toll of the epidemic when shown deaths by date of occurrence. Individuals who saw the evolution of occurred deaths announced an expected total with, on average, 14% more cases and 11% more deaths. This effect is larger in the low prior subsample. For the self-reported intentions of staying home – which may differ from actual behavior – the results are consistent with information presented by actual date of death having a positive impact. Showing the graph by date of occurrence is associated with a decrease in the number of times people expect to leave their homes (four percentage points or 10% lower probability of leaving three or more times), particularly for the low prior subsample (11 percentage points or 32%).[17]

Notwithstanding the limited statistical power due to our small sample and the relatively small differences in the information provided, we interpret these results as evidence that the delays with

[15] We assign a value of 3.5 for the "3-4 times" category, and 5 times for "5+". We show results in Figure S3 using indicators for each of the possible response categories for all our outcomes.

[16] The low prior subsample are those who reported a case load lower than 50,000 (47.7%), while the high prior group are those that reported over 50,000 cases (Figure S4). The true number was 56,594 (see https://twitter.com/HLGatell/status/1263264663283908609?s=20, last accessed June 29, 2020).

[17] We obtain positive coefficients for the high prior subsample, although not precisely estimated. This may be consistent with a fatalistic response (Akesson et al., 2020). As modeled by Kerwin (2020) for HIV in Africa, perceptions of a higher risk of contagion may lead to low marginal costs of risk-taking behavior, leading individuals to take on more risks in a "rationally fatalistic" response. While we cannot definitely confirm it, our results for the high prior subsample in panel C of Table 2 are consistent with this response and warrant further exploration in the future.

which deaths are reported are very likely to affect perceptions about the state of the epidemic and, consequently, compliance with social distancing. These findings also suggest that individuals do not fully incorporate reporting delays when forming beliefs about the epidemic, even in our highly educated sample. We proceed incorporating these insights into an equilibrium model.

4 Model of Equilibrium Behavior

We present an equilibrium model to illustrate the impact of reporting delays on epidemic dynamics through the endogenous behavioral response of agents. We follow 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.[18] The main difference is how agents update expectations about Covid-19 prevalence . Given a prior, agents form plans about consumption and leisure over their life-cycle. The government informs about deaths, which agents take as accurate.[19] Agents then update their prior in order to rationalize the number of deaths as reported by the government. This also changes their planning over economic decisions in the remainder of their life-cycle, thus affecting the dynamics of the epidemic.

States. Each period represents one day in discrete time. There is a continuum of ex-ante identical agents that can spend time at and away from home. Let j be an agent's health status. The initial state $j = s$ (**susceptible**) is never infected. Being outside may lead to $j = i$ (**infected**). Infected agents can contaminate susceptible ones (with a uniform mixing contact rate). With probability γ contagiousness ceases and a $j = c$ (**recovering**) process follows. Agents exit this state with probability θ, and a share $1 - \delta$ become $j = r$ (**recovered**), and a share δ become $j = d$ (**dead**). We assume that recovered agents are permanently immune. The future is discounted at rate β.

Utility and hours. Each agent is endowed with a single unit of labor every period, divided into work/leisure hours n outside and hours at home $h = 1 - n$. Flow utility is derived from hours

[18] Extensions include macroeconomic implications, savings, non-pharmacy initiatives, testing, vaccines, optimal lockdowns, and age and asset heterogeneity (Eichenbaum et al., 2020a,b; Acemoglu et al., 2020; Alvarez et al., 2020; Kaplan et al., 2020).

[19] Information could also focus on cases instead of deaths, and the results would follow through. We assume that agents do not perfectly anticipate the frictions associated with how the government provides information. With perfect knowledge, beliefs about prevalence would be accurate even with delays. This is unlikely given the content of the government's press conferences and due to the degree of sophistication required of individuals.

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. Hence:

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

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. We assume that hours outside are constrained to $\bar{n} < 1$ for infected and recovering individuals. Value functions are then given by:

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

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

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

$$V(r) = \max_{n\in(0,1)} \left\{ u(n) + \beta V(r) \right\} \quad \text{(value recovered)}$$

$$V(d) = 0 \quad \text{(value dead)}$$

Laws of motion. Letting $n(j,t)$ be the optimal hours outside for states $j = s, i, c, r$, the laws of motion are given by:

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

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

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

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

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

Total population is normalized to 1, so that:

$$1 = M_t(s) + M_t(i) + M_t(c) + M_t(r) + M_t(d), \quad \forall t.$$

Aggregate probability of infection. The (instantaneous) Poisson rate of infection is given by:

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

where Λ is the biological transmissibility of the disease. Hence:[20]

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

Information and priors. Agents accurately know all the parameters of the model, but are unaware of the initial mass of infectious individuals $M_0(i)$, forming a prior $\tilde{M}_0(i)$, which may be different from the truth. In this case, forecasts of the epidemic dynamics made by agents will be biased. In particular, given a history of labor supply $\{n(s,t), n(i,t)\}_{t=0}^{t'} \equiv \left\{n(s)^{t'}, n(i)^{t'}\right\}$ for any $t' \geq 0$, the following probabilities emerge as potentially different:

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

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

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

4.1 Definition of an equilibrium

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

[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 $Prob(t < \bar{t}) = 1 - \exp\left(-\hat{\Pi}\bar{t}\right)$. For a single period, $\bar{t} = 1$, we then have that $Prob(t < 1) = 1 - \exp\left(-\hat{\Pi}\right)$.

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

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

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

4.2 Model analysis

The static solution for hours outside is given by:

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

This also corresponds to hours outside for fully recovered individuals, $n(r,t) = n^*$. Assuming that the maximum hours outside for infected and recovering individuals is capped by the static optimal, these individuals will supply: $n(i,t) = n(c,t) = \bar{n} < n^*$.

This then yields closed-form solutions:

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

As for susceptible agents, first order conditions imply:

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

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, $V(s,t+1) > V(i)$ for any $\pi\left(n, \tilde{\Pi}_t\right) > 0$. This implies that hours supplied by susceptible individu-

als are $n(s,t) < n^*$ 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.

4.3 Update of beliefs based on deaths reported by the government

If agents do not receive information during the course of the epidemic, the equilibrium outcome from the definition in subsection 4.1 follows and the epidemic runs its course based on agents' behavior and potentially misspecified beliefs. But if the government provides information on current deaths $D_t \neq \tilde{M}_t^{prior}(d)$, then agents update their information about prevalence in order to match their forecast with what the government is announcing. Agents will then get an updated set of beliefs for the masses at t given by $\tilde{M}_t^{posterior}(j)$ for each $j = s, i, c, r, d$, where now $D_t = \tilde{M}_t^{posterior}(d)$. 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 of the equilibrium.

4.4 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 $D_t \neq M_t(d)$, such that:

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

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, $D_t = \tilde{M}_t^{posterior}(d) \Rightarrow \tilde{M}_t(j) \neq M_t(j)$ for $j = s, i, c, r, d$.

4.5 Simulation of an epidemic with delays in deaths reported

Given a parametrization and specification of delays, we numerically simulate the model following the definition of the equilibrium in subsection 4.1 using value function iteration.

Calibration. 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 (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 outbreak. The parameter b captures a drop in total hours outside the home during the epidemic of 50% as suggested by Google Mobility data. Lastly, the baseline contagion rate parameter Λ is set to generate a basic reproduction number of two (Marioli et al., 2020). Table S3 summarizes these parameters.

Reporting delays. We capture the distribution of the difference in days between deaths as reported and deaths as occurred as given in the data for the whole country. Specifically, we use the following formula:

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

where the coefficients $\{p_t, ..., p_{t-60}\}$ capture the same density as what we observe in a histogram of the data. This effectively allows for a delay of up to two months between when a death occurred and when it was reported. The distribution of delays shows an average of 8.6 days and a standard deviation of 11.5 days (see Figure 1).

End of the epidemic. We allow for the introduction of a vaccine, which is expected ex-ante. After becoming available at time $t^{vacc} = 350$ days, transmissibility immediately becomes zero.

4.6 Results

Figure 2 compares the epidemic dynamics in a model without delays (solid line) relative to one with delays as described above (dashed line). Delays lead to a faster progression of the epidemic, implying 25% higher daily deaths at the peak (266 additional deaths) and 3,706 additional total deaths by the end of the outbreak (Table S4). This implies that eliminating delays would decrease total deaths by 13,289 at day 120 of the epidemic. Furthermore, the slow adjustment of hours at the onset contributes to a higher prevalence at the peak, jumping from 0.67% to 0.90% of the population. Once the government informs about deaths, agents realize that the spread of the disease is serious and thus adjust their behavior more abruptly. This causes a larger collapse of

hours spent outside than in the case without delays. Relative to pre-pandemic levels, hours away from home dip up to 46% without delays, but 50% with delays.

The bottom row in Figure 2 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, this leads to an underestimation of the severity of the disease, and therefore a lower behavioral response from agents.[21]

Extreme cases of delays. Aggregate delays mask large variations at finer geographical levels (Gutierrez et al., 2020). Taking the extreme cases among the 32 states in Mexico, delays in the state of Tamaulipas are on average four times larger than in Querétaro (Figure S6). To highlight our effects, we simulate the model under the same parameters as before, but using a delay distribution modeled after each of these states. We, again, allow for delays of up to 60 days.[22]

Figure 3 shows the results. The epidemic progresses much faster when delays are larger, as in Tamaulipas. Agents are slower to react to the progression of the disease, which contributes to a larger peak, with a share of 1.06% of the population infected relative to 0.73% if delays occur as in Querétaro (Table S5). Due to the acceleration in contagion, peak deaths increase to 1,522 in the case of large delays, which is 38% more than in the case of small delays. By the end of the epidemic, the additional deaths relative to the case without delays are more than three times greater with large delays than with shorter delays (5,440 vs 1,600), or in absolute terms, reducing delays in Tamaulipas to the level observed in Querétaro lowers the toll of total deaths by the 120th day by 15,644, or by 20,802 if delays are completely eliminated.

The last two plots in Figure 3 help explain the difference in epidemic dynamics between both extreme cases. The next-to-last panel shows how delays evolve, with Querétaro stabilizing at around three days while Tamaulipas stabilizes at 14 days. In the last panel, we show that at the beginning of the epidemic, agents would underestimate the prevalence of the infected population by almost a

[21] Our results are consistent with the predictions generated by standard epidemiological models. For example, an SIR model based on Capistran et al. (2021) would predict 70,000 deaths by day 120 of the epidemic as opposed to 67,000 in our baseline model. However, using data as reported to estimate the model would generate only 51,000 deaths by day 120 (Gutierrez et al., 2020). Furthermore, since SIR models fix behavior exogenously, analysis on how the epidemic may change due to frictions in data reports cannot be investigated.

[22] If delays are larger than that in Tamaulipas, then we may be undercounting deaths. However, this is unlikely given the delay distribution. Furthermore, the model only uses the distribution of delays (which should be the same when undercounting deaths) and not the actual death count.

full log point (-63%) in the presence of large delays relative to a 0.4 log point (-33%) in the case of shorter delays.

These extreme examples further suggest that tackling the issue of delays may require more localized efforts. For instance, improving death registry in Querétaro may be easier than in Tamaulipas, where delays are so much larger, which could suggest that a more cost-effective intervention in this state would require using statistical prediction methods to adjust the data for delays.

5 Conclusion

The Covid-19 pandemic has been a large experiment on the use of high-frequency and detailed data as a policy lever during a public health shock. The extent to which disseminating information may help mitigate the effects of an epidemic depends largely on its accuracy and reliability, and how individuals may update their priors and respond. In settings with low state capacity, collecting accurate real-time information may be particularly challenging. In this paper, we showed that delays in death reports during Covid-19 are sizable in Mexico and affect individual beliefs and behavior, leading to more severe epidemic outcomes.

From a policy perspective, our results highlight the importance of accurate real-time information, or at least being clear about the shortcomings of the available data. Increasing collection efforts may be very costly, and while government resources during a pandemic may be better spent on other palliative measures, our findings suggest that simple, low-cost interventions – such as using statistical prediction techniques to correct for delays – may improve outcomes.[23] Overall, our empirical and simulation results underscore the role of accuracy when using information as a public health tool for mitigating an epidemic.

[23] In that scenario, there may be additional considerations, such as a preference to exaggerate if concerns about public health are larger than economic ones (de Véricourt et al., 2021). However, this may backfire if individuals are rationally fatalistic in their responses (Akesson et al., 2020).

References

Abel, M., T. Byker, and J. Carpenter (2021). Socially optimal mistakes? Debiasing Covid-19 mortality risk perceptions and prosocial behavior. *Journal of Economic Behavior & Organization 183*, 456–480.

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.

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.

Brañas Garza, P., D. Jorrat, A. Alfonso-Costillo, A. M. Espin, T. García, and J. Kovářík (2020). Exposure to the Covid-19 pandemic and generosity. Technical report, Munich Personal Repec Archive.

Breza, E., F. C. Stanford, M. Alsan, B. Alsan, A. Banerjee, A. G. Chandrasekhar, S. Eichmeyer, T. Glushko, P. Goldsmith-Pinkham, K. Holland, et al. (2021). Doctors' and nurses' social media ads reduced holiday travel and covid-19 infections: A cluster randomized controlled trial. Technical report, National Bureau of Economic Research.

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., 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 (2021). Opinions as facts. Technical report, University of Chicago, Becker Friedman Institute for Economics Working Paper.

Campos-Mercade, P., A. N. Meier, F. H. Schneider, and E. Wengström (2021). Prosociality predicts health behaviors during the Covid-19 pandemic. *Journal of Public Economics 195*, 104367.

Capistran, M. A., A. Capella, and J. A. Christen (2021). Forecasting hospital demand in metropolitan areas during the current Covid-19 pandemic and estimates of lockdown-induced 2nd waves. *PloS one 16*(1).

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

de Véricourt, F., H. Gurkan, and S. Wang (2021). Informing the public about a pandemic. Forthcoming in *Management Science*.

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.

Falco, P. and S. Zaccagni (2020). Promoting social distancing in a pandemic: Beyond the good intentions. Technical report, SSRN 3696804.

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.

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*.

Kerwin, J. T. (2020). Scared straight or scared to death? fatalism in response to disease risks. Technical report, University of Minnesota, Department of Applied Economics.

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.

Table 1:
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)
$\ln(\text{Reported Deaths})_{s,t-1}$	-0.0852***	-0.0821***	-0.0841**	-0.144**	-0.149**	-0.172**
	(0.0244)	(0.0248)	(0.0337)	(0.0571)	(0.0565)	(0.0817)
$\ln(\text{Occurred Deaths})_{s,t-1}$	0.0103	0.0133	0.0179	-0.0254	-0.0300	0.0384
	(0.0303)	(0.0296)	(0.0331)	(0.0713)	(0.0723)	(0.0875)
Controls:						
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
$H_0 : \beta_1 = \beta_2$	0.063	0.065	0.123	0.324	0.328	0.211

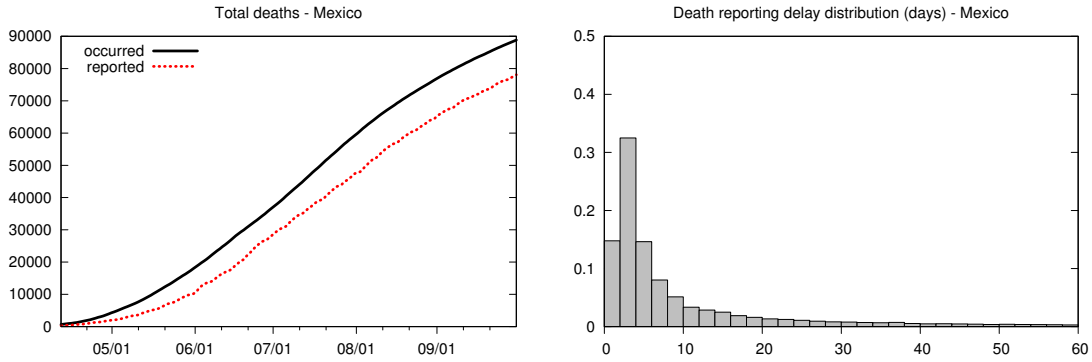
Notes: This table shows how reported and occurred deaths correlate with the growth in Covid-19 cases as reported in the data. Observations are at the state-week level and we show estimates of equation 1. Columns 1-3 show the growth rate from week t to $t+1$, while columns 4-6 consider the rate from t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

Table 2:
Estimates of Informational Treatments on Perceptions and Behavior

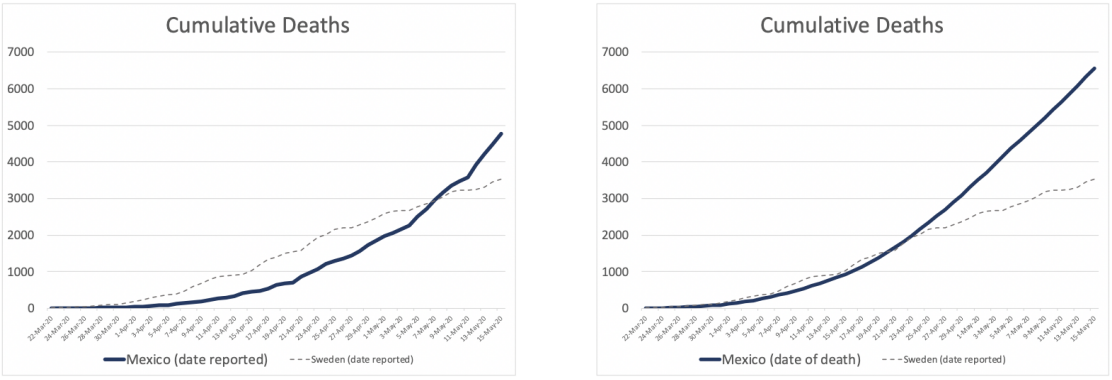
	(1) Full sample	(2) Low prior	(3) High prior	(4) Full sample	(5) Low prior	(6) High prior
Panel A: Pandemic's evolution	Compared to Sweden					
	Faster or much faster			Much faster		
Information by date occurred	0.195***	0.169***	0.198***	0.258***	0.220***	0.266***
	(0.024)	(0.035)	(0.034)	(0.030)	(0.044)	(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. Regressions include control variables as listed in Table S2. Robust standard errors are reported in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Figure 1:
Delays in Death Reports in Mexico and the Information Treatments in the Survey



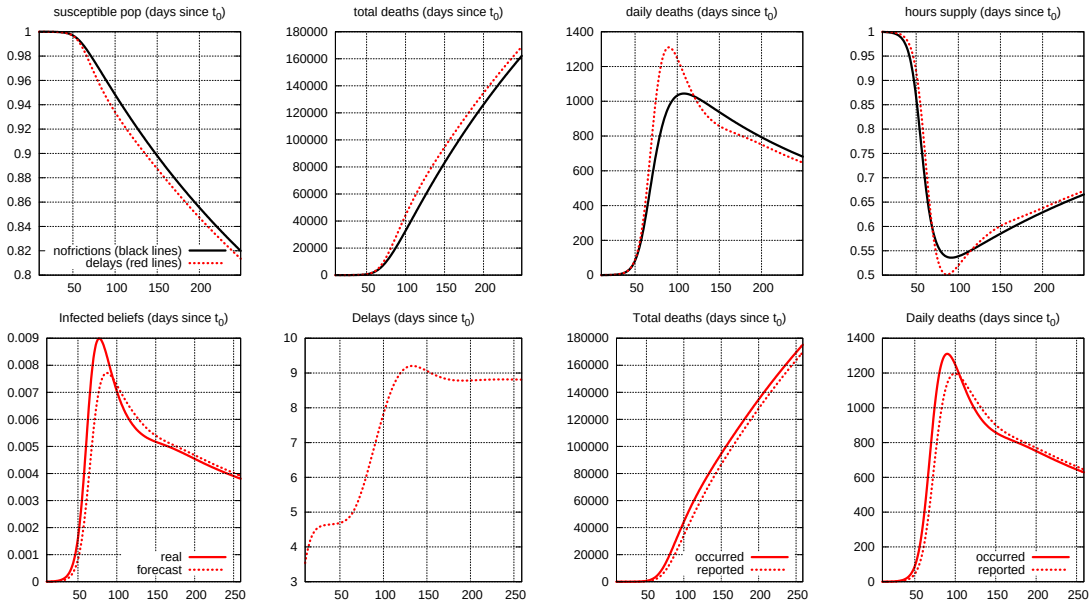
(a) Reporting delays: deaths by date reported and by date occurred



(b) Treatment arms in the survey: cumulative deaths by date reported and date occurred

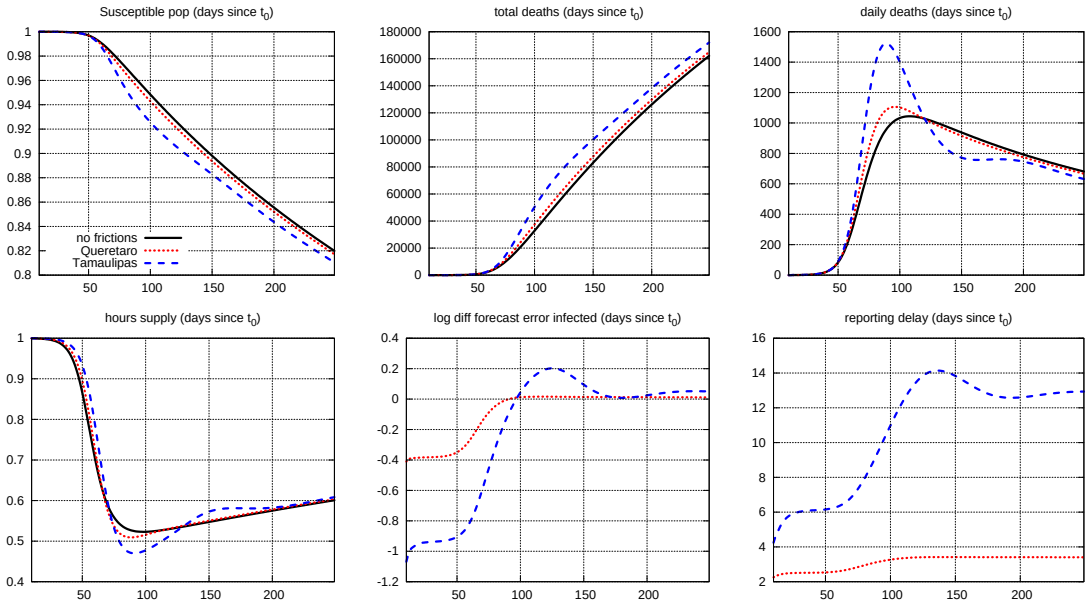
Notes: These plots depict reporting delays in Mexico and the information treatments in the survey. Focusing on the top panel, the plot on the left shows total deaths in Mexico up to September 30, 2020. The solid line corresponds to cumulative death counts based on the date of occurrence, while the dotted 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. In the bottom panel, we show the information treatments. The plot on the left shows cumulative deaths in Mexico based on the date they were reported. The plot on the right shows them by when they actually occurred. 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.

Figure 2:
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 4.1. The top row shows results for a situation without reporting delays (solid line) and with delays calibrated to the Mexican data (dashed line). The plots in the top panel show the mass of susceptible individuals, total deaths, daily number of deaths, and hours supplied outside the home. Focusing on the bottom row, the plot on the furthest left shows the beliefs agents have about the mass of infected individuals over time from the onset of the epidemic. We then show the evolution of delays, and both total and daily deaths as occurred and as reported in the scenario with reporting delays.

Figure 3:
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 4.1. The top panel and the leftmost plot in the bottom panel show results for a situation without reporting delays (solid line), with short delays as in Querétaro (short-dashed line), and large 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. The last two plots in the bottom row compute the forecast error of the mass of infected individuals for the short and long delays scenarios as well as the reporting delays in each case.

Supplementary Information Appendix

A Additional Details on Covid-19 Data in Mexico

We illustrate the government data on Covid-19 reports provided on a daily basis by the Mexican government with an example for clarity in Table S1. 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.

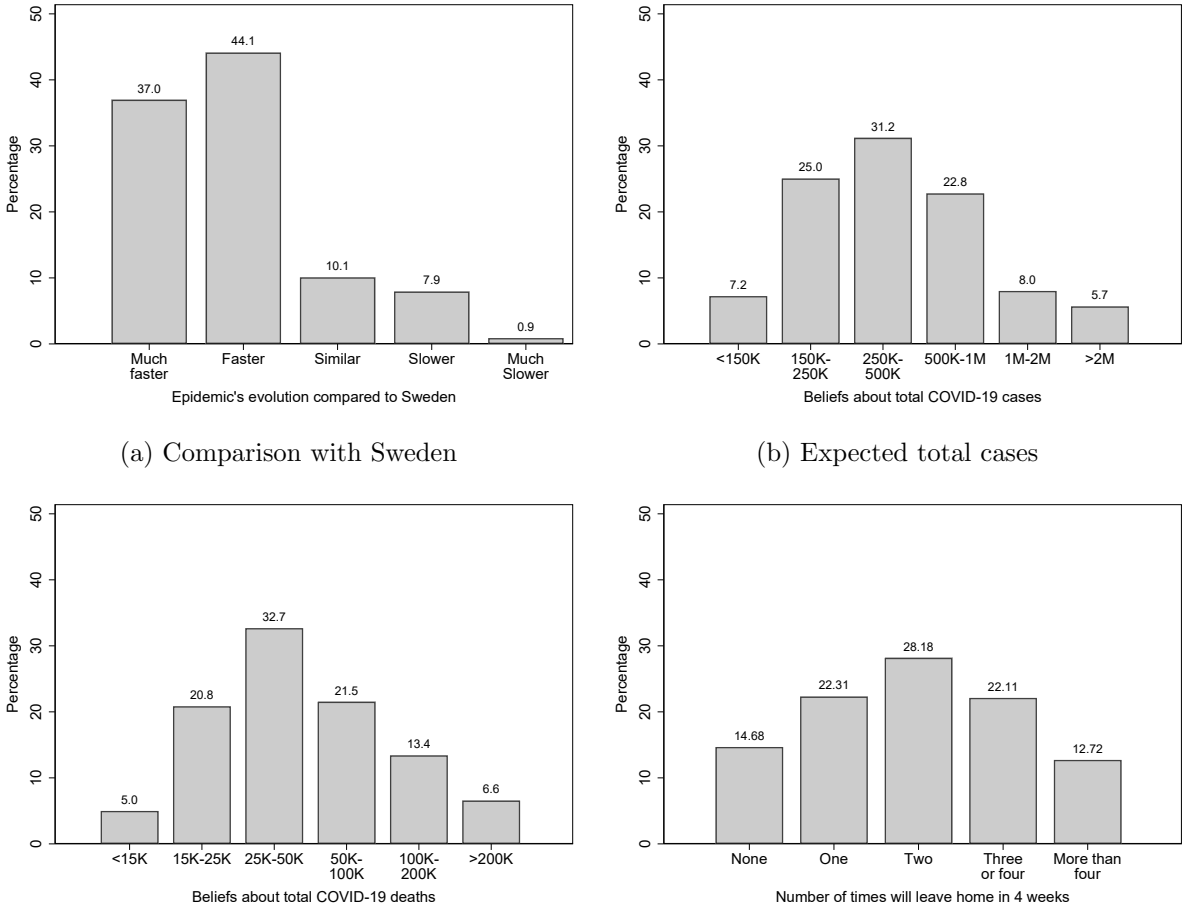
Table S1:
Example of the Government Database of Covid-19 Patients

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.

B Supplementary Figures and Tables for the Survey

Figure S1:
Histograms of Risk Perceptions and Behavior



(a) Comparison with Sweden

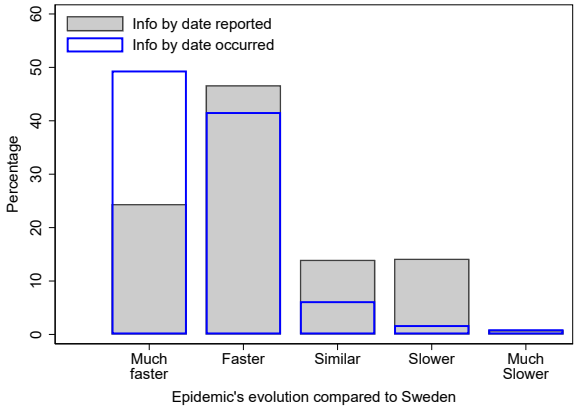
(b) Expected total cases

(c) Expected total deaths

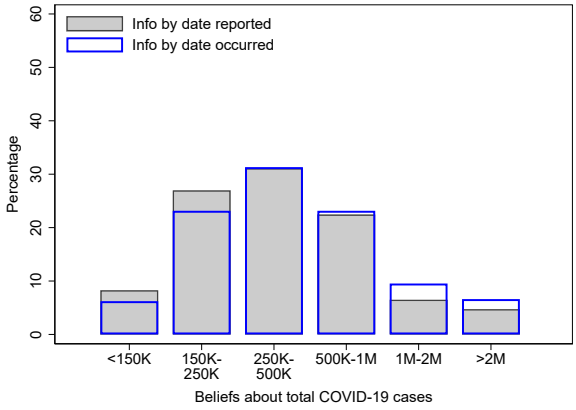
(d) Times leaving the house in 4 weeks

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.

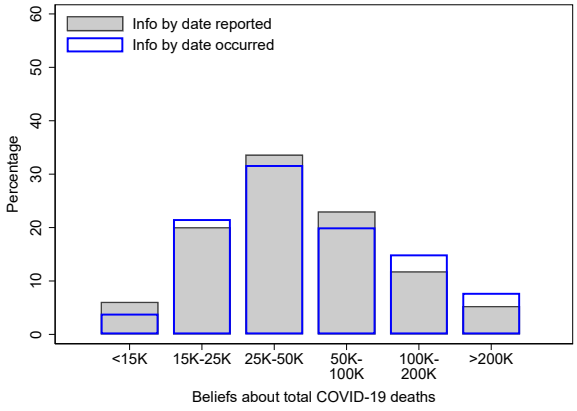
Figure S2:
Histograms of Risk Perceptions and Behavior by Informational Treatments



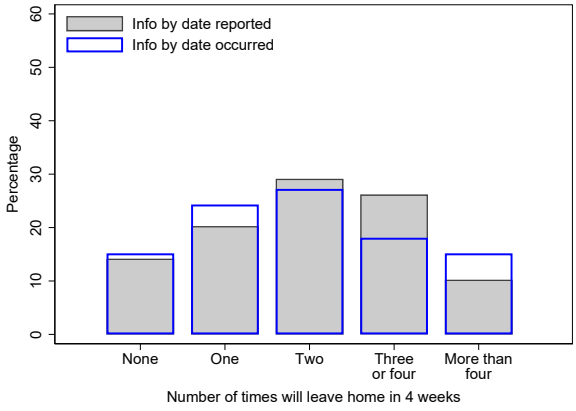
(a) Comparison with Sweden



(b) Expected total cases



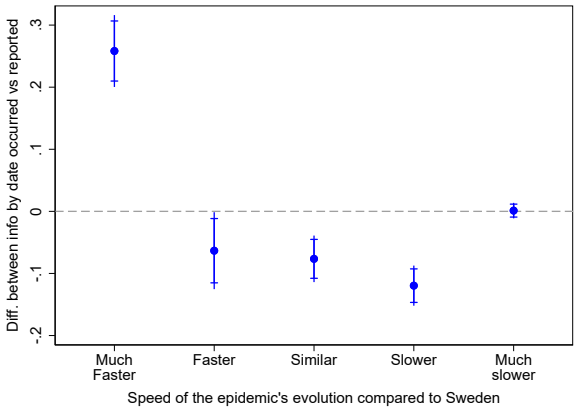
(c) Expected total deaths



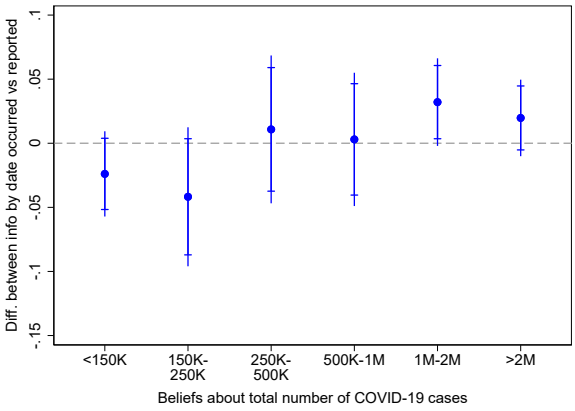
(d) Times leaving the house in 4 weeks

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.

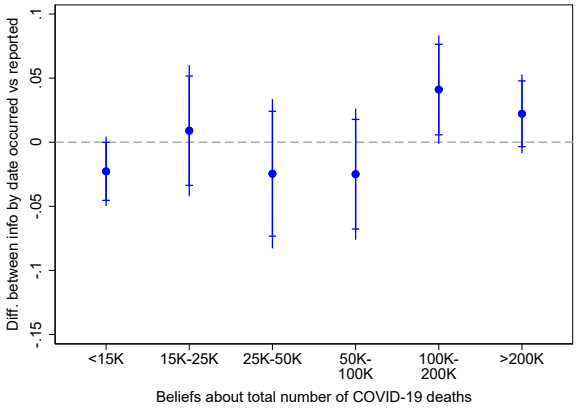
Figure S3:
Estimates of Informational Treatments for Full Set of Responses



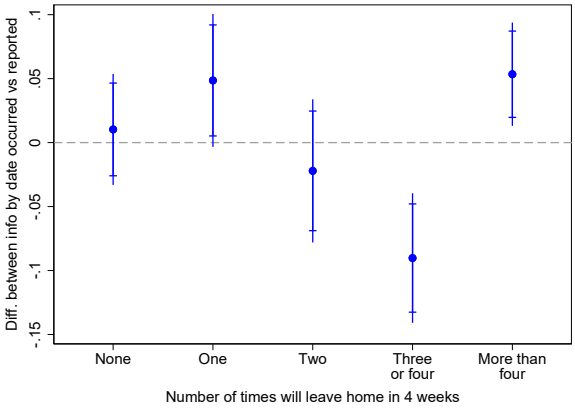
(a) Comparison with Sweden



(b) Expected total cases



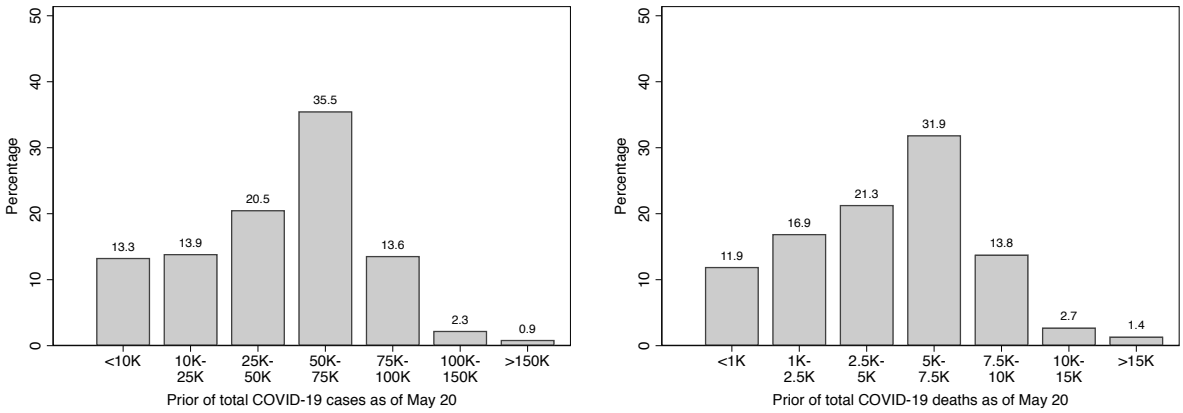
(c) Expected total deaths



(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 S4:
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 S2:
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)
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)
Has HH members in vulnerable age groups (50+)	0.640	0.625	-0.015
	(0.481)	(0.485)	(0.030)
Has HH members 50-60 years old	0.461	0.471	0.010
	(0.499)	(0.500)	(0.031)
Has older HH members (60+)	0.313	0.257	-0.056**
	(0.464)	(0.437)	(0.028)
Has HH members 60-70 years old	0.215	0.202	-0.012
	(0.411)	(0.402)	(0.025)
Has HH members over 70 years old	0.159	0.080	-0.080***
	(0.366)	(0.271)	(0.020)
Observations	508	514	1,022

Notes: This table shows means and standard deviations for covariates asked in the survey. 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. An F-test of joint significance yields a p-value of 0.210; a p-value of 0.112 when decomposing vulnerable household members into two age groups; and 0.014 when decomposing into three age groups.
*** p<0.01, ** p<0.05, * p<0.1

C Additional Details and Results on the Model

C.1 Additional details on the model

Updating beliefs based on government information. We assume that the government provides information on current deaths $D_t \neq \tilde{M}_t^{prior}(d)$. 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 dead), the prevalence measured as the number of infected at $t-3$ can be back-traced. Given the structure of the model and the state transitions from (mass susceptible)-(mass dead), 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 $t-3$. 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:

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

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

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

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

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

With this information, agents can forward on the system (mass susceptible)-(mass dead) using $\left\{n(s)^{t-1}, n(i)^{t-1}\right\}$ to get an updated set of beliefs for the masses at t given by $\tilde{M}_t^{posterior}(j)$ for each $j = s, i, c, r, d$, where now $D_t = \tilde{M}_t^{posterior}(d)$. 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 4.1: behavior schedules $\{n(j, t')\}_{t'=t}^{\infty}$ are updated in accordance with the new belief $\tilde{M}_t^{posterior}(j)$ for each state $j = s, i, c, r, d$.

C.2 Details of model calibration for Mexico

We summarize our calibration parameters as stated in the text in Table S3.

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 S5. 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.

C.3 Additional details on results presented in main text

We show some statistics from the simulations presented in the main text. Figure S4 considers the simulation where delays are modeled after the delay distribution for the whole country. Fig-

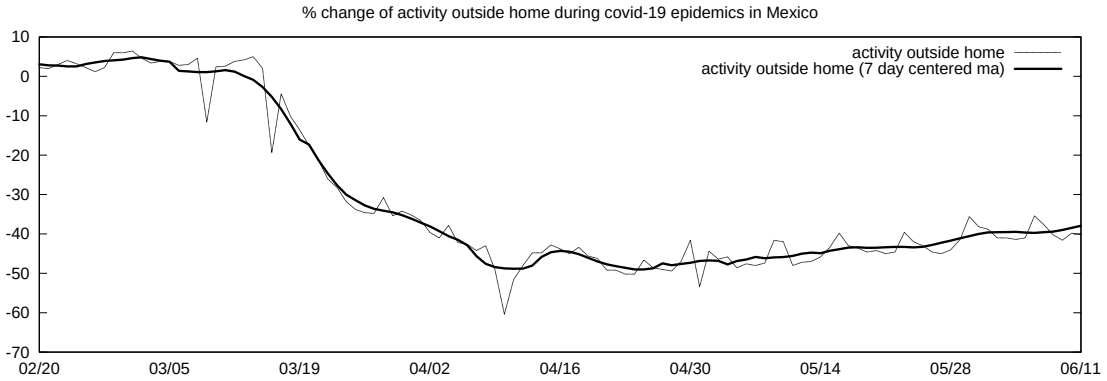
[1] The dataset can be accessed at https://en.www.inegi.org.mx/programas/enut/2014/.
[2] Google Mobility data can be accessed at https://www.google.com/covid19/mobility/.

Table S3:
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.

Figure S5:
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.

ure S5 does the same for the simulations with small and long delays. We also show the empirical distribution of delays in our extreme cases of Querétaro and Tamaulipas in Figure S6.

C.4 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 $n(j,t)$;

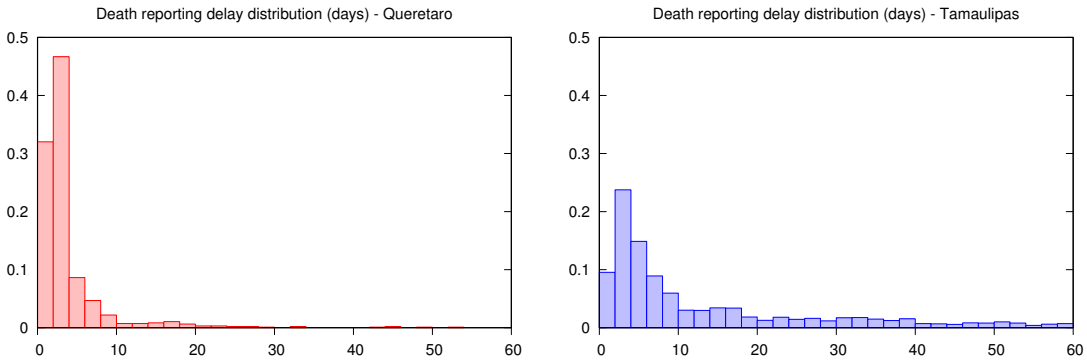
3. Compute the path of $M_t(j)$ for each j;

Table S4:
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 S6:
Distribution of Reporting Delays for Querétaro and Tamaulipas



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

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.5 Robustness over the choice of parameters

Table S6 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).

C.6 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.

Table S5:
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 (Querétaro)						
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 Querétaro) or long delays (Tamaulipas). The "reported deaths" row presents statistics on deaths as reported in the models with delays.

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.6.1 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 S7 and Table S7 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.

Table S6:
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

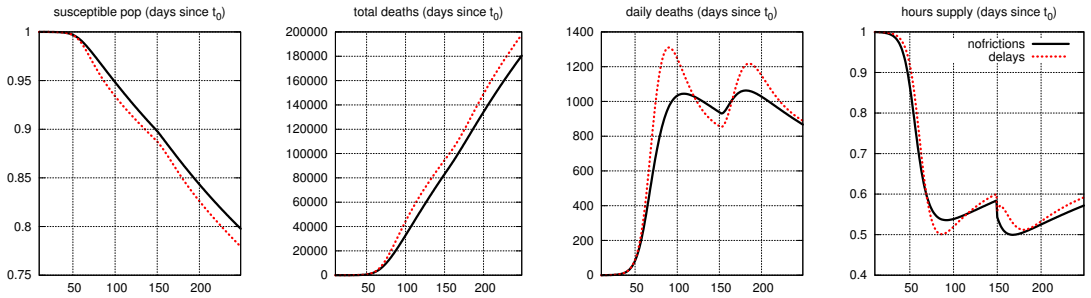
Notes: This table shows results from changing parameters of the model. We consider a higher and lower death rate, infection rate and resolving probability. For each case, we show estimates from model with and without delays in death reporting. We present the estimates for the peak number of infections (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.6.2 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.

Figure S7:
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 4.1. 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 S7:
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.

given by:

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

$$V(b,1) = \max_{n\in(0,1)} \{ u(n) + \beta V(b,2) \}$$

$$V(b,2) = \max_{n\in(0,1)} \{ u(n) + \beta V(b,3) \}$$

$$V(b,3) = \max_{n\in(0,1)} \{ u(n) + \beta V(b,4) \}$$

$$V(b,4) = \max_{n\in(0,1)} \{ u(n) + \beta V(i) \}$$

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

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

$$V(r) = \max_{n\in(0,1)} \{ u(n) + \beta V(r) \}$$

$$V(d) = 0$$

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:

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

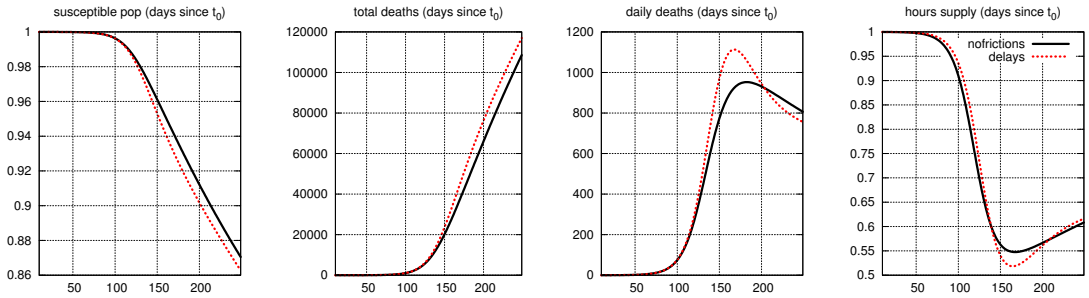
And finally, when faced with a situation that requires belief updating, agents take into account the existence of these intermediate states.

Figure S8 and Table S8 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.

C.6.3 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

Figure S8:
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 4.1. 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 S8:
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.

foresight of this change in medical technology means also a change in the value functions that incorporate such improvement:

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

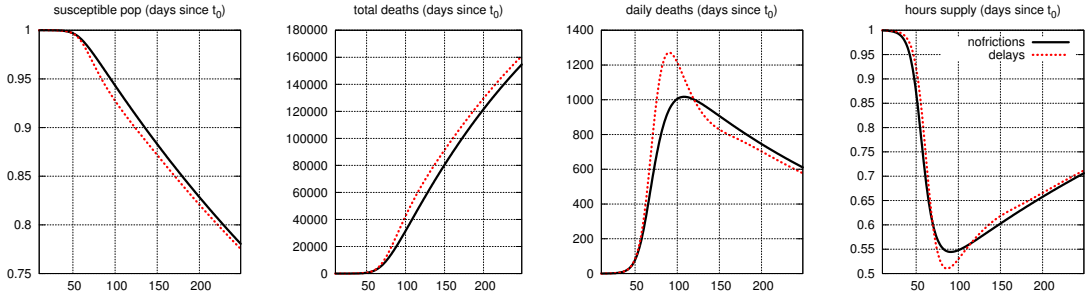
$$V(i,t) = \max_{n\in(0,\bar{n})} \left\{ u(n) + \beta \left[\gamma V(c,t+1) + (1-\gamma) V(i,t+1) \right] \right\}$$

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

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 S9 and Table S9, 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.

Figure S9:
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 4.1. 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.

Table S9:
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.

C.7 Benchmarking against standard epidemiological models for Mexico

Although our model is not specifically designed to replicate the data on the progression of the epidemic for Mexico, we find the comparison of our baseline results with standard epidemic models potentially informative.

We focus on two models used by the Mexican government:

- a simple statistical model that fits a generalized logistic curve to the data (project available at https://salud.conacyt.mx/coronavirus/investigacion/proyectos/gompertz.html)

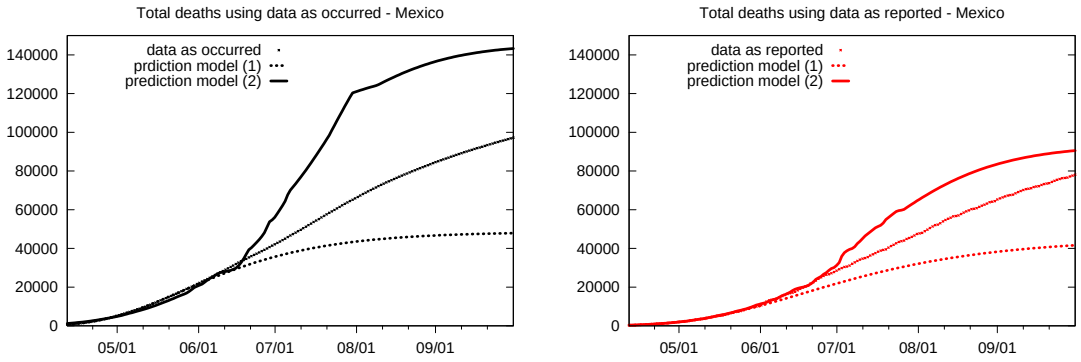
- a version of a standard SIR model with additional compartments, where the exogenous contact rate is adjusted piecewise between dates, and the model parameters are Bayesian-estimated using data for cases and deaths (project available at https://salud.conacyt.mx/coronavirus/investigacion/proyectos/ama.html and more details in Capistran et al. 2021)

To estimate these models, we use data for the whole country until the end of May 2020. Since we highlight delays in this paper, we show results when feeding the models with data as reported and with data as occurred. We do not change any of the calibrated parameters in the SIR model. Also, because this model is based on stochastic realizations of different shocks, we present results based on the median outcome out of 25,000 simulations, as done in Capistran et al. (2021).

Table S10 and Figure S10 present the estimates. The results show that both models can adjust well the data within the sample, that is, until May 2020. The first model, based on a generalized logistic function seems to underestimate outside the sample data predictions, while the second model, based on an estimated version of an SIR model, overestimates the observations from the data. Note how both model predictions dramatically change when using data as reported relative to data as occurred to generate the estimates.

Relative to the baseline model used in this paper, we find that our results are mostly comparable with the standard SIR model when using deaths as occurred, although some important differences should be highlighted. At day 120, the baseline model in our paper predicts 67 thousand deaths while this SIR model predicts 70 thousand deaths. In the data, we observe a cumulative number of 54 thousand deaths by day 120 (July 17th, 2020). By day 200 of the epidemic (October 4th, 2020), the predictions and data observations are 135 thousand and 98 thousand, respectively. Predictions of the two models start diverging for longer horizons. This discrepancy occurs since behavior in the SIR model is fixed, thus a relatively small contact rate estimated in later periods implies that the disease quickly fades away. This is something that does not happen with an equilibrium behavior model as that in our baseline, which instead generates a "slow burn" tension between looser behavior and lower prevalence.

Figure S10:
Dynamics of Covid-19 Deaths in Data and Epidemiological Models



Notes: These graphs show total deaths using data as occurred (left panel) and as reported (right panel). Each series corresponds to the actual data, data generated from predictions of an adjusted generalized logistic model (prediction model 1), and data from the median prediction of 25,000 simulations from an SIR model (prediction model 2).

Table S10:
Predictions of the Epidemic for Mexico: Standard Epidemiological Models and our Baseline Model

	Maximum deaths (daily)	Total deaths (at day 120)	Total deaths (at day 200)	Total deaths (at day 300)
Baseline model with delays	1,310	66,891	134,733	199,559
Adjusted generalized logistic model				
Reported deaths	398	26,853	41,767	44,740
Occurred deaths	585	35,678	47,543	48,591
Median of simulated SIR model				
Reported deaths	3,240	50,649	91,367	96,481
Occurred deaths	3,335	70,128	142,554	145,506

Notes: This table shows comparisons of our baseline model with delays against standard epidemiological models used by the Mexican government. The first model is an adjusted generalized logistic fit. The second model is an SIR model for which we show the median outcome of 25,000 simulations.