The Effect of Large-Scale Anti-Contagion Policies on the Coronavirus (COVID-19) Pandemic

Solomon Hsiang[*1,2], Daniel Allen[†1], Sébastien Annan-Phan[†1,3], Kendon Bell[†1,4], Ian Bolliger[†1,5], Trinetta Chong[†1], Hannah Druckenmiller[†1,3], Andrew Hultgren[†1,3], Luna Yue Huang[†1,3], Emma Krasovich[†1], Peiley Lau[†1,3], Jaecheol Lee[†1,3], Esther Rolf[†1,6], Jeanette Tseng[†1], and Tiffany Wu[†1]

[1]Global Policy Laboratory, Goldman School of Public Policy, UC Berkeley
[2]National Bureau of Economic Research & Centre for Economic Policy Research
[3]Agricultural & Resource Economics, UC Berkeley
[4]Manaaki Whenua – Landcare Research
[5]Energy & Resources Group, UC Berkeley
[6]Electrical Engineering & Computer Science Department, UC Berkeley

First version of text: March 21, 2020
This version of text: March 22, 2020
Data updated: March 19, 2020

[*]Correspondence: shsiang@berkeley.edu. [†] Equal contribution (alphabetical). We thank Brenda R. Chen for her role in initiating this work. Funding: SAP, EK, PL, JT are supported by a gift from the Tuaropaki Trust. TC is supported by an AI for Earth grant from National Geographic and Microsoft. DA, AH, IB are supported through joint collaborations with the Climate Impact Lab. KB is supported by the Royal Society Te Apārangi Rutherford Postdoctoral Fellowship (New Zealand). HD and ER are supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE 1106400 and 1752814, respectively. Opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not reflect the views of supporting organizations. The authors declare no conflicts of interest. All data and code used in this analysis are available at https://github.com/bolliger32/gpl-covid. Updates posted at http://www.globalpolicy.science/covid19.

Abstract

Governments around the world are responding to the novel coronavirus (COVID-19) pandemic[1] with unprecedented policies designed to slow the growth rate of infections. Many actions, such as closing schools and restricting populations to their homes, impose large and visible costs on society. In contrast, the benefits of these policies, in the form of infections that did not occur, cannot be directly observed and are currently understood through process-based simulations.[2–4] Here, we compile new data on 936 local, regional, and national anti-contagion policies recently deployed in the ongoing pandemic across localities in China, South Korea, Iran, Italy, France, and the United States (US). We then apply reduced-form econometric methods, commonly used to measure the effect of policies on economic growth, to empirically evaluate the effect that these anti-contagion policies have had on the growth rate of infections. In the absence of any policy actions, we estimate that early infections of COVID-19 exhibit exponential growth rates of roughly 45% per day. We find that anti-contagion policies collectively have had significant effects slowing this growth, although policy actions in the US appear to be too recent to have a substantial impact since the magnitude of these effects grows over time. Our results suggest that similar policies may have different impacts on different populations, but we obtain consistent evidence that the policy packages now deployed are achieving large and beneficial health outcomes. We estimate that, to date, current policies have already prevented or delayed on the order of eighty-million infections. These findings may help inform whether or when ongoing policies should be lifted or intensified, and they can support decision-making in the over 150 countries where COVID-19 has been detected but not yet achieved high infection rates.[5]

Introduction

The 2019 novel coronavirus[1] (COVID-19) pandemic is forcing societies around the world to make consequential policy decisions with limited information. After containment of the initial outbreak failed, attention turned to implementing large-scale social policies designed to slow contagion of the virus,[6] with the ultimate goal of slowing the rate at which life-threatening cases emerge so as to not exceed the capacity of existing medical systems. In general, these policies aim to decrease opportunities for virus transmission by reducing contact among individuals within or between populations, such as by closing schools, limiting gatherings, and restricting mobility. Such actions are not expected to halt contagion completely, but instead are meant to slow the spread of COVID-19 to a manageable rate. These large-scale policies are developed using epidemiological simulations[2,4,7–17] and a small number of natural experiments in past epidemics.[18] However, the actual impacts of these policies on infection rates in the ongoing pandemic are unknown. Because the modern world has never experienced a pandemic from this pathogen, nor deployed anti-contagion policies of such scale and scope, it is crucial that direct measurements of policy impacts be used alongside numerical simulations in current decision-making.

Populations in almost every country are now currently weighing whether, or when, the health benefits of anti-contagion policies are worth the costs they impose on society. For example, restrictions imposed on businesses are increasing unemployment,[19] travel bans are bankrupting airlines,[20] and school closures may have enduring impacts on affected students.[21] It is therefore not surprising that some populations hesitate before implementing such dramatic policies, particularly when these costs are visible while their health benefits – infections and deaths that would have occurred but instead were avoided or delayed – are unseen. Our objective is to measure this direct benefit; specifically, how much these policies slowed the growth rate of infections. We treat recently implemented policies as hundreds of different natural experiments proceeding in parallel. Our hope is to learn from the recent experience of six countries where the virus has advanced enough to trigger large-scale policy actions, in part so that societies and decision-makers in the remaining 180+ countries can access this information immediately.

Here we directly estimate the effects of local, regional, and national policies on the growth rate of infections across localities within China, South Korea, Iran, Italy, France, and the US (see Figure 1 and Appendix Table A1). We compile publicly available sub-national data on daily infection rates and the timing of policy deployments, including (1) travel restrictions, (2) social distancing through cancellation of events and suspensions of educational/commercial/religious activities, (3) quarantines and lockdowns, and (4) additional policies such as emergency declarations or expansions of paid sick leave, from the earliest available dates to the present (March 18, 2020; see complete descriptions in the Appendix). Because the pandemic is still in its early stages, populations in these countries remain almost entirely susceptible to COVID-19, causing the natural spread of infections

to exhibit almost perfect exponential growth.[7,14,22] The rate of this exponential growth may change daily and is determined by epidemiological factors, such as disease infectivity and contact networks, as well as policies that induce behavior changes.[7,8,22] We cannot experimentally manipulate policies ourselves, but because they are being deployed while the epidemic unfolds, we can measure their impact empirically. We examine how the growth rate of infections each day in a given locality changes in response to the collection of ongoing policies applied to that locality on that day.

We employ well-established "reduced-form" econometric techniques[23,24] commonly used to measure the effect of policies[25,26] or other events (e.g., wars[27] or environmental changes[28]) on economic growth rates. Similarly to early COVID-19 infections, economic output generally increases exponentially with a variable rate that can be affected by policy or other conditions. Unlike process-based epidemiological models,[7–9,12,22,29,30] the reduced-form statistical approach to inference that we apply does not require explicit prior information about fundamental epidemiological parameters or mechanisms, many of which remain unknown in the current pandemic. Rather, the collective influence of these factors is empirically recovered from the data without modeling their individual effects explicitly (see Methods). Prior work on influenza,[31] for example, has shown that such statistical approaches can provide important complementary information to process-based models.

To construct the dependent variable, we transform location-specific, sub-national time-series of infections into first-differences of their natural logarithm, which is the *per day growth rate of infections* (see Methods). We use data from first- or second-level administrative units and data on active or cumulative cases, depending on availability (see Appendix Section 2). We then employ widely-used panel regression models[23,24] to estimate how the daily growth rate of infections changes over time within a location when different combinations of large-scale social policies are enacted (see Methods). Our econometric approach accounts for differences in the baseline growth rate of infections across locations due to differences in demographics, socio-economic status, culture, or health systems across localities within a country; it accounts for systemic patterns in growth rates within countries unrelated to policy, such as the effect of the work-week; it is robust to systematic under-surveillance; and it accounts for changes in procedures to diagnose positive cases (see Methods and Appendix Section 2). The reduced-form statistical techniques we use are designed to measure the total magnitude of the effect of *changes in policy*, without attempting to explain the origin of baseline growth rates or the specific epidemiological mechanisms linking policy changes to infection growth rates (see Methods). Thus, this approach does not provide the important mechanistic insights generated by process-based models; however, it does effectively quantify the key policy-relevant relationships of interest using recent real-world data when fundamental epidemiological parameters are still uncertain.

Results

We estimate that in the absence of policy, early infection rates of COVID-19 grow 45% per day on average, implying a doubling time of approximately two days. Country-specific estimates range from 25.23% per day ($p < 0.05$) in China to 65.04% per day ($p < 0.001$) in Iran, although an estimate only using data from Wuhan, the only Chinese city where a meaningful quantity of pre-policy data is available, is 55% per day ($p < 0.001$). Growth rates in South Korea, Italy, France, and the US are very near the 45% average value (Figure 2A). These estimated values differ from the observed growth rates because the latter are confounded by the effects of policy. In the early stages of most epidemics, a large proportion of the population remains susceptible to the virus, and if the spread of the virus is left uninhibited by policy or behavioral change, exponential growth will continue until the fraction of the susceptible population declines meaningfully.[7,29] This decline results from members of the population leaving the transmission cycle, due to either recovery or death.[29] At the time of writing, the minimum susceptible population fraction in any of the administrative units analyzed is 99.4% of the total population (Lodi, Italy: 1,445 infections in a population of 230,000). This suggests that all administrative units in all six countries would likely be in a regime of uninhibited exponential growth if policies were removed today.

Consistent with predictions from epidemiological models,[2,18,32] we find that the combined effect of all policies within each country reduces the growth rate of infections by a substantial and, except in the US, statistically significant amount (Figure 2B). For example, a locality in Italy with a baseline growth rate of 0.38 (national avg.) that deployed all policy actions used in Italy would be expected to lower its daily growth rate by 0.18 to 0.20. In general, the estimated total effects of policy packages are large enough that they can in principle offset a large fraction of, or even eliminate, the baseline growth rate of infections—although in several countries many localities are not currently deploying the full set of policies. Our estimate for the total growth effect of all US policies is quantitatively substantial (-0.25) but not statistically significant. US estimates are highly uncertain due to the short period of time for which data are available and because the time elapsed since these actions may be too short to observe a significant impact. In China, where policies have been enacted for over seven weeks, we observe that policy impacts have grown over time during the first three weeks of deployment (-0.11 to -0.33). In all other countries except China, we only estimate an average effect for the entire interval of observation, due to the short temporal length of the sample.

The estimates above describe the superposition of all policies deployed in each country, i.e. they represent, for each country, the average effect of policies on infection growth rates that we would expect to observe, if all policies enacted anywhere in the country were implemented simultaneously in a region of the country. We also estimate the effects of individual types of policies or clusters of policies that are grouped based on their similarity in goal (e.g., closing libraries and closing

museums are grouped) or timing (e.g., policies that are generally deployed simultaneously in a certain country). In many cases, our estimates for these effects are statistically noisier than the estimates for all policies combined (presented above) because we are estimating multiple effects simultaneously. Thus, we are less confident in individual estimates and in their relative rankings. Estimated effects differ between countries, and policies are neither identical nor perfectly comparable in their implementation across countries or, in many cases, across different localities within the same country. Nonetheless, overall we estimate that almost all policies likely contribute to slowing the growth rate of infections (Figure 2c), except two policies (social distancing in France and Italy) where point estimates are slightly positive, small in magnitude, and not statistically different from zero.

We combine the estimates above with our data on the timing of hundreds of policy deployments to estimate the total effect to date of all policies in our sample. To do this, we use our estimates above to predict the growth rate of infections in each locality on each day given the policies in effect at that location on that date (Figure 3, blue markers). We then use the same model to predict what counterfactual growth rates would be on that date if all policies were removed (Figure 3, red), which we refer to as a "no policy" scenario. The difference between these two predictions is our estimated effect that all anti-contagion policies actually deployed had on the growth rate of infections on that date. We estimate that since the beginning of our sample, on average, all anti-contagion policies combined have slowed the average daily growth rate of infections -0.166 per day ($\pm 0.015, p < 0.001$) in China, -0.276 ($\pm 0.066, p < 0.001$) in South Korea, -0.158 ($\pm 0.071, p < 0.05$) in Italy, -0.292 ($\pm 0.037, p < 0.001$) in Iran, -0.132 ($\pm 0.053, p < 0.05$) in France and -0.044 ($\pm 0.059, p = 0.45$) in the US. Taken together, these results suggest that anti-contagion policies currently deployed in the first five countries are achieving their intended objective of slowing the pandemic, broadly confirming epidemiological simulations. We estimate that anti-contagion policies have not yet had a substantial nor significant impact suppressing overall infection growth rates in the US.

At a particular moment in time, the total number of COVID-19 infections depends on the growth rate of infections on all prior days. Thus, persistent decreases in growth rates have a compounding effect on total infections, at least until a shrinking susceptible population slows growth through a different mechanism. To provide a sense of scale and context for our main results in Figures 2 and 3, we integrate the growth rate of infections in each locality from Figure 3 to estimate total infections to date, both with actual anti-contagion policies and in the "no policy" counterfactual scenario. To account for the declining size of the susceptible population in each administrative unit, we couple our econometric estimates for the effects of policies to a simple Susceptible-Infected-Removed (SIR) model of infectious disease dynamics[7,22] (see Methods). This allows us to extend our projections beyond the initial exponential growth phase of infections, a threshold which our results suggest would currently be exceeded in several countries in the "no policy" scenario.

Our results suggest that ongoing anti-contagion policies have already substantially reduced the

number of COVID-19 infections observed in the world today (Figure 4). Our central estimates suggest there would be roughly 74-million more cumulative cases in China, 5-million more in South Korea, 1.2-million more in Italy, 2.6-million more in Iran, 650,000 more in France, and 20,000 more in the US had these countries never enacted any anti-contagion policies since the start of the pandemic. The relative magnitudes of these impacts partially reflects the intensity and extent of policy deployment (e.g. how many localities deployed policies) and the duration for which they have been applied. Several of these estimates are subject to large uncertainties (see intervals in Figure 4).

Discussion

Overall, our results indicate that large-scale anti-contagion policies are achieving their intended objective of slowing the growth rate of COVID-19 infections. Because infection rates in the countries we study would have initially followed rapid exponential growth had no policies been applied, our results suggest that these ongoing policies are currently providing large health benefits. For example, we estimate that there would be roughly 621× the current number of infections in South Korea, 36× in Italy, and 153× in Iran if large-scale policies had not been deployed during the early weeks of the pandemic. Consistent with process-based simulations of COVID-19 infections,[2,4,10–12,14,17,29] our empirical analysis of existing policies indicates that seemingly small delays in policy deployment likely produce dramatically different health outcomes.

While the quantity of currently available data poses challenges to our analysis, our aim is to use what limited data exist to estimate the first-order impacts of unprecedented policy actions in an ongoing global crisis. As more data become available, empirical research findings will become more precise and may capture more complex interactions. For example, this analysis does not account for potentially important interactions between populations in nearby localities,[7,33] nor the structure of mobility networks.[3,4,10,12,17,34] Nonetheless, we hope the results we are able to obtain at this early stage of the pandemic can support critical decision-making, both in the countries we study and in the other 150+ countries where COVID-19 infections have been reported.

Based on our results from China, where the most post-policy time has elapsed and where a relatively uniform set of policies were imposed during a narrow window of time, it appears that roughly three weeks are required for policies to achieve their full effect. In other countries, these temporal dynamics are more difficult to disentangle with currently available data, in part because less post-policy data is available and also because countries continue to deploy new policies, making it more challenging to precisely measure the lagged effects of earlier policies. Future work should investigate these timing changes after more time has passed and new data become available.

A key advantage of our reduced-form "top down" statistical approach is that it captures the real-world behavior of affected populations without requiring that we explicitly model all underlying

mechanisms and processes. This property is useful in early stages of the current pandemic when many process-related parameters remain unknown. However, our results cannot and should not be interpreted as a substitute for process-based epidemiological models specifically designed to provide guidance in public health crises. Rather, our results complement existing models, for example, by helping to calibrate key model parameters. We believe both forward-looking simulations and backward-looking empirical evaluations should be used to inform decision-making.

Here we have focused our analysis on large-scale social policies, specifically, to understand their impact on infection rate growth within a locality. However, contact tracing, international travel restrictions, and medical resource management, along with many other policy decisions, will play key roles in the global response to COVID-19. Our results do not speak to the efficacy of these other policies.

Our analysis accounts for some known changes in the availability of testing for COVID-19 and changes in testing procedures; however, it is likely that other unobserved changes in patterns of testing could affect our results. For example, if growing awareness of COVID-19 caused an increasing fraction of infected individuals to be tested over time, then unadjusted infection growth rates later in our sample would be biased upwards. Because an increasing number of policies are active later in these samples as well, this bias would cause our current findings to understate the overall effectiveness of anti-contagion policies.

It is also possible that changing public information during the period of our study has some unknown effect on our results. If individuals alter their behavior in response to new information unrelated to anti-contagion policies, such as news reports about COVID-19, this could alter the growth rate of infections and thus affect our estimates. Because the quantity of new information is increasing over time, if this information reduces infection growth rates, it would cause us to overstate the effectiveness of anti-contagion policies. We note, however, that if public information is increasing in response to policy actions, then it should be considered a pathway through which policies alter infection growth, not a form of bias. Investigating these potential effects is beyond the scope of this analysis, but it is an important topic for future investigations.

Lastly, we note that the results presented here are not sufficient, on their own, to determine which anti-contagion policies are ideal for individual populations, nor whether the social costs of individual policies are larger or smaller than the social value of their health benefits. Computing a full value of health benefits also requires understanding how different growth rates of infections and total active infections affect mortality rates, as well as determining a social value for all of these impacts. Furthermore, this analysis does not quantify the sizable social costs of anti-contagion policies, a critical topic for future investigations.

Methods

Data Collection and Processing

We have provided a brief summary of our data collection processes here (see Appendix Section 2 for more details, including access dates). Epidemiological and policy data for each of the six countries in our sample were collected from a variety of in-country data sources, including government public health websites, regional newspaper articles, and Wikipedia crowd-sourced information. The available epidemiological and policy data varied across the six countries, and preference was given to collecting data at the most granular administrative unit level. The country-specific panel datasets are at the region level in France, the state level in the US, the province level in South Korea, Italy and Iran, and the city level in China. Below, we describe our data sources.

China We acquired epidemiological data from an open source GitHub project[1] that scrapes time series data from Ding Xiang Yuan. We extended this dataset back in time to January 10 by manually collecting official daily statistics from the central and provincial (Hubei, Guangdong, and Zhejiang) Chinese government websites. We compiled policies by collecting data on the start dates of travel bans and lockdowns at the city-level from the "2020 Hubei lockdowns" Wikipedia page[2], the Wuhan Coronavirus Timeline project on Github[3], and various other news reports. As we suspect that most Chinese cities have been treated by at least one anti-contagion policy, due to their reported trends in infections, we have dropped cities where we cannot find a policy deployment date to avoid miscategorizing the policy status of cities.

South Korea We manually collected and compiled the epidemiological dataset in South Korea, based on provincial government reports, policy briefings, and news articles. We compiled policy actions from press releases from the Korean Centers for Disease Control and Prevention (KCDC), the Ministry of Foreign Affairs, local governments' websites, and news articles.

Iran We used epidemiological data from the table "New COVID-19 cases in Iran by province"[4] in the "2020 coronavirus pandemic in Iran" Wikipedia article, which have been compiled from the data provided on the Iranian Ministry of Health website (in Persian). We relied on news media reporting and two timelines of pandemic events in Iran[5] [6] to collate policy data.

[1] https://github.com/BlankerL/DXY-COVID-19-Data
[2] https://en.wikipedia.org/wiki/2020_Hubei_lockdowns
[3] https://github.com/Pratitya/wuhan2020-timeline
[4] https://en.wikipedia.org/wiki/2020_coronavirus_pandemic_in_Iran
[5] https://www.thinkglobalhealth.org/article/updated-timeline-coronavirus
[6] https://en.wikipedia.org/wiki/2020_coronavirus_pandemic_in_Iran

Italy We utilized epidemiological data from the GitHub repository[7] maintained by the Italian Department of Civil Protection (Dipartimento della Protezione Civile). For policies, we primarily relied on the English version of the COVID-19 dossier "Chronology of main steps and legal acts taken by the Italian Government for the containment of the COVID-19 epidemiological emergency" written by the Department of Civil Protection (Dipartimento della Protezione Civile)[8].

France We used the region-level epidemiological dataset provided by France's government website[9] and supplemented it with scraped number of confirmed cases by region on France's public health website, which is updated daily.[10] We obtained data on France's policy response to the COVID-19 pandemic from the French government website,[11] press releases from each regional public health site,[12] and Wikipedia[13].

United States We used state-level epidemiological data from the GitHub repository[14] associated with the interactive dashboard from Johns Hopkins University (JHU). For policy responses, we relied on a number of sources, including the U.S. Center for Disease Control (CDC), individual state health departments, as well as various press releases from county and city-level government or media outlets.

Policy Data Policies in administrative units were coded as binary variables, where the policy is coded as either 1 (after the date that the policy was implemented, and before it is removed) or 0 otherwise, for the affected administrative units. There were instances when a policy implementation only affected a portion of the administrative units (e.g. half of the counties within the state). In an attempt to accurately represent the locality and impact of policy implementation, policy variables were weighted by the percentage of population within the administrative unit that was treated by the policy. The most recent estimates available of population data for countries' administrative units were used (see the Population Data section in the Appendix). Additionally, in order to standardize policy types across countries, we mapped country-specific policies to one of our broader policy categories used as variables in our analysis. In this exercise, we collected 130 policies for China, 37 for South Korea, 195 for Italy, 26 for Iran, 59 for France, and 498 for the United States (see Appendix Table A1).

[7] https://github.com/pcm-dpc/COVID-19

[8] http://www.protezionecivile.it/documents/20182/1227694/Summary+of+measures+taken+against+the+spread+of+C-19/c16459ad-4e52-4e90-90f3-c6a2b30c17eb

[9] https://www.data.gouv.fr/en/datasets/fr-sars-cov-2/

[10] https://www.santepubliquefrance.fr/maladies-et-traumatismes/maladies-et-infections-respiratoires/infection-a-coronavirus/articles/infection-au-nouveau-coronavirus-sars-cov-2-covid-19-france-et-monde

[11] https://www.gouvernement.fr/info-coronavirus

[12] https://www.ars.sante.fr/

[13] https://fr.wikipedia.org/wiki/Pand%C3%A9mie_de_maladie_%C3%A0_coronavirus_de_2020_en_France

[14] https://github.com/CSSEGISandData/COVID-19

Epidemiological Data We collected information on cumulative confirmed cases, cumulative recoveries, cumulative deaths, active cases, and any changes to domestic COVID-19 testing regimes. For our regression analysis (Figure 2), we use active cases when they are available (for China and South Korea) and cumulative confirmed cases otherwise. We document quality control steps in detail in Appendix Section 2. Notably, for China and South Korea we acquire more granular data than the the data hosted on the John Hopkins University (JHU) interactive dashboard[15]; we confirm that the number of confirmed cases closely match between the two data sources (see Appendix Figure A2). To conduct the econometric analysis, we merge the epidemiological and policy data to form a single data set for each country.

Econometric analysis

Reduced-Form Approach The reduced-form econometric approach that we apply here is a "top down" approach that describes the behavior of aggregate outcomes y in data (here, infection rates). This approach can identify plausibly causal effects[23,24] induced by exogenous changes in independent policy variables z (e.g. school closure) without explicitly describing all underlying mechanisms that link z to y and without observing intermediary variables x (e.g. behavior) that might link z to y nor other determinants of y unrelated to z (e.g. demographics), denoted w. Let $f(\cdot)$ describe a complex and unobserved process that generates infection rates y:

$$y = f(x_1(z_1, \ldots, z_K), \ldots, x_N(z_1, \ldots, z_K), w_1, \ldots, w_M) \tag{1}$$

Process-based epidemiological models aim to capture elements of $f(\cdot)$ explicitly, and then simulate how changes in z, x, or w affect y. This approach is particularly important and useful in forward-looking simulations where future conditions are likely to be different than historical conditions. However, a challenge faced by this approach is that we may not know the full structure of $f(\cdot)$, for example if a pathogen is new and many key biological and societal parameters remain uncertain. Crucially, we may not know the effect that large-scale policy (z) will have on behavior ($x(z)$) or how this behavior change will affect infection rates ($f(\cdot)$).

Alternatively, one can differentiate Equation 1 with respect to the kth policy z_k:

$$\frac{\partial y}{\partial z_k} = \sum_{j=1}^{N} \frac{\partial y}{\partial x_j} \frac{\partial x_j}{\partial z_k} \tag{2}$$

which describes how changes in the policy affects infections through all N potential pathways mediated by $x_1, \ldots, x_N$. Usefully, Equation 2 does not depend on w. If we can observe y and z directly and estimate $\frac{\partial y}{\partial z_k}$ with data, then intermediate variables x also need not be observed

[15] `https://github.com/CSSEGISandData/COVID-19`

nor modeled. The reduced-form econometric approach[23,24] thus attempts to measure $\frac{\partial y}{\partial z_k}$ directly, exploiting exogenous variation in policies z.

Model Active infections grow exponentially during the initial phase of an epidemic, when the proportion of immune individuals in a population is near zero. Assuming a simple Susceptible-Infected-Recovered (SIR) disease model (e.g. ref. [22]), the growth in infections during the early period is

$$\frac{dI}{dt} = (S\beta - \gamma)I_t \underset{S\to 1}{=} (\beta - \gamma)I_t, \tag{3}$$

where I_t is the number of infected individuals at time t, β is the transmission rate (new infections per day per infected individual), γ is the removal rate (proportion of infected individuals recovering or dying each day) and S is the fraction of the population susceptible to the disease. The second equality holds in the limit $S \to 1$, which describes the current conditions during the beginning of the COVID-19 pandemic. The solution to this ordinary differential equation is the exponential function

$$\frac{I_{t_2}}{I_{t_1}} = e^{g\cdot(t_2 - t_1)}, \tag{4}$$

where the growth rate $g = \beta - \gamma$ and t_1 are the initial conditions. Taking the natural logarithm and rearranging, we have

$$\log(I_{t_2}) - \log(I_{t_1}) = g \cdot (t_2 - t_1). \tag{5}$$

Anti-contagion policies are designed to alter g, through changes to β, by reducing contact between susceptible and infected individuals. Holding the time-step between observations fixed at one day ($t_2 - t_1 = 1$), we thus model g as a time-varying outcome that is a linear function of a time-varying policy

$$g_t = \log(I_t) - \log(I_{t-1}) = \theta_0 + \theta \cdot policy_t + \epsilon_t, \tag{6}$$

where θ_0 is the average growth rate absent policy, $policy_t$ is a binary variable describing whether a policy is deployed at time t, and θ is the average effect of the policy on growth rate g. ϵ_t is a mean-zero disturbance term that captures inter-period changes not described by $policy_t$. Using this approach, infections each day are treated as the initial conditions for integrating Equation 4 through to the following day.

We compute the first differences $\log(I_t) - \log(I_{t-1})$ using active infections where they are available, otherwise we use cumulative infections, noting that they are almost identical during this early period (except in China, where we use active infections). We then match these data to policy variables that we construct using the novel data sets we assemble and apply a reduced-form approach to estimate a version of Equation 6, although the actual expression has additional terms detailed below.

Estimation To estimate a multi-variable version of Equation 6, we estimate a separate regression for each country c. Observations are for sub-national units indexed by i observed for each day t. Because not all localities began testing for COIVD-19 on the same date, these samples are unbalanced panels. To ensure data quality, we restrict our analysis to localities after they have reported at least ten cumulative infections.

We estimate a multiple regression version of Equation 6 using ordinary least squares. We include a vector of sub-national unit-fixed effects θ_0 (i.e. varying intercepts captured as coefficients to dummy variables) to account for all time-invariant factors that affect the local growth rate of infections, such as differences in demographics, socio-economic status, culture, or health systems.[24] We include a vector of day-of-week-fixed effects δ to account for weekly patterns in the growth rate of infections that are common across locations within a country. We include a separate single-day dummy variable each time there is an abrupt change in the availability of COVID-19 testing or a change in the procedure to diagnose positive cases. Such changes generally manifest as a discontinuous jump in infections and a re-scaling of subsequent infection rates (e.g. See China in Figure 1), effects that are flexibly absorbed by a single-day dummy variable because the dependent variable is the first-difference of the logarithm of infections. Denote the vector of these testing dummies μ.

Lastly, we include a vector of P_c country-specific policy variables for each location and day. These policy variables take on values between zero and one (inclusive) where zero indicates no policy action and one indicates a policy is fully enacted. In cases where a policy variable captures the effects of collections of policies (e.g. museum closures and library closures), a binary policy variable is computed for each, then they are averaged, so the coefficient on these variables are interpreted as the effect if all policies in the collection are fully enacted. In some cases (for Italy and the US), policy data is available at a more spatially granular level than infection data (e.g. city policies and state-level infections in the US). In these cases, we code binary policy variables at the more granular level and use population-weights to aggregate them to the level of the infection data. Thus, policy variables may take on continuous values between zero and one, with a value of one indicating that the policy is fully enacted for the entire population.

For each country, our general multiple regression model is thus

$$g_{cit} = \log(I_{cit}) - \log(I_{ci,t-1}) = \theta_{0,ci} + \delta_{ct} + \mu_{cit} + \sum_{p=1}^{P_c} (\theta_{cp} \cdot policy_{pcit}) + \epsilon_{cit} \tag{7}$$

where observations are indexed by country c, sub-national unit i, and day t. The parameters of interest are the country-by-policy specific coefficients θ_{pc}. We verify that our residuals ϵ_{cit} are approximately normally distributed (Appendix Figure A1) and we estimate uncertainty over all parameters by clustering our standard errors at the day level.[23] This approach non-parametrically

accounts for arbitrary forms of spatial auto-correlation or systematic misreporting in regions of a country on any given day (it generates larger estimates for uncertainty than clustering by i). When we report the effect of all policies combined (e.g. Figure 2B) we are reporting the sum of coefficent estimates for all policies $\sum_{p=1}^{P_c} \theta_{cp}$, accounting for the covariance of errors in these estimates when computing the uncertainty of this sum.

Note that our estimates of θ and θ_0 in Equation 7 are robust to systematic under-reporting of infections, a major concern in the ongoing pandemic, due to the construction of our dependant variable. If only a fraction ψ of infections are being reported such that we observe $\tilde{I} = \psi I$ rather an actual infections I, then the left-hand-side of Equation 7 will be

$$\begin{aligned}
\log(\tilde{I}_t) - \log(\tilde{I}_{t-1}) &= \log(\psi I_t) - \log(\psi I_{t-1}) \\
&= \log(\psi) - \log(\psi) + \log(I_t) - \log(I_{t-1}) \\
&= \log(I_t) - \log(I_{t-1}) = g_t
\end{aligned}$$

and is therefore unaffected by the under-reporting. Thus systematic under-reporting does not affect our estimates for the effects of policy θ.

There are some country-specific adjustments to Equation 7 due to idiosyncratic differences between samples. In China, we code policy parameters using weekly lags based on the date that the policy is first implemented in locality i. As discussed in the main text, this is done to understand the temporal dynamics of the response to policy in the one country where policy has been enacted the longest and in the most consistent way. Weekly lags are used because the incubation period COVID-19 is thought to be 5-6 days.[4] Econometrically, this means the effect of a policy implemented one week ago is allowed to differ arbitrarily from the effect of a policy implemented two weeks ago, etc. These effects are all estimated simultaneously. Also in China, we omit day-of-week effects because there is no evidence to suggest they are present in the data – this could be due to the fact that the outbreak of COVID-19 began during a national holiday and workers never returned to work. In Iran, we estimate a separate effect of policies implemented in Tehran that is allowed to differ from the effect in other locations by creating Tehran-specific dummy variable that is interacted with both policy variables. This is implemented because of the stark and significantly different effect of policies in Tehran relative to effects in other parts of the country.

Projections

Daily growth rates of infections To estimate the instantaneous daily growth rate of infections if policies were removed, we obtain fitted values from Equation 7 and compute a predicted value for the dependent variable when all P_c policy variables are set to zero. Thus, these estimated growth rates $\hat{g}_{cit}^{no\ policy}$ capture the effect of all locality-specific factors on the growth rate of infections (e.g.

demographics), day-of-week-effects, and adjustments based on the way in which infection cases are reported. This counterfactual does not account for changes in information that are triggered by policy deployment, since those should be considered a pathway through which policies affect outcomes, as discussed in the main text. When we report an average "no policy" growth rate of infections (Figure 2A), it is the average value of these predictions for all observations in the original sample. Location-and-day specific counterfactual predictions ($\hat{g}_{cit}^{no\ policy}$), accounting for the covariance of errors in estimated parameters, are shown as red markers in Figure 3.

Cumulative infections To provide a sense of scale for the estimated cumulative benefits of effects shown in Figure 3, we link our reduced-form empirical estimates to the key structures in a simple SIR system and simulate this dynamical system from the start of the pandemic to the present in each country. The system is defined as the following:

$$\frac{dS_t}{dt} = -\beta_t S_t I_t \tag{8}$$

$$\frac{dI_t}{dt} = (\beta_t S_t - \gamma) I_t \tag{9}$$

$$\frac{dR_t}{dt} = \gamma I_t \tag{10}$$

where S is the susceptible population and R is the removed population. Here β is a time-evolving parameter, determined via our empirical estimates as described below. Accounting for changes in S becomes increasingly important as the size of cumulative infections ($I_t + R_t$) becomes a substantial fraction of the local subnational population, which occurs in some "no policy" scenarios. Our reduced-form analysis provides estimates for the growth rate of active infections ($\hat{g}$) for each locality and day, in a regime where $S \approx 1$. Thus we know

$$\frac{dI_t}{dt} / I_t \Big|_{S \approx 1} = \hat{g}_t = \beta_t - \gamma \tag{11}$$

but we do not know the values of either of the two right-hand-side terms, which are required to simulate Equations 8-10. To estimate γ, we note that the left-hand-side term of Equation 10 is

$$\frac{dR_t}{dt} \approx \frac{d}{dt}(\text{cumulative_recoveries} + \text{cumulative_deaths})$$

which we can observe in our data for China and South Korea. Computing first differences in these two variables (to differentiate with respect to time), summing them, and then dividing by active cases gives us estimates of γ from Equation 10 (medians: China=0.076, Korea=0.029). These values differ slightly from the classical SIR interpretation of γ because, in the public data we are able to obtain, individuals are coded as "recovered" when they no longer test positive for COVID-

19, whereas in the classical SIR model this occurs when they are no longer infectious. We adopt the average of these two medians, setting $\gamma = .052$. We use medians rather than simple averages because low values for I induce a long right-tail in daily estimates of γ and medians are less vulnerable to this distortion. We then use our empirically based reduced-form estimates of $\hat{g}$ (both with and without policy) combined with Equations 8-11 to project total cumulative cases in all countries, shown in Figure 4. We simulate infections and cases for each administrative unit in our sample beginning on the first day for which we observe 10 or more cases (for that unit) using a time-step of 4 hours. We estimate uncertainty by resampling from the estimated variance-covariance matrix of all parameters.

References

[1] Wu, F. *et al.* A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265–269 (2020).

[2] Ferguson, N. M. *et al.* Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand .

[3] Chinazzi, M. *et al.* The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science* (2020).

[4] Kraemer, M. U. *et al.* The effect of human mobility and control measures on the COVID-19 epidemic in China. preprint, Epidemiology (2020).

[5] WHO novel coronavirus (COVID-19) situation. `https://experience.arcgis.com/experience/685d0ace521648f8a5beeeee1b9125cd`. Accessed: 2020-03-19.

[6] With new state decrees, 1 in 5 americans to be ordered indoors. `https://www.nytimes.com/2020/03/20/world/coronavirus-news-usa-world.html?action=click&module=Spotlight&pgtype=Homepage`. Accessed: 2020-03-20.

[7] Chowell, G., Sattenspiel, L., Bansal, S. & Viboud, C. Mathematical models to characterize early epidemic growth: A review. *Physics of life reviews* **18**, 66–97 (2016).

[8] Yang, W., Zhang, D., Peng, L., Zhuge, C. & Hong, L. Rational evaluation of various epidemic models based on the COVID-19 data of China. *arXiv preprint arXiv:2003.05666* (2020).

[9] Zhao, S., Musa, S. S., Fu, H., He, D. & Qin, J. Simple framework for real-time forecast in a data-limited situation: the zika virus ZIKV outbreaks in Brazil from 2015 to 2016 as an example. *Parasites & vectors* **12**, 344 (2019).

[10] Li, R. *et al.* Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (COVID-19). preprint, Infectious Diseases (except HIV/AIDS) (2020). URL `http://medrxiv.org/lookup/doi/10.1101/2020.02.14.20023127`.

[11] Tuite, A. R. & Fisman, D. N. Reporting, epidemic growth, and reproduction numbers for the 2019 novel coronavirus (2019-nCoV) epidemic. *Annals of Internal Medicine* (2020).

[12] Tang, B. *et al.* Estimation of the transmission risk of the 2019-nCoV and its implication for public health interventions. *Journal of Clinical Medicine* **9**, 462 (2020).

[13] Jefferson, T. *et al.* Physical interventions to interrupt or reduce the spread of respiratory viruses. *Cochrane database of systematic reviews* (2011).

[14] Kucharski, A. J. *et al.* Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *The Lancet Infectious Diseases* (2020).

[15] Ishola, D. A. & Phin, N. Could influenza transmission be reduced by restricting mass gatherings? Towards an evidence-based policy framework. *Journal of Epidemiology and Global Health* **1**, 33 (2011).

[16] Read, J. M., Bridgen, J. R., Cummings, D. A., Ho, A. & Jewell, C. P. Novel coronavirus 2019-nCoV: early estimation of epidemiological parameters and epidemic predictions. *medRxiv* (2020).

[17] Lin, Q. *et al.* A conceptual model for the outbreak of coronavirus disease 2019 (COVID-19) in Wuhan, China with individual reaction and governmental action. *International Journal of Infectious Diseases* (2020).

[18] Hatchett, R. J., Mecher, C. E. & Lipsitch, M. Public health interventions and epidemic intensity during the 1918 influenza pandemic. *Proceedings of the National Academy of Sciences* **104**, 7582–7587 (2007).

[19] The staggering rise in jobless claims this week. `https://www.nytimes.com/interactive/2020/03/19/upshot/coronavirus-jobless-claims-states.html`. Accessed: 2020-03-19.

[20] Global fear of flying spawns crisis for airlines. `https://www.wsj.com/articles/global-fear-of-flying-spawns-crisis-for-airlines-11583772004`. Accessed: 2020-03-20.

[21] Almond, D., Currie, J. & Duque, V. Childhood circumstances and adult outcomes: Act II. *Journal of Economic Literature* **56**, 1360–1446 (2018).

[22] Ma, J. Estimating epidemic exponential growth rate and basic reproduction number. *Infectious Disease Modelling* (2020).

[23] Angrist, J. D. & Pischke, J.-S. *Mostly harmless econometrics: An empiricist's companion* (Princeton university press, 2008).

[24] Greene, W. H. *Econometric analysis* (Pearson Education India, 2003).

[25] Romer, C. D. & Romer, D. H. The macroeconomic effects of tax changes: estimates based on a new measure of fiscal shocks. *American Economic Review* **100**, 763–801 (2010).

[26] Metcalf, G. E. & Stock, J. H. Measuring the macroeconomic impact of carbon taxes. In *American Economic Review: Papers and Proceedings* (2020).

[27] Cerra, V. & Saxena, S. C. Growth dynamics: the myth of economic recovery. *American Economic Review* **98**, 439–57 (2008).

[28] Burke, M., Hsiang, S. M. & Miguel, E. Global non-linear effect of temperature on economic production. *Nature* **527**, 235–239 (2015).

[29] Maier, B. F. & Brockmann, D. Effective containment explains sub-exponential growth in confirmed cases of recent COVID-19 outbreak in Mainland China. *arXiv preprint arXiv:2002.07572* (2020).

[30] Pitzer, V. E. *et al.* Predicting the impact of vaccination on the transmission dynamics of typhoid in south asia: a mathematical modeling study. *PLoS neglected tropical diseases* **8** (2014).

[31] Kandula, S. *et al.* Evaluation of mechanistic and statistical methods in forecasting influenza-like illness. *Journal of The Royal Society Interface* **15**, 20180174 (2018).

[32] Bootsma, M. C. J. & Ferguson, N. M. The effect of public health measures on the 1918 influenza pandemic in U.S. cities. *Proceedings of the National Academy of Sciences* **104**, 7588–7593 (2007).

[33] Charu, V. *et al.* Human mobility and the spatial transmission of influenza in the United States. *PLoS computational biology* **13**, e1005382 (2017).

[34] Wu, J. T., Leung, K. & Leung, G. M. Nowcasting and forecasting the potential domestic and international spread of the 2019-ncov outbreak originating in wuhan, china: a modelling study. *The Lancet* **395**, 689–697 (2020).

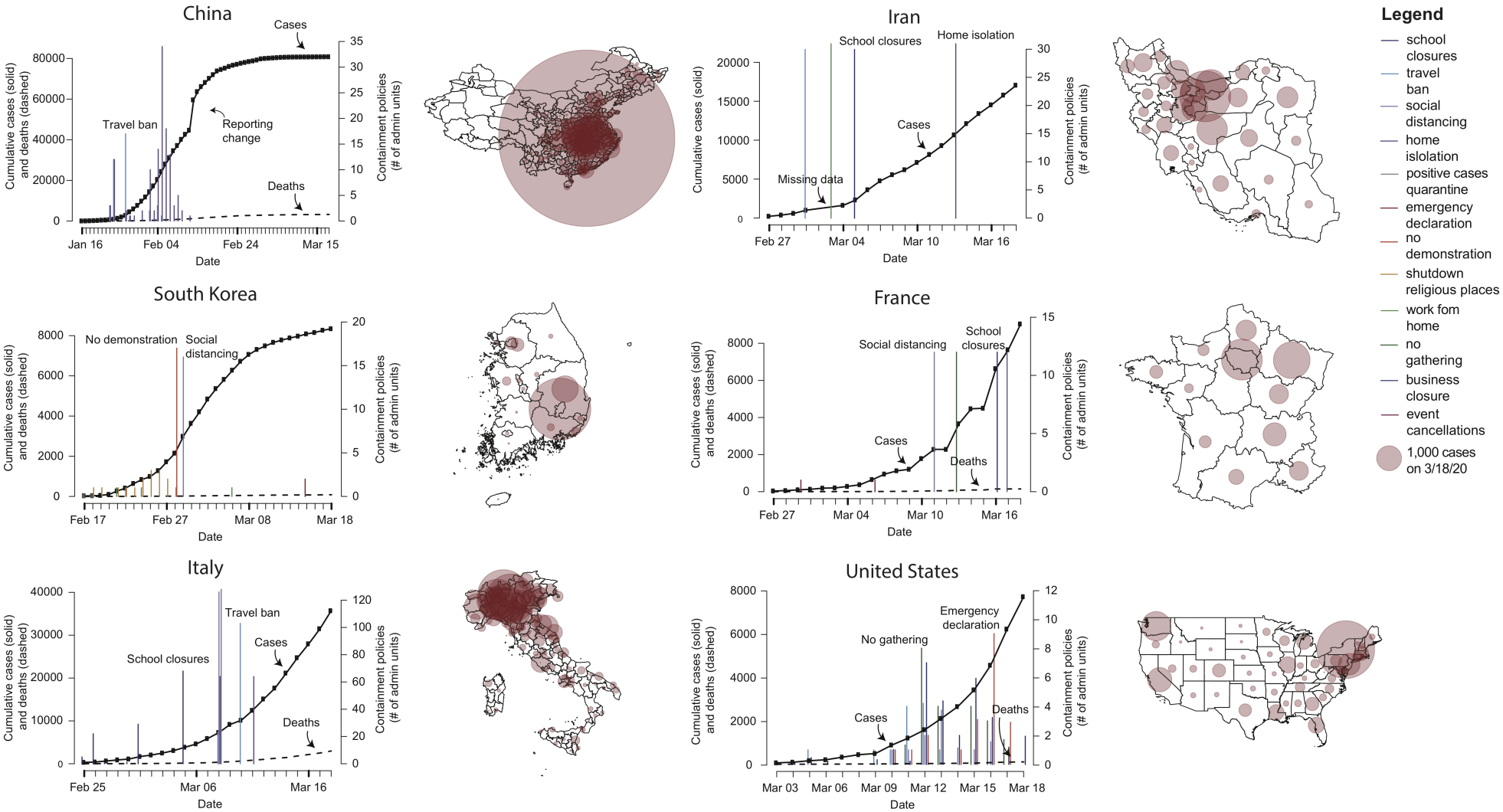


Figure 1: **Data on COVID-19 cases and large-scale anti-contagion policies in six countries.** The left-hand-side plots show cumulative confirmed cases of COVID-19 (solid black line w. squares) and deaths attributed to the disease (dashed black line) over time (left axes). Deployment of anti-contagion policies are indicated by vertical lines, with height corresponding to the number of administrative units that instituted the policy on a given day (right axes). For display purposes, not all policies are shown for each country; we highlight up to five policy types in each country, giving priority to those that were enacted in the most administrative units (see SI for all policies). The right-hand-side maps show the number of confirmed cases by administrative unit, as of March 18, 2020. The area of each circle scales with the number of cases. Country-specific longitudinal data are at the region level in France, the state level in the US, the province level in South Korea, Italy, and Iran, and the city level in China. For China, there are some discrepancies between the administrative units used in the epidemiological data and the official administrative units that the map is based on. We use the former in our analysis, but display the administrative units common to both data sets here.

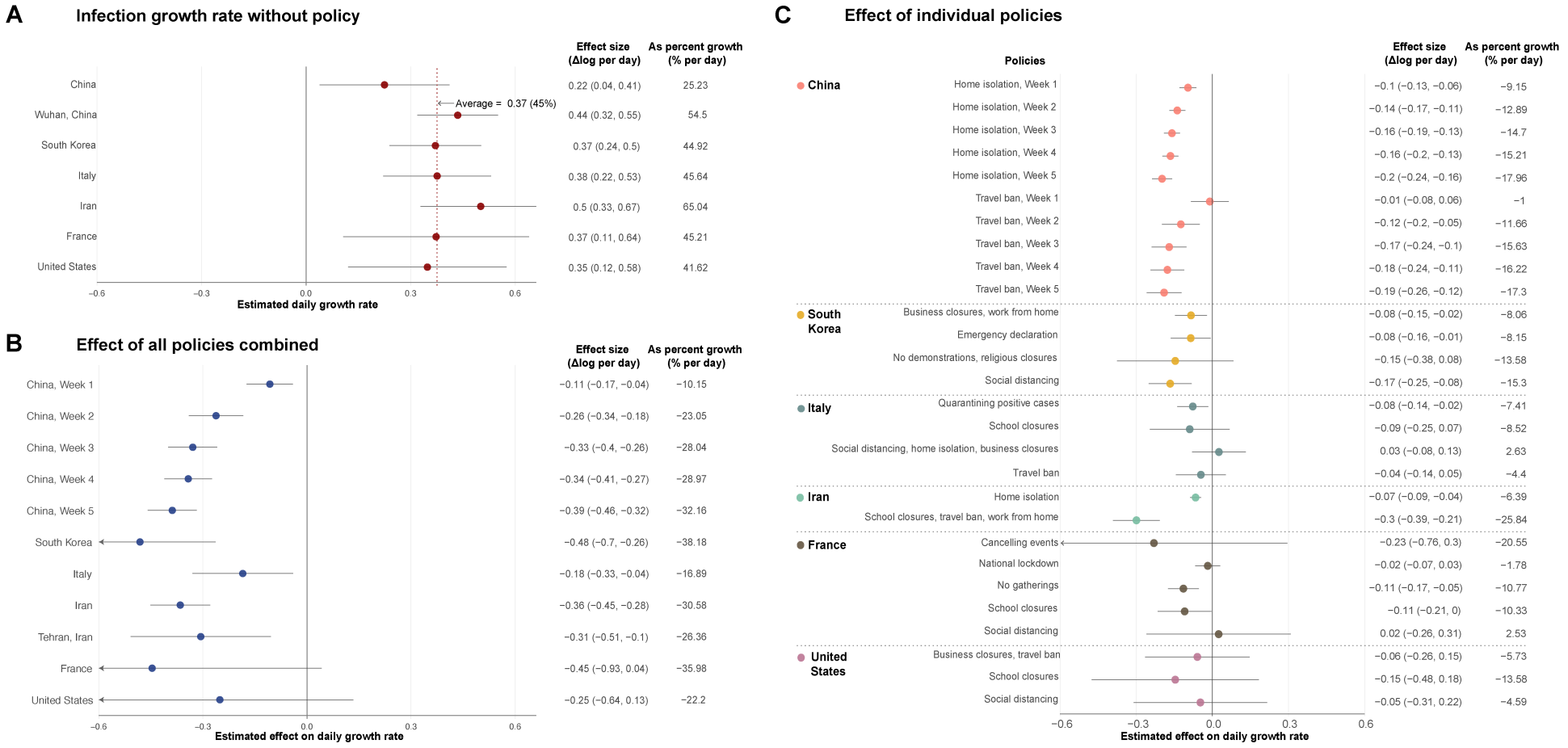


Figure 2: **Estimated "no policy" infection growth rates and the estimated impact of anti-contagion policies.** Markers are country-specific estimates, whiskers are 95% CI that account for spatial autocorrelation and other country-by-day events. Columns report effect sizes as a change in the continous-time growth rate ($\Delta \log(infections)$ per day) and the day-over-day percentage growth rate (% change per day). (**A**) Estimates of daily COVID-19 infection rate growth that would occur in the absence of any anti-contagion policy. The vertical red dashed line is the average estimated daily infection rate across the six countries. We display a separate estimate for Wuhan, China because it is the only city in China with substantial pre-policy infection data; but only the national China estimate is included in the average. (**B**) Estimated total impact of all anti-contagion policies in each country on infection growth rates. These estimates describe the combined impact of all policies deployed in a country, for a locality where policies are fully utilized and affect the entire population. Weekly impacts in China are the week-by-week impact for the same set of policies which remained deployed for multiple weeks. Results are marginally statistically significant for France ($p < 0.09$). (**C**) Estimated effects of individual policies (or collections of policies) on the daily growth rate of infections. Policies are sometimes grouped by similarity in goal (e.g. closing libraries and closing museums are grouped) or timing (e.g. policies that are deployed simultaneously in a given country) to reduce the number of estimated parameters. Effects are all estimated simultaneously within a country. For China, we simultaneously estimate separate effects for each week after the policy was implemented (e.g. "China, week 2" is the change in daily growth rates caused by policies implemented 8-14 days prior).

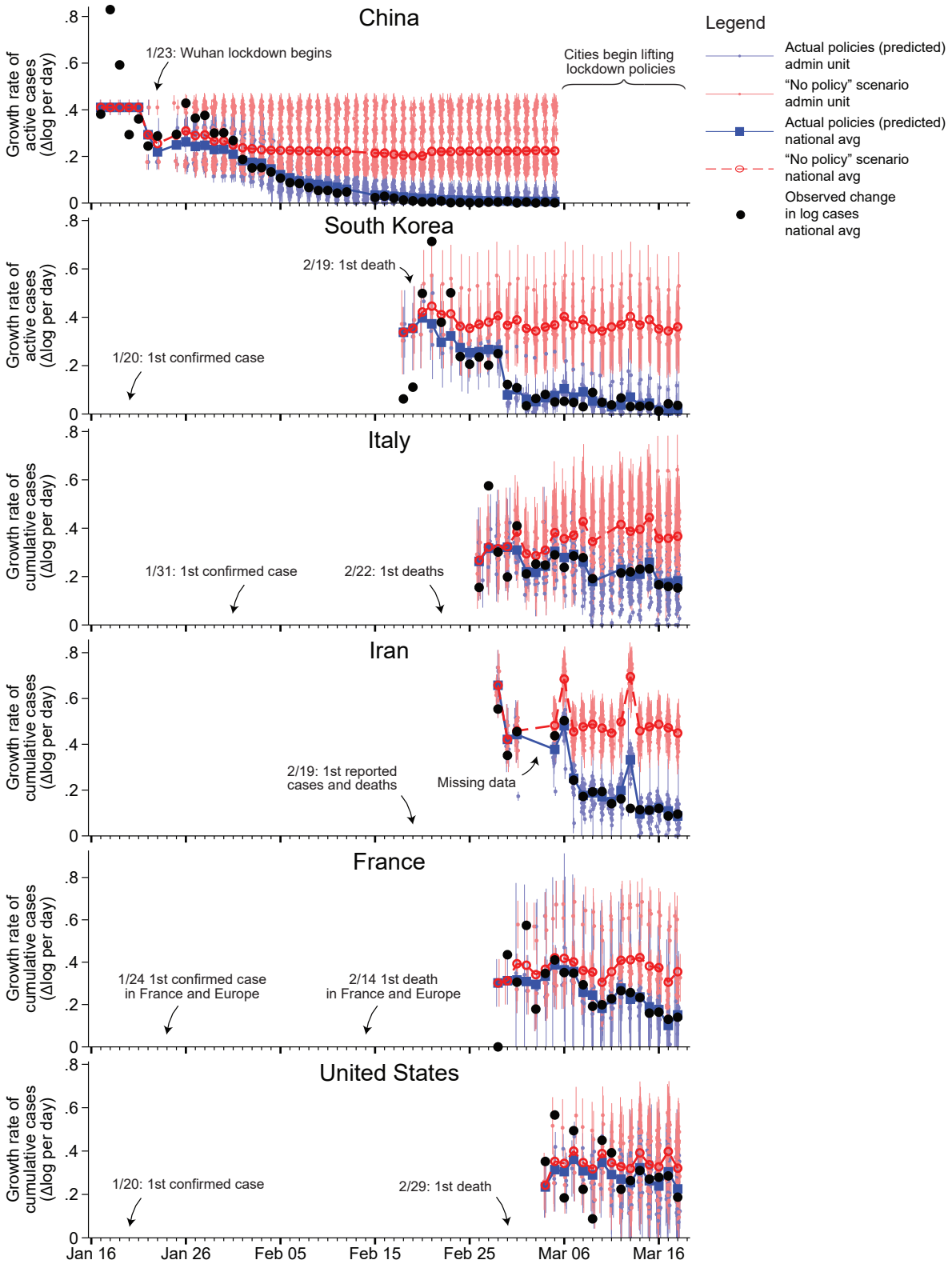


Figure 3: **Estimated infection growth rates based on actual anti-contagion policies and in a "no policy" counterfactual scenario.** Predicted daily growth rates of active (China and South Korea) or cumulative (all others) COVID-19 infections based on the observed timing of all policy deployments within each country (blue) and in a scenario where no policies were deployed (red). The difference between these two predictions is our estimated effect of actual anti-contagion policies on the growth rate of infections. Small markers are daily estimates for sub-national administrative units (vertical lines are 95% CI). Large markers are national average values for all sub-national units in our sample on that day. Black circles are observed changes in $\log(infections)$, averaged across the same administrative units. Predictions are only for observations in our sample, and we omit observations before sub-national units report ten cumulative cases. To focus our analysis on the impact of new policies, we omit data from China after March 5, 2020 because policies began to be rolled back during this period.

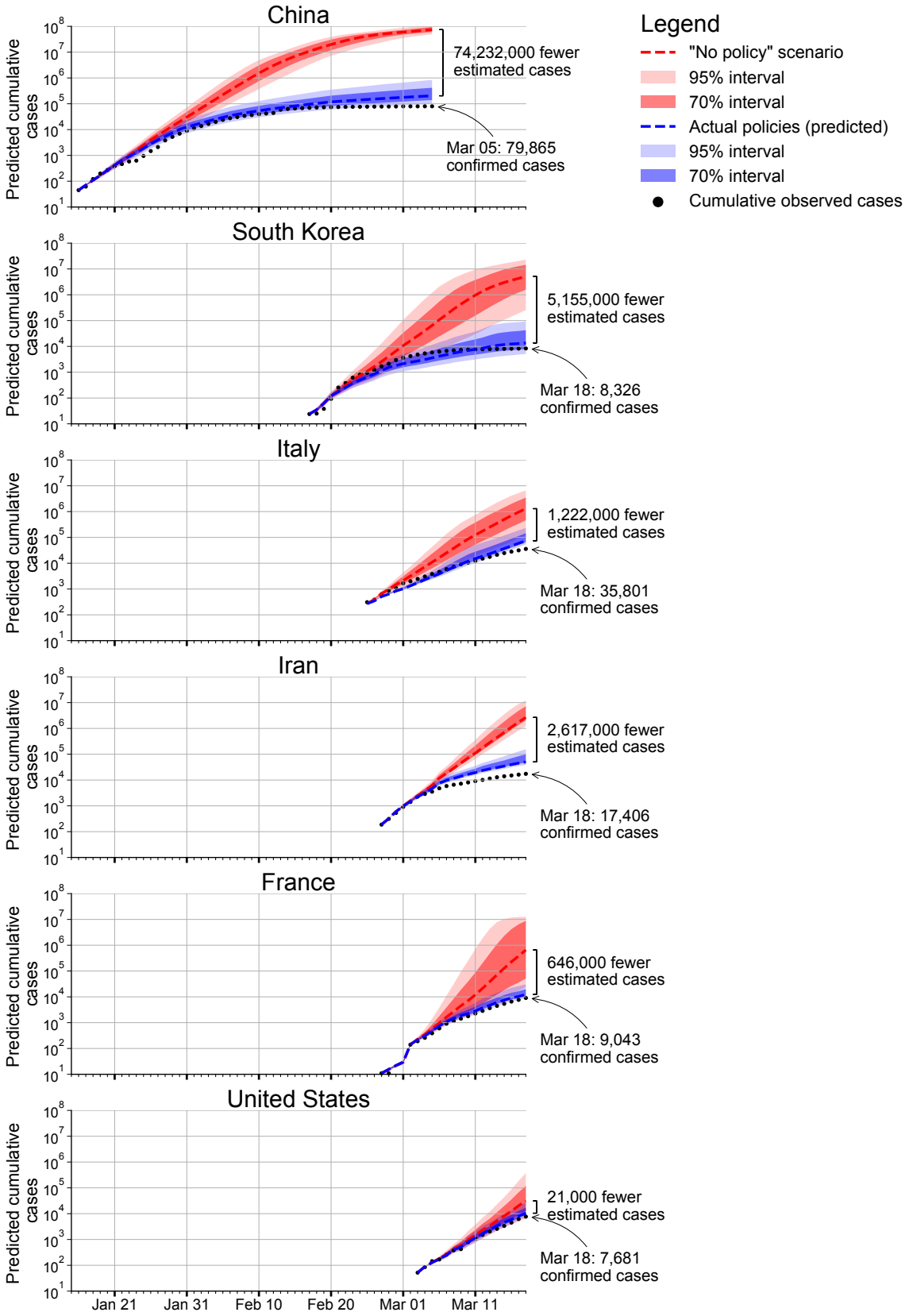


Figure 4: **Estimated cumulative COVID-19 infections with and without anti-contagion policies.** The predicted cumulative number of COVID-19 infections based on each country's actual policy deployments (blue) and in the "no policy" counterfactual scenario (red). Sub-national infection growth rates from Figure 3 are integrated adjusting for SIR system dynamics in each sub-national unit (see Methods). Shaded areas show uncertainty based on 1,000 simulations where estimated parameters are resampled from their joint distribution (dark = inner 70% of predictions; light = inner 95%). Black circles show the cumulative number of reported infections observed in the data. In both scenarios, the sample is restricted to units we analyze in Figures 2 and 3. Note that infections are not projected for administrative units that never report infections in the data, but which plausibly would have experienced infections in a "no policy" scenario. The jump in infections in France on March 2, 2020 occurs due to administrative units entering the sample.

Appendix for "The Effect of Large-Scale Anti-Contagion Policies on the Coronavirus (COVID-19) Pandemic"

Solomon Hsiang,*[1,2] Daniel Allen,[†1] Sébastien Annan-Phan,[†1,3] Kendon Bell,[†1,4] Ian Bolliger,[†1,5] Trinetta Chong,[†1] Hannah Druckenmiller,[†1,3] Andrew Hultgren,[†1,3] Luna Yue Huang,[†1,3] Emma Krasovich,[†1] Peiley Lau,[†1,3] Jaecheol Lee,[†1,3] Esther Rolf,[†1,6] Jeanette Tseng,[†1] and Tiffany Wu[†1]

[1]Global Policy Laboratory, Goldman School of Public Policy, UC Berkeley
[2]National Bureau of Economic Research & Centre for Economic Policy Research
[3]Agricultural & Resource Economics, UC Berkeley
[4]Manaaki Whenua, Landcare Research
[5]Energy & Resources Group, UC Berkeley
[6]Electrical Engineering & Computer Science Department, UC Berkeley

*Correspondence: shsiang@berkeley.edu. [†] Equal contribution (alphabetical).

The following sections provide additional information to support the article and to document our data collection and processing procedures.
All data and code are available at https://github.com/bolliger32/gpl-covid.
Additional resources and updates are available at www.globalpolicy.science/covid19.
For repository related questions, please contact bolliger@berkeley.edu.
For data related questions, please contact peiley@berkeley.edu.

This document last updated: March 21, 2020. Data last updated: March 19, 2020.

1. Appendix Tables and Figures

Table A1: Number of policies tabulated by administrative divisions of each country.

Country	Adm0	Adm1	Adm2	Adm3	Total
China	-	4 Province	126 City	-	130 All
France	8 Country	2 Région	49 Département	-	59 All
Iran	5 Country	16 Province	-	-	21 All
Italy	12 Country	24 Region	81 Province	77 Municipality	194 All
South Korea	20 Country	22 Province	-	-	42 All
United States	16 Country	191 State	38 County	245 City	490 All
Total	**61**	**259**	**294**	**322**	**936**

Table A1: Policy data have been collected at various levels of administrative divisions in each country. “Adm0” represents the country level, and higher “Adm*” numbers indicate smaller administrative subdivisions. Each policy is counted at the highest level of specificity of the regions where the policy is applied. For example, if a country has ten regions at the “Adm1” level, and a policy is applied across five of those regions, the policy is counted as five separate “Adm1” policies rather than a single “Adm0” policy.

Figure A1: Error distributions for estimated growth rates of COVID-19 cases by country.

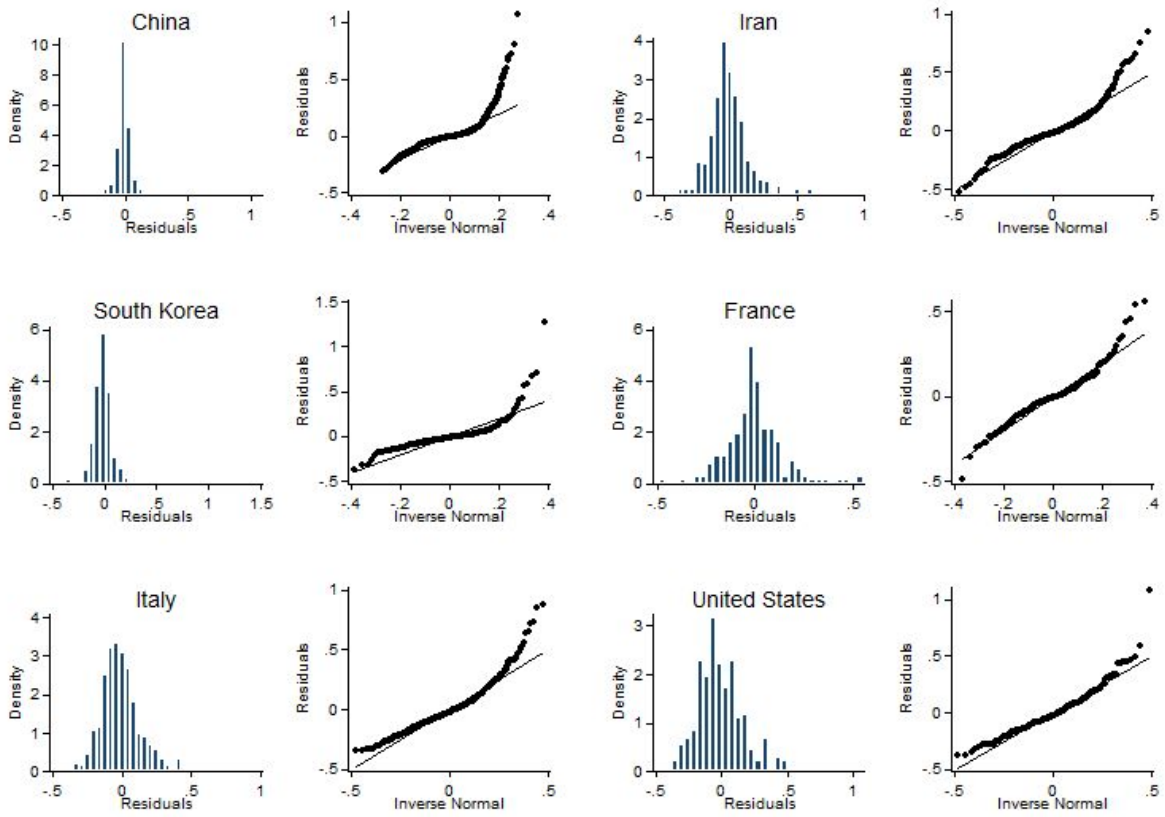
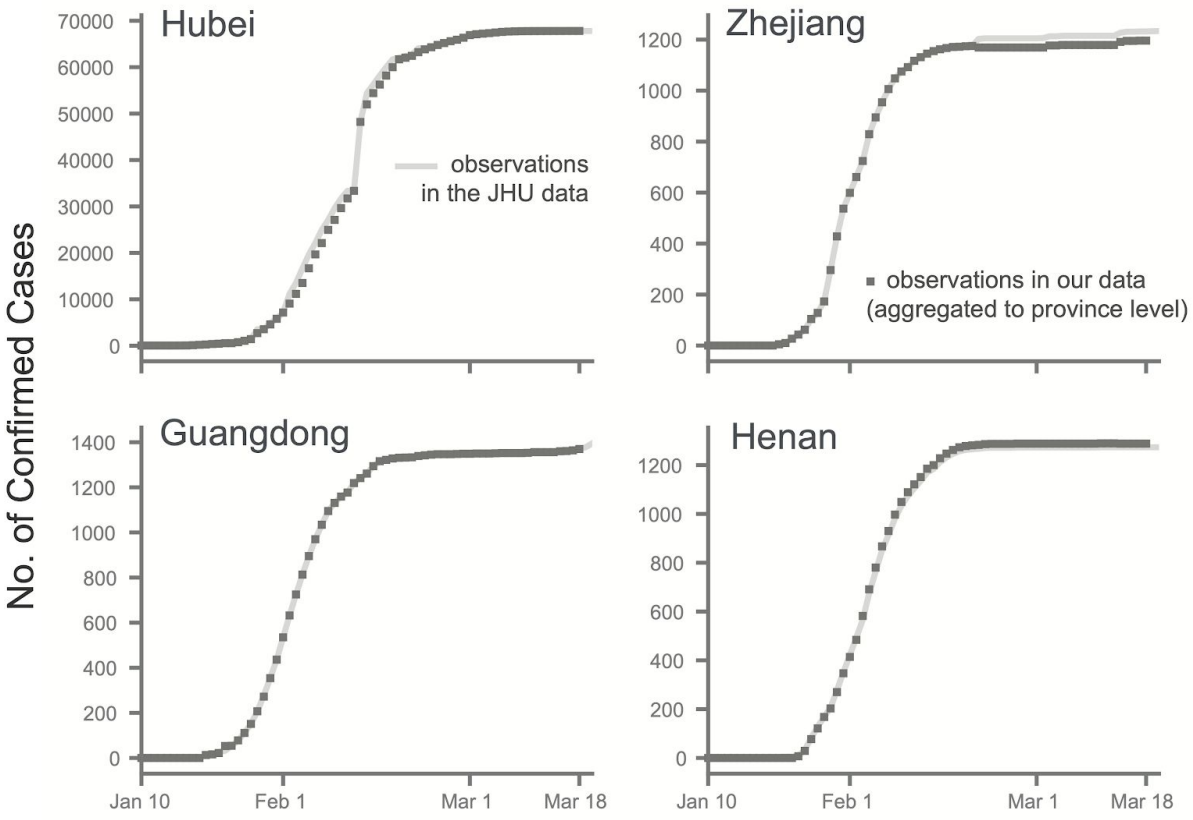


Figure A1: These plots show the error structure for each country-specific econometric model used to predict the daily growth of active or cumulative COVID-19 cases under the country's actual policy regime, as compared to the counterfactual world where no policies were enacted. See the full model under the Methods - Econometric analysis section as well as the results in Figure 3 of the main paper.

Figure A2: Validating our disaggregated epidemiological data against data from the Johns Hopkins Center for Systems Science and Engineering.

Panel A: The four provinces in China with the highest number of cumulative confirmed cases as of March 18, 2020.



Panel B: The entire country of South Korea.

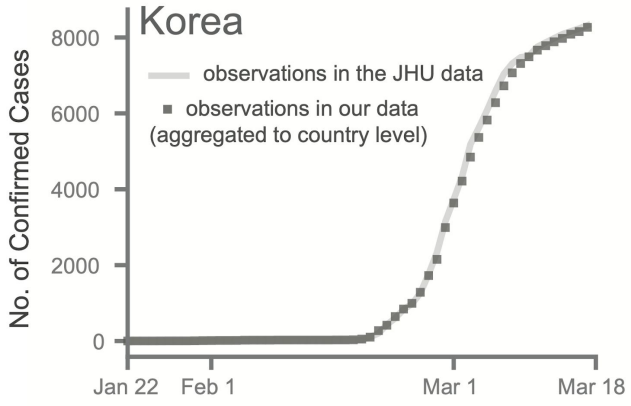


Figure A2: As an additional check, we compared the cumulative number of confirmed cases from a handful of regions in our collated epidemiological dataset to the same statistics from the 2019 Novel Coronavirus COVID-19 (2019-nCoV) Data Repository by the Johns Hopkins Center for Systems Science and Engineering (JHU CSSE).[1] We conducted this comparison for the two countries that we had the most data for and at two different administrative levels. In China, we aggregate city level data up to the province level, and in Korea we aggregate provincial level data up to the country level. The numbers tracked each other for the entire time series we have collected thus far.

2. Data Acquisition and Processing

This section describes the data acquisition and processing procedure for both epidemiological and policy data used in this paper. The sources for both types of data come from a variety of in-country data sources, which include government public health websites, regional newspaper articles, and Wikipedia crowd-sourced information. We have supplemented this data with international data compilations. A list of the epidemiological and policy data compiled for this analysis can be found here.

Epidemiological Data

The epidemiological datasets and sources used in this paper are described below. The main health variables of interest:

1. "*cum_confirmed_cases*": The total number of confirmed positive cases in the administrative area since the first confirmed case.
2. "*cum_deaths*": The total number of individuals that have died from COVID-19.
3. "*cum_recoveries*: The total number of individuals that have recovered from COVID-19.
4. "*cum_hospitalized*": The total number of hospitalized individuals.
5. "*cum_hospitalized_symptom*": The total number of symptomatic hospitalized individuals.
6. "*cum_intensive_care"* : The total number of individuals that have received intensive care.
7. "*cum_home_confinement*": The total number of individuals that have been self-quarantined in their homes as a result of a positive test.
8. "*active_cases*": The number of individuals who currently still test positive on the date of the observation.
9. "*active_cases_new*": The number of new cases since the previous date.
10. "*cum_tests*": The total number of tests (includes both positive and negative results) conducted in an administrative unit.

Additional metadata accompanying the health outcome variables:

1. "*date*": The date of observation.
2. "*adm0_name*": The ISO3 code to which this observation belongs.
3. "*adm1_name*": The name of the "Adm1" region to which this observation belongs.

[1] https://github.com/CSSEGISandData/COVID-19

4. "*adm2_name*": If the dataset contains observations at the "Adm2" level, then this is the name of the "Adm2" region to which this observation belongs.
5. "*adm[1,2]_id*": Any alphanumeric ID scheme to identify different administrative units (e.g. FIPS code).
6. "*lat*": The latitude of the centroid of the administrative unit.
7. "*lon*": The longitude of the centroid of the administrative unit.
8. "*policies_enacted*": The number of active policies that are in place for the administrative unit as of that date. This variable is not population weighted.
9. "*testing_regime*": A categorical variable used to identify when an administrative region (or country) changed their COVID-19 testing regime. This is zero-indexed, with the ordering only indicating chronological progression (there is no external meaning to Regime 2 vs. Regime 1 vs. Regime 0, and there is no consistency enforced for coding across countries). For example, if China changes their testing regime twice, all observations prior to the first regime change would be coded "*testing_regime*=0," all observations in between the two changes would be coded "*testing_regime*=1," and all observations after the second change would be coded "*testing_regime*=2."

Data Imputation:

In instances where health outcome observations are missing or suffer from data quality issues, we have imputed to fill in the missing values. Imputed health outcome variables are denoted by "*[health_outcome]_imputed.*" For the majority of our analyses we do not use imputed data; France is the exception where we impute two days of missing data. We do this to ensure we have variation in policy variables for use in the analysis.

We impute by:

1. Taking the natural log of the non-missing observations pertaining to that health outcome variable.
2. Linearly interpolating over the missing dates for that health outcome variable.
3. Exponentiating the interpolated values back into levels and rounding to the nearest integer.

China

We have collated a city level time series health outcome dataset in China for 339 cities from January 10, 2020 to present-day.

For data from January 24, 2020 onwards, we relied on the public dataset Ding Xiang Yuan[2] (DXY) that reports daily statistics across Chinese cities. Since DXY only publishes the most recent (cross-sectional) statistics (and not the historical data), we used the time series dataset scraped from DXY in an open source GitHub project[3]. The web scraper program checks for updates at least once a

[2] https://ncov.dxy.cn/ncovh5/view/pneumonia
[3] https://github.com/BlankerL/DXY-COVID-19-Data

day for the statistics published on DXY and records any changes in the number of cumulative confirmed cases, cumulative recoveries or cumulative deaths.

We assumed that no updates to the statistics meant there had been no new cases. We dropped a small number of cases that had been recorded but not assigned to a specific city (many of these cases are imported ones from other cities). We also dropped confirmed cases in prison populations (we assumed the spread of COVID-19 in prisons was not affected by the implementation of city-level lockdowns or travel ban policies).

For city level health outcomes prior to January 24, 2020, we manually collected official daily statistics from the central[4] and provincial (Hubei,[5] Guangdong,[6] and Zhejiang[7]) Chinese government websites. We did not collect city level health outcomes recorded prior to January 24, 2020 in provinces that had fewer than ten confirmed cases at that date. We made this decision since our analysis dropped observations with fewer than ten cumulative confirmed cases to prevent noisy data during the early transmission phase from disproportionately biasing the estimated results.

After merging the two datasets, we conducted a few quality checks:

(1) We checked that cumulative confirmed cases, cumulative recoveries, and cumulative deaths were increasing over time. In instances when cumulative outcomes decreased over time, we assumed that the recent numbers were more reliable, and treated the earlier number of cumulative cases as missing (this was often due to data entry errors or cases where patients that were reported to have been diagnosed with COVID-19, but were later found out to actually have tested negative). The magnitude of these errors was relatively small. We filled in any missing data with the imputation methodology described in the health data overview section.

(2) We validated our city level dataset by aggregating observations up to the provincial level and comparing the time trends from the aggregated dataset to that of the provincial dataset collated by Johns Hopkins University.[8] We confirmed that the two datasets matched very closely (see Figure A2 Panel A).

Testing Regime Changes:
As of the time of writing, the criteria for being diagnosed with COVID-19 had changed twice in China.[9] On February 13, 2020, China recategorized patients who exhibited symptoms, as determined through a chest scan, as part of the "confirmed" cases count even if they had not tested positive in the PCR test. This was due to concerns that the PCR test had relatively high false negative rates. On February

[4] http://www.nhc.gov.cn/xcs/yqtb/list_gzbd.shtml
[5] http://wjw.hubei.gov.cn/bmdt/ztzl/fkxxgzbdgrfyyq/xxfb/index_26.shtml
[6] http://wsjkw.gd.gov.cn/zwyw_yqxx/index_5.html
[7] http://www.zj.gov.cn/col/col1228996608/index.html
[8] https://github.com/CSSEGISandData/COVID-19
[9] https://www.cnbc.com/2020/02/26/confusion-breeds-distrust-china-keeps-changing-how-it-counts-coronavirus-cases.html

20, 2020, China reversed this decision. We included this information in the dataset because it could have potentially changed the levels and short-term growth rates of the number of confirmed cases.

France

We have collated a regional level time series health outcome dataset in France from February 15, 2020 to present-day.

We used the number of confirmed COVID-19 cases by *région* from France's government website.[10] The sources listed for this dataset were the French public health website,[11] the Ministry of Solidarity and Health,[12] French newspapers that reported government information,[13] and regional public health websites.[14] Given that this dataset was not published on a daily basis, we supplemented it by scraping the number of confirmed cases by *région* on the French public health website, which has been updated every day.[15]

Testing Regime Changes:
As of the time of writing, there have been no changes in France's testing regime.

South Korea

We have collated a provincial level time series health outcome dataset in South Korea from January 20, 2020 to March 14, 2020.

Most provinces in South Korea have been publishing data on their number of confirmed coronavirus cases. Daegu,[16] Gyeongsangbuk-do,[17] Jeollabuk-do,[18] Gyeongsangbuk-do,[19] and Sejong[20] provinces

---

[10] https://www.data.gouv.fr/en/datasets/fr-sars-cov-2/

[11] https://www.santepubliquefrance.fr/maladies-et-traumatismes/maladies-et-infections-respiratoires/infection-a-coronavirus/

[12] https://solidarites-sante.gouv.fr/soins-et-maladies/maladies/maladies-infectieuses/coronavirus/article/points-de-situation-coronavirus-covid-19

[13] https://france3-regions.francetvinfo.fr/

[14] https://www.ars.sante.fr/

[15] https://www.santepubliquefrance.fr/maladies-et-traumatismes/maladies-et-infections-respiratoires/infection-a-coronavirus/articles/infection-au-nouveau-coronavirus-sars-cov-2-covid-19-france-et-monde

[16] http://www.daegu.go.kr/dgcontent/index.do?menu_id=00936642&menu_link=/icms/bbs/selectBoardArticle.do&bbsId=BBS_02112

[17] https://www.gb.go.kr/Main/open_contents/section/wel/page.do?mnu_uid=5857&LARGE_CODE=360&MEDIUM_CODE=90&SMALL_CODE=10mnu_order=2

[18] http://www.jeonbuk.go.kr/board/list.jeonbuk?boardId=BBS_0000105&menuCd=DOM_000000105010004000&contentsSid=1189&cpath=

[19] https://www.gb.go.kr/

have been reporting the number of confirmed cases on a daily basis. For these provinces, we recorded this published health data.

Given that the province of Gangwon-do[21] does not report provincial level health data, we refer to the daily number of new cases reported by each of its counties (Chuncheon-si,[22] Wonju-si,[23] Gangneung-si,[24] Taebaek-si,[25] Sokcho-si,[26] and Samcheok-si[27]). As a result, we manually collected the number of new confirmed cases from each county's webpage and aggregated the numbers to the provincial level.

The remaining provinces (Seoul,[28] Gyeonggi-do,[29] Incheon,[30] Busan,[31] Ulsan,[32] Gwangju,[33] Chungcheongnam-do,[34] Chungcheongbuk-do,[35] Gyeongsangnam-do,[36] Jeju,[37] and Jeollanam-do[38]) did not explicitly publish the number of cumulative confirmed cases. However, they did publish patient-level data, including the date of when patients had tested positive. For these provinces, we constructed the measure of cumulative confirmed cases by counting the number of daily confirmed cases and adding it to the previous date's total.

Most provinces did not publish the number of deaths. Instead, we checked the daily policy briefings posted on the government homepages mentioned in the footnotes and manually collected mortality data. In instances when mortality data were not found in the briefings, we obtained the mortality data from other official sources, such as through social media sources (e.g. Facebook) and blogs maintained by local governments. Lastly we supplement these sources with mortality data reported in news articles.

Testing regime changes:
We collected information on testing regime changes from the homepage of the Korean Center for Disease Control and Prevention (KCDC). In the press release menu, the KCDC uploaded daily briefing announcements which contained information on testing criteria and changes to the testing regime.[39]

---

[20] https://www.sejong.go.kr/bbs/R3273/list.do?cmsNoStr=17465
[21] https://www.provin.gangwon.kr/covid-19.html
[22] https://www.chuncheon.go.kr/index.chuncheon?menuCd=DOM_000000599001000000
[23] https://www.wonju.go.kr/intro.jsp
[24] https://www.gn.go.kr/
[25] http://www.taebaek.go.kr/intro.jsp
[26] http://www.sokcho.go.kr/intro.html
[27] http://www.samcheok.go.kr/02179/02696.web
[28] http://www.seoul.go.kr/coronaV/coronaStatus.do
[29] https://www.gg.go.kr/contents/contents.do?ciIdx=1150&menuId=2909
[30] https://www.incheon.go.kr/health/HE020409
[31] http://www.busan.go.kr/corona19/index
[32] http://www.ulsan.go.kr/corona.jsp
[33] https://www.gwangju.go.kr/
[34] http://www.chungnam.go.kr/coronaStatus.do
[35] http://www1.chungbuk.go.kr/covid-19/index.do
[36] http://xn--19-q81ii1knc140d892b.kr/main/main.do
[37] https://www.jeju.go.kr/corona19.jsp
[38] https://www.jeonnam.go.kr/coronaMainPage.do
[39] https://www.cdc.go.kr/board/board.es?mid=a30402000000&bid=0030

Initially, the South Korean government only tested people who: 1) demonstrated respiratory symptoms within 14 days after visiting Wuhan South China Seafood Wholesale Market and 2) those who had pneumonia symptoms within 14 days after returning from Wuhan.[40]

As the outbreak spread, the KCDC broadened the criteria for testing. Starting January 28, 2020, the agency isolated 1) those who had fever or respiratory symptoms upon returning from Hubei province and 2) those who had symptoms of pneumonia upon returning from mainland China.[41,42] We coded this as the first change in the testing regime.

The second testing regime change occurred on February 4, 2020, when the KCDC announced that people who had had any "routine contacts" with confirmed cases were required to self quarantine for a 14-day period. The agency defines two categories of contacts: close contacts and routine contacts. The former is defined as a person who has been within two meters of, in the same room as, or exposed to any respiratory secretions of an infected individual. The latter refers to whether the individual conducted any activity in the same place and time as the infected person. Prior to this regime change, KCDC separated those two cases and applied different quarantine policies; starting February 4, 2020, any routine contacts were also required to be self-quarantined. [43]

Shortly thereafter, South Korea aggressively expanded the scope of their testing. Starting February 7, 2020, the KCDC broadened the definition of suspected cases to 1) anyone who developed a fever or respiratory symptoms within 14 days after returning from China, 2) anyone who developed a fever or respiratory symptoms within 14 days after being in close contact with a confirmed case, and 3) anyone suspected of contracting COVID-19 based on their travel history to affected countries and their clinical symptoms.[44] Moreover,the KCDC announced that the test would be free for all suspected cases and

---

[40] https://www.cdc.go.kr/board/board.es?mid=a20501000000&bid=0015&list_no=365654&act=view
[41] https://www.cdc.go.kr/board/board.es?mid=a20501000000&bid=0015&list_no=365874&act=view
[42] NB: The KCDC English website explains the testing regime change in a more condensed format: "Any citizens identified with a fever or respiratory symptoms and have visited Wuhan will be isolated and tested at a nationally designated isolation hospital, and any foreigners staying in Korea will be conducted in cooperation with police."
https://www.cdc.go.kr/board/board.es?mid=a30402000000&bid=0030&act=view&list_no=365888&tag=&nPage=1
[43]
http://www.mohw.go.kr/react/al/sal0301vw.jsp?PAR_MENU_ID=04&MENU_ID=0403&page=1&CONT_SEQ=352662
[44]
https://www.cdc.go.kr/board/board.es?mid=a30402000000&bid=0030&act=view&list_no=366114&tag=&nPage=1
NB: The date of this press release is February 8, 2020, but the definition of "suspected cases" was effective starting from February 7, 2020.

confirmed cases.[45] As a result of these efforts, KCDC announced that they would begin to test 3,000 people daily, a marked increase from only 200 people a day.[46]

The KCDC revised their guidelines on February 20, 2020 in order to test more people. Their press release stated: "Suspected cases with a medical professional's recommendation, regardless of travel history, will get tested. Additionally, those who are hospitalized with unknown pneumonia will also be tested. Lastly, anybody in contact with a diagnosed individual will need to self-isolate, and will only be released when they test negative on the thirteenth day of isolation."[47]

As the number of patients grew rapidly, the KCDC decided to focus on more vulnerable groups. In their February 29, 2020 press release, the agency stated: "The KCDC has asked local government and health facilities to focus on tests and treatment, especially targeting those aged 65+ and those with underlying conditions who need early detection and treatment." This change was coded as our last testing regime change in the dataset.[48]

Italy

We have collated a regional and provincial level time series health outcome dataset in Italy from February 24, 2020 to present-day.

This data came from the GitHub repository maintained by the Italian Department of Civil Protection (*Dipartimento della Protezione Civile*). Health outcomes included the number of confirmed cases, the number of deaths, the number of recoveries, and the number of active cases. These figures have been updating daily at 5 or 6 pm (Central European Time). The regional level dataset was pulled directly from "*dati-regioni/dpc-covid19-ita-regioni.csv,*" and the provincial level dataset was pulled from "*dati-province/dpc-covid19-ita-province.csv.*"

Testing regime changes:
The testing regime change in Italy occurred when the Director of Higher Health Council announced on February 26, 2020 that COVID-19 testing would only be performed on symptomatic patients, as the majority of the previous tests performed were negative.

---

[45] NB: The testing fee was already somewhat affordable; a person needed to pay 160,000 KRW (about $130 USD). A related article can be found here: https://www.edaily.co.kr/news/read?newsId=02604326625668224&mediaCodeNo=257

[46] http://www.mohw.go.kr/upload/viewer/skin/doc.html?fn=1581054767217_20200207145247.pdf&rs=/upload/viewer/result/202003/

[47] https://www.cdc.go.kr/board/board.es?mid=a30402000000&bid=0030&act=view&list_no=366247&tag=&nPage=4#

[48] https://www.cdc.go.kr/board/board.es?mid=a30402000000&bid=0030&act=view&list_no=366406&tag=&nPage=2

Iran

We have collated a provincial level time series health outcome dataset in Iran from February 19, 2020 to present-day.

The Iranian government had been announcing its new daily number of COVID-19 confirmed cases at the provincial level on the Ministry of Health's website. This data has been compiled daily in the table "New COVID-19 cases in Iran by province"[49] located in the "2020 coronavirus pandemic in Iran" article on Wikipedia.

We spot-checked the data in the Wikipedia table against the Iranian Ministry of Health announcements[50] using a combination of Google Translate and a comparison[51] of the numbers in the announcements (which were written in Persian script) to the Persian numbers.

Testing regime changes:
On March 6, 2020, the Ministry of Health announced[52] a national coronavirus plan, which included contacting families by phone to identify potential cases, along with the disinfecting of public places. The plan was to begin in the provinces of Qom, Gilan, and Isfahan, and then would be rolled out nationwide. On March 13, 2020, the government announced a military-enforced home isolation policy throughout the nation.[53] This announcement included nationwide disinfecting of public places. While a follow-up announcement of the March 6 high testing regime stating its complete rollout was not found, the March 13 announcement did reference the implementation of the public spaces component of the earlier plan across the country. We thus assumed that the high testing regime had also been fully rolled out on March 13, 2020.

United States

We have collated a state level time series health outcome dataset in the United States from January 22, 2020 to present-day.

The data comes from the Github repository associated with the Johns Hopkins University (JHU) interactive dashboard (Dong, Du & Gardner 2020, Lancet). As of the time of writing, the data are available here. The repository and dashboard are updated essentially in real-time; at least daily.

Testing regime changes:

---

[49] https://en.wikipedia.org/wiki/2020_coronavirus_pandemic_in_Iran
[50] Example of Ministry of Health data http://behdasht.gov.ir/index.jsp?siteid=1&fkeyid=&siteid=1&pageid=54782&newsview=200716
[51] Google Translate sometimes translates various Persian numbers as "1". Persian numbers compared here: https://www.languagesandnumbers.com/how-to-count-in-persian/en/fas/
[52] https://www.dailymail.co.uk/news/article-8082443/ANOTHER-senior-Iranian-official-dies-coronavirus.html
[53] https://www.theguardian.com/world/2020/mar/13/revolutionary-guards-enforce-coronavirus-controls-iran

To determine the testing regime, we used estimated daily counts of the cumulative number of tests conducted in every state, as aggregated by the largely crowdsourced effort named "The Covid Tracking Project" (covidtracking.com). We estimated the total number of tests as the sum of confirmed positive and negative cases. For some states and some days, there have been no negative case counts, in which case we utilize just the confirmed positive cases. We also ensured that the confirmed number of positive cases agreed with the counts in the JHU dataset.

We programmatically filtered for possible testing regime changes by filtering for any consecutive days during which the testing rate increased at least 250% from one day to the next, and where this jump was an increase of at least 150 total tests over one day. After visually inspecting the candidates, we removed detected testing regime changes for North Carolina and Connecticut, as these states did not demonstrate spikes in their testing rate, but rather a more gradual and steady rate in the increase of testing.

(NB: the last download from covidtracking.com was March 19, 19:30 PST. We have been updating the process and the removal of detected testing regime changes periodically, so this may change.)

Policy Data

The policy events, datasets, and sources used in this paper are described below. For each country, the relevant country-specific policies identified were then mapped to a harmonized policy categorization used across all countries.

The policy categories are coded as binary variables, where "*[policy_variable]*" = 0 before the policy is implemented in that area, and "*[policy_variable]*" = 1 on the date the policy is implemented (and for all subsequent dates until the policy is lifted). The main policy categories identified across the six different countries fall into four broad classes:

1. Restricting travel:
    a. "*travel_ban_local*" : A policy that restricts people from entering or exiting the administrative area (e.g county or province) treated by the policy.
    b. "*travel_ban_intl_in*": A policy that either bans foreigners from specific countries from entering the country, or requires travelers coming from abroad to self-isolate upon entering the country.
    c. "*travel_ban_intl_out*": A policy that suspends international travel to specific foreign countries that have high levels of COVID-19 outbreak.
    d. "*travel_ban_country_list*": A list of countries for which the national government has issued a travel ban or advisory. This information supplements the policy variable "*travel_ban_intl_out.*"
    e. "*transit_suspension*": A policy that suspends any non-essential land-, rail-, or water-based passenger or freight transit.

2. Distancing through cancellation of events and suspension of educational/commercial/religious activities:
    a. "*school_closure*": A policy that closes school and other educational services in that area.
    b. "*business_closure*": A policy that closes all offices, non-essential businesses, and non-essential commercial activities in that area. "Non-essential" services are defined by area.
    c. "*religious_closure"*: A policy that prohibits gatherings at a place of worship, specifically targeting locations that are epicenters of COVID-19 outbreak. See the section on Korean policy for more information on this policy variable.
    d. "*work_from_home*": A policy that requires people to work remotely. This policy may also include encouraging workers to take holiday/paid time off.
    e. "*event_cancel*": A policy that cancels a specific pre-scheduled large event (e.g. parade, sporting event, etc). This is different from prohibiting all events over a certain size.
    f. "*no_gathering*": A policy that prohibits any type of public or private gathering. (whether cultural, sporting, recreational, or religious). Depending on the country, the policy can prohibit a gathering above a certain size, in which case the number of people is specified by the "*No_gathering_size*" variable.
    g. "*no_demonstration*": A policy that prohibits protest-specific gatherings. See the section on Korean policy for more information on this policy variable.
    h. "*social_distance*": A policy that encourages people to maintain a safety distance (often between one to two meters) from others. This policy differs by country, but includes other policies that close cultural institutions (e.g. museums or libraries), or encourage establishments to reduce density, such as limiting restaurant hours.
3. Quarantine and lockdown:
    a. "*pos_cases_quarantine*": A policy that mandates that people who have tested positive for COVID-19, or subject to quarantine measures, have to confine themselves at home. The policy can also include encouraging people who have fevers or respiratory symptoms to stay at home, regardless of whether they tested positive or not.
    b. "*home_isolation*": A policy that prohibits people from leaving their home regardless of their testing status. For some countries, the policy can also include the case when people have to stay at home, but are allowed to leave for work- or health-related purposes. For the latter case, when the policy is moderate, this is coded as '*home_isolation* = 0.5.'
4. Additional policies
    a. "*emergency_declaration*": A decision made at the city/municipality, county, state/provincial, or federal level to declare a state of emergency. This allows the affected area to marshal emergency funds and resources as well as activate emergency legislation.
    b. "*paid_sick_leave*": A policy where employees receive pay while they are not working due to the illness.

Optional policies:

In the cases when the aforementioned policies are optional, we denote this as "*[policy_variable]_opt.*"

Population weighting of policy variables:

In the cases when only a portion of the administrative unit (e.g. half of the counties within the state) are affected by the implementation of the policy, we weight the policy variable by the percentage of population within the administrative unit that is treated by the policy. This is denoted as "*[policy_variable]_popwt,*" and the value that this variable can take on is a continuous number between 0 and 1. Sources for the population data are detailed in a later section.

China

We obtain data on China's policy response to the COVID-19 pandemic by culling data on the start dates of travel bans and lockdowns at the city-level from the "2020 Hubei lockdowns" Wikipedia page, [54] the Wuhan Coronavirus Timeline project on Github,[55] and various news reports.

To combat the spread of COVID-19, the Chinese government imposed travel restrictions and quarantine measures, starting with the lockdown of the city of Wuhan, the origin of the pandemic, on January 23, 2020. Immediately following the Wuhan lockdown, neighboring cities followed suit, banning travel into and out of their borders, shutting down businesses, and placing residents under household quarantine. The same policy measures were implemented in cities across China for the next three weeks.

Some lockdowns occurred during the national Chinese New Year holiday (January 24 - 30, 2020) when schools and most workers were on break. On January 27, 2020, China extended the official holiday to February 2, 2020, while many additional provinces delayed resuming work and opening schools for even longer.[56] The Chinese New Year holiday is analogous to containment policies such as school closures and restrictions on non-essential work. We do not specifically estimate the effect of this holiday extension, as most cities were in lockdown during the extended holiday, and a lockdown is a more restrictive containment measure. A lockdown requires all residents to stay home, except for medical reasons or essential work, and only allows one person from each household to go outside once every one to five days (exact policy varied by city).

France

We obtain data on France's policy response to the COVID-19 pandemic from the French government website, press releases from each regional public health site, and Wikipedia.

---

[54] https://en.wikipedia.org/wiki/2020_Hubei_lockdowns
[55] https://github.com/Pratitya/wuhan2020-timeline
[56] https://www.china-briefing.com/news/china-extends-lunar-new-year-holiday-february-2-shanghai-february-9-contain-coronavirus-outbreak/

The French government website contains a timeline of all national policy measures.[57] Each regional public health agency (*l'Agence Régionale de Santé*) in France posts press releases with information on the policies the région or départements within the région will implement to mitigate the spread and impact of the COVID-19 outbreak.[58] The Wikipedia page on the 2020 coronavirus pandemic in France has collated information on the major policy measures taken in response to the COVID-19 pandemic.[59]

Starting February 29, 2020, France banned mass gatherings of more than 5,000 people nationwide, while some major sporting events were cancelled and a handful of schools closed to mitigate the spread of the virus. As more COVID-19 cases were confirmed during the following week, additional sporting events were canceled, more schools decided to close, and certain cities and *départements* limited mass gatherings to no more than 50 people, excluding shops, business, restaurants, bars, weddings, and funerals. Some régions closed early childhood establishments (e.g. nurseries, daycare centers) and prohibited visitors to elderly care facilities. On March 8, 2020, France banned mass gatherings of more than 1,000 people nationwide. Other schools, cities, and départements followed suit with additional school closures and limiting mass gatherings. On March 11, 2020, France prohibited all visits to elder care establishments. Starting March 16, 2020, France closed all schools nationwide.

We have coded various policies that cancel events and large gatherings as such: any cancellations of professional sporting and other specific pre-scheduled events as the policy variable "*event_cancel*." The "*no_gathering*" policy variable represents policy measures that banned all events or mass gatherings of a certain size, e.g. no gatherings of over 1,000 people. The "*social_distance*" policy variable includes measures preventing visits to elder care establishments, closures of public pools and tourist attractions, and teleworking plans for workers.

South Korea

We obtained data on South Korea's policy response to the COVID-19 pandemic from various news sources, as well as press releases from the Korean Centers for Disease Control and Prevention (KCDC), the Ministry of Foreign Affairs, and local governments' websites. The policy variables coded in the dataset are: "*business_closure,*" "*business_closure_opt,*" "*emergency_declaration,*" "*no_demonstration,*" "*religious_closure*," "*school_closure,*" "*social_distance_opt,*" "*travel_ban_intl_in_opt,*" "*travel_ban_intl_out_opt,*" and "*work_from_home_opt.*"

The KCDC announced on February 28, 2020 that health-related public facilities were recommended to be closed;[60] hence, the policy variable "*business_closure*" was coded as one from the announcement date. Even though it was technically a recommendation, we did not code this policy as an optional

[57] https://www.gouvernement.fr/info-coronavirus
[58] https://www.ars.sante.fr/
[59] https://fr.wikipedia.org/wiki/Pand%C3%A9mie_de_maladie_%C3%A0_coronavirus_de_2020_en_France
[60] http://www.mohw.go.kr/react/al/sal0301vw.jsp?PAR_MENU_ID=04&MENU_ID=0403&page=8&CONT_SEQ=353184

one because a majority of facility types listed in the press release (senior centers, job centers, children's centers, etc.) are under public administration, so these facilities likely would have followed recommendations. Indeed, some news articles have reported that all children's centers in Busan are closed[61] and over 3,600 facilities in Seoul.[62]

We have another business variable, "*business_closure_opt*", which applies to two provinces: Seoul and Gyeonggi-do. On March 11, 2020, the mayor of Seoul advised that popular commercial establishments such as karaoke places, clubs, and cyber cafes be closed.[63] Seven days later, the governor of Gyeonggi-do issued an executive order limiting the usage of commonly frequented commercial establishments and requiring a higher standard of cleanliness.[64] We coded this as an optional business closure given that the policy discourages usage of these facilities but did not explicitly order them to shut down.

Daegu and Gyeongsangbuk-do have been two of the regions hardest hit by COVID-19. The government of South Korea declared an emergency for those two areas on March 15, 2020.[65] We incorporated this information into the variable "*emergency_declaration*."

The variable "*no_demonstration*" reflects the efforts of some regions limiting any protests calling for slowing the spread of the outbreak. On February 24, 2020, Incheon stopped a protest in front of the Incheon Metropolitan City Hall.[66] Two days later, Seoul prohibited protests in downtown areas where massive demonstrations used to take place.[67]

Many province level COVID-19 policies have targeted religious gatherings at Shincheonji Church of Jesus, since its religious gatherings have been linked to the explosion in the number of cumulative confirmed cases. Provincial governments tried to shut down Shincheonji-related places of worship, and the related policy implementation is encoded in the variable "*religious_closure*." The regions which utilized this policy option are: Daegu,[68] Gyeongsangbuk-do,[69] Seoul,[70] Jeju,[71] Gyeonggi-do,[72]

---

[61] http://www.kookje.co.kr/news2011/asp/newsbody.asp?code=0300&key=20200313.33001005312
[62] http://www.hani.co.kr/arti/area/capital/929213.html
[63] http://news.seoul.go.kr/welfare/archives/512657
[64] https://gnews.gg.go.kr/briefing/brief_gongbo_view.do?BS_CODE=s017&number=43714&period_1=&period_2=&search=0&keyword=&subject_Code=BO01&page=1
[65] http://ncov.mohw.go.kr/tcmBoardView.do?brdId=3&brdGubun=31&dataGubun=&ncvContSeq=1241&contSeq=1241&board_id=311&gubun=BDC
[66] http://press.incheon.go.kr/citynet/jsp/sap/SAPNewsBizProcess.do?command=searchDetailSvp&sido=&matOfYmd=20200224&matSno=10&flag=&viFlag=in
[67] http://mediahub.seoul.go.kr/archives/1270860
[68] http://www.ctimes.co.kr/news/articleView.html?idxno=6843
[69] https://www.msn.com/ko-kr/news/national/%EA%B2%BD%EB%B6%81-%EC%8B%A0%EC%B2%9C%EC%A7%80-1612%EB%AA%85-%EC%A4%91-221%EB%AA%85-%ED%99%95%EC%A7%84%C2%B7%C2%B7%C2%B731%EB%B2%88%EC%9D%B4-156%EB%AA%85-%EC%98%AE%EA%B2%BC%EB%8B%A4/ar-BB10C1am

Jeollanam-do,[73] Gyeongsangnam-do,[74] Incheon,[75] Ulsan,[76] Busan,[77] Jeollabuk-do,[78] Chungcheongbuk-do,[79] Gwangju,[80] Chungcheongnam-do,[81] and Daejeon.[82]

The policy variable "*school_closure*" has been turned on for the entirety of the Korean time series dataset. This is because all schools were already on vacation during the beginning of the outbreak, and the government then postponed their start dates. At the time of writing, the Ministry of Education announced that schools would be kept closed until April 3, 2020.[83] Therefore, this policy variable is always equal to 1 in the dataset.

"*social_distance_opt*" has been turned on from February 29, 2020, when KCDC recommended social distancing as one of the main tools to deal with the outbreak. In their press release, they recommended that "people maintain personal hygiene and practice 'social distancing' until the beginning of March, an important point of this outbreak."[84] In the case of Daegu, the hardest-hit region in the country, we coded the variable as 1 starting from February 22, 2020, based on the statement, "It is recommended for residents in Daegu to minimize gathering events and outdoor activities."[85]

The first travel restriction for incoming travelers ("*travel_ban_intl_in_opt*") was implemented on January 28, 2020. It is worth noting that it was not a total prohibition of incoming visitors; rather, it means inbound travellers were subject to COVID-19 specific emergency measures. KCDC mentioned that starting on January 28, 2020 "any travellers depart[ing] from China [would] be a subject to

---

[70] http://www.c-herald.co.kr/news/articleView.html?idxno=2156
[71] http://www.jemin.com/news/articleView.html?idxno=646993
[72] http://www.kookje.co.kr/news2011/asp/newsbody.asp?code=0300&key=20200224.99099008869
[73] http://www.kwangju.co.kr/article.php?aid=1582729200690279004
[74] http://woman.chosun.com/mobile/news/view.asp?cate=C01&mcate=M1004&nNewsNumb=20200264476#_enliple
[75] https://www.gov.kr/portal/ntnadmNews/2102077
[76] https://www.yna.co.kr/view/RPR20200225010600353
[77] https://www.ajunews.com/view/20200311082759689
[78] http://www.segye.com/newsView/20200309517306
[79] http://www.cbnews.kr/news/articleView.html?idxno=113191
[80] http://www.bosa.co.kr/news/articleView.html?idxno=2122251
[81] http://www.dtnews24.com/news/articleView.html?idxno=572551
[82] https://blog.naver.com/PostView.nhn?blogId=storydaejeon&logNo=221834017110&redirect=Dlog&widgetTypeCall=true&directAccess=false
[83] https://www.moe.go.kr/boardCnts/view.do?boardID=294&boardSeq=80044&lev=0&searchType=null&statusYN=W&page=1&s=moe&m=020402&opType=N
[84] https://www.cdc.go.kr/board/board.es?mid=a30402000000&bid=0030&act=view&list_no=366406&tag=&nPage=2
[85] https://www.cdc.go.kr/board/board.es?mid=a30402000000&bid=0030&act=view&list_no=366299&tag=&nPage=3

strengthened screening and quarantine measures."[86] On February 12, 2020, KCDC broadened the list of countries subject to the stricter measures to include Hong Kong and Macau.[87] Subsequently, KCDC added Italy and Iran (on March 11, 2020)[88]; France, Germany, Spain, UK, and Netherlands (on March 15, 2020)[89] ; and any remaining European countries (March 15, 2020)[90] to their country list.

This restriction was not limited to inbound travellers. The government also issued advisories on countries where the number of infections had increased, which has been encoded as the variable "*travel_ban_intl_out_opt*." The first outbound travel alert due to COVID-19 was announced on January 28, 2020: The Ministry of Foreign Affairs (MOFA) issued a Level 2 (Yellow) alert for any travel to mainland China, Hong Kong, and Macau.[91] Later, MOFA added Italy on February 28, 2020,[92] Japan on March 9, 2020,[93] and all European countries on March 16, 2020.[94] It should be noted that the Level 2 alert does not enable the government to prohibit travel to these destinations, which is why the policy was coded as "optional."

There are four types of travel advisories distributed by the South Korean government: Level 1, Navy; Level 2, Yellow; Level 3, Red; and Level 4, Black.[95] Travel under the Level 4 alert is prohibited, and the government utilizes legal instruments to enforce the restriction. If people leave the country under the black alert, they will be subject to fines up to ten million KRW, or imprisonment up to a year. However, there is no enforcement instrument for the advisories up to Level 3. In that sense, we stated above that the banning policy does not mean prohibiting travel. Nevertheless, we coded the yellow alert as

---

86 https://www.cdc.go.kr/board/board.es?mid=a30402000000&bid=0030&act=view&list_no=365875&tag=&nPage=3

87 https://www.cdc.go.kr/board/board.es?mid=a30402000000&bid=0030&act=view&list_no=366154&tag=&nPage=1

88 https://www.cdc.go.kr/board/board.es?mid=a30402000000&bid=0030&act=view&list_no=366523&tag=&nPage=1

89 https://www.cdc.go.kr/board/board.es?mid=a30402000000&bid=0030&act=view&list_no=366537&tag=&nPage=1

90 https://www.cdc.go.kr/board/board.es?mid=a30402000000&bid=0030&act=view&list_no=366568&tag=&nPage=1

91 http://www.0404.go.kr/dev/newest_view.mofa?id=ATC0000000007598&pagenum=1&mst_id=MST0000000000040&ctnm=&div_cd=&st=title&stext=

92 http://www.0404.go.kr/dev/newest_view.mofa?id=ATC0000000007690&pagenum=1&mst_id=MST0000000000040&ctnm=&div_cd=&st=title&stext=

93 http://www.0404.go.kr/dev/newest_view.mofa?id=ATC0000000007709&pagenum=1&mst_id=MST0000000000040&ctnm=&div_cd=&st=title&stext=

94 http://www.0404.go.kr/dev/newest_view.mofa?id=ATC0000000007723&pagenum=1&mst_id=MST0000000000040&ctnm=&div_cd=&st=title&stext=

95 http://www.0404.go.kr/walking/walking_intro.jsp

the first travel ban in our dataset, since Level 2 alerts are issued relatively rarely, such as during a significant demonstration[96] or military coup.[97] As a result, we coded the Level 2 alert due to COVID-19 into the dataset for the policy analysis.

The policy variable "*work_from_home_optional*" indicates when KCDC began recommending that people work from home. On March 15, 2020, the KCDC press release stated: "Since contact with confirmed cases in an enclosed space increases the possibility of transmission, it is recommended to work at home or adjust desk locations so as to keep a certain distance among people in the office. More detailed guidelines for local governments and high-risk working environments will be distributed soon."[98]

Italy

We have obtained data on Italy's policy responses to the COVID-19 pandemic primarily from the English version of the COVID-19 dossier "Chronology of main steps and legal acts taken by the Italian Government for the containment of the COVID-19 epidemiological emergency"[99] written by the Department of Civil Protection (*Dipartimento della Protezione Civile*), most recently updated on March 12, 2020. This dossier details the majority of the municipal, regional, provincial, and national policies rolled out between the start of the pandemic to present-day. We have supplemented these policy events with news articles that detail which administrative areas were specifically impacted by the additional policies.

The first major policy rollout was on February 23, 2020, when 11 municipalities across two provinces in Northern Italy were placed on lockdown. These policies included closing schools, cancelling public and private events and gatherings, closing museums and other cultural institutions, closing non-essential commercial activities, and prohibiting the movement of people into or out of the municipalities.

The second major policy rollout was on March 1, 2020, when two provinces and three regions in Northern Italy were placed on partial lockdown. These policies also included closing schools, cancelling public and private events and gatherings, closing museums, closing non-essential commercial activities, as well as limiting the number of people at places of worship, restricting operating hours of bars and restaurants, and encouraging people to work remotely.

---

[96] http://www.0404.go.kr/dev/notice_view.mofa?id=ATC0000000007416

[97] http://www.0404.go.kr/dev/notice_view.mofa?id=8679&pagenum=1&st=title&stext=%EC%97%AC%ED%96%89%EA%B2%BD%EB%B3%B4

[98] https://www.cdc.go.kr/board/board.es?mid=a30402000000&bid=0030&act=view&list_no=366523&tag=&nPage=1

[99] http://www.protezionecivile.it/documents/20182/1227694/Summary+of+measures+taken+against+the+spread+of+C-19/c16459ad-4e52-4e90-90f3-c6a2b30c17eb

The third major policy roll-out was on March 5, 2020, when all schools across the country were closed.

The fourth major policy roll-out was on March 8, 2020 when the region of Lombardy and 13 provinces in Northern Italy were placed on lockdown. These policies included the cancellation of public and private events and gatherings, closing of museums, encouraging people to work remotely, limiting the number of people at places of worship, restricting opening hours of bars and restaurants, mandating quarantine of people who tested positive for COVID-19, prohibiting the movement of people into or out of the affected area, and restricting movement within the affected area to only work- or health-related purposes. Commercial activities were still allowed, as long as they maintained a safety distance of one meter apart per person within the establishment. All civil and religious ceremonies, including weddings and funeral ceremonies, were suspended. During this same policy roll-out, the rest of the country faced less stringent policies: cancelling of public and private events, closing of museums, and requiring restaurants and commercial establishments to maintain a safety distance of one meter apart per person within the establishment.

The fifth major policy roll-out was announced on March 9, 2020, and went into effect on March 10, 2020, when lockdown policies applied to Northern Italy were rolled out to the entire country. Lastly, on March 11, 2020, the lockdown was changed to also cover the closing of any non-essential businesses and further restricted people from leaving their home.

Iran

For Iran's policy response to the COVID-19 pandemic, we relied on news media reporting as the primary source of policy information (mostly due to translation restrictions). We also relied on two timelines of pandemic events in Iran to help guide the policy search.[100] [101]

The first major outbreak in Iran was connected to a major Shia pilgrimage in the city of Qom that brought Shiite pilgrims from Iran and throughout the Middle East, where they came to kiss the Fatima Masumeh shrine. It is possible that the disease was brought to Qom by a merchant traveling from Wuhan, China.[102] In addition, it is believed that the Iranian government knew of the COVID-19 outbreak prior to its February 21, 2020 parliamentary elections, but downplayed the risks associated with the disease as not to suppress voter turnout (given concerns that a low turnout would reflect poorly on its legitimacy).[103] The disease, initially centered in Qom and neighboring Tehran, spread rapidly throughout the country.

---

[100] https://www.thinkglobalhealth.org/article/updated-timeline-coronavirus
[101] https://en.wikipedia.org/wiki/2020_coronavirus_pandemic_in_Iran
[102] https://www.newyorker.com/news/our-columnists/how-iran-became-a-new-epicenter-of-the-coronavirus-outbreak
[103] https://www.newyorker.com/news/our-columnists/how-iran-became-a-new-epicenter-of-the-coronavirus-outbreak

As the number of cases grew, the Iranian government started to increase the stringency of its response. The first case was reported on February 19, 2020 (two individuals who both were reported to have died that day). The next day, school closures were announced in the province of Qom and travel in the region was discouraged. By February 22, 2020 the government closed schools in 14 provinces and closed down major gathering sites such as football matches and theaters. By March 5, 2020 schools were closed nationwide and government employees were required to work from home. Home isolation was implemented by the military on March 13, 2020, which the media described as "the near-curfew follows growing exasperation among MPs that calls for Iranian citizens to stay at home had been widely ignored, as people continued to travel before the Nowruz New Year holidays." [104]

United States

For the United States' policy response to the COVID-19 pandemic, we relied on a number of sources, including the U.S. Center for Disease Control (CDC), individual state health departments, as well as various press releases from county and city-level government or media outlets. The CDC has posted and continually updated a Community Mitigation Framework that encompasses both mandatory and recommended policies at a national level.[105] [106] This framework was interpreted by individual states as they each declared their own States of Emergency at various dates, and subsequently released their own community mitigation plans. Some of the first states to release such plans include Massachusetts, California, Florida, Washington, and New York.[107] Each respective Community Mitigation Framework included both mandatory and optional policies to prevent the COVID-19 spread. In addition to both national and federal level policies and recommendations, cities and counties have also taken on the role of providing guidance and implementing policies to mitigate the spread of COVID-19.

There have been a wide range in responses across states since the first case of COVID-19 was announced in Washington State on January 14, 2020. Upon this, the CDC began releasing recommendations to those at risk of being exposed to the virus. The initial recommendations included travel warnings and restricted travel to countries with confirmed cases and sustained COVID-19 spread. These travel restrictions grew to include inbound and outbound travel bans to a list of 26 countries, in both Europe and Asia.[108]

Other policies have included social distancing, which has been widely recommended or enforced at various levels of government. This method involves avoiding crowds, staying home, limiting or avoiding visiting vulnerable populations (such as long-term care facilities) and standing at least six feet away from others in public spaces.[109] [110] Some regions have implemented school closures at both

---

[104] https://www.theguardian.com/world/2020/mar/13/revolutionary-guards-enforce-coronavirus-controls-iran

[105] https://www.cdc.gov/coronavirus/2019-ncov/whats-new-all.html

[106] https://www.cdc.gov/coronavirus/2019-ncov/downloads/community-mitigation-strategy.pdf

[107] https://www.cdc.gov/coronavirus/2019-ncov/community/index.html

[108] https://www.cdc.gov/coronavirus/2019-ncov/travelers/after-travel-precautions.html

[109] https://www.nytimes.com/2020/03/16/smarter-living/coronavirus-social-distancing.html

the K-12 and higher education level. Business closures have also been recommended or enforced, such that employees should work from home, unless their work is considered essential to the greater public (e.g. health care, grocery stores). To support employees working remotely or staying home when sick, a number of states have also mandated paid sick leave for those who are affected by COVID-19. Free testing has also been implemented in certain states, so that anyone experiencing symptoms or has been exposed to the virus can now get tested for free.[111]

We coded various policies that cancel events and large gatherings as follows: the cancellation of large events, specifically the election postponement in Louisiana, is categorized as "*event_cancel*." The separate "*no_gathering*" policy variable represents policy measures that banned all events or mass gatherings of a certain size, i.e. no gatherings over a certain number of people (where this number has varied by region). The "*social_distance*" category includes measures that prevent visits to elderly care facilities, close public facilities such as libraries, and require workers to work remotely. The "*emergency_declaration*" encompasses the declarations of a state of emergency at the city, county, state, and federal level. This declaration allows the affected area to immediately marshal emergency funds and resources and activate emergency legislation, while also giving the public an indication of the gravity of the situation.

Population Data

In order to construct population weighted policy variables and to determine the susceptible fraction of the population for disease projections under the realized and the "no policy" counterfactual scenarios, we obtained the most recent estimates of population for each administrative unit included in our analysis. The sources of that population data are documented below.

China

City-level population data have been extracted from a compiled dataset of the 2010 Chinese City Statistical Yearbooks. We matched the city level population dataset to the city level COVID-19 epidemiology dataset. As the two datasets use slightly different administrative divisions, we only matched 295 cities that exist in both datasets, and grouped the remaining 39 cities in our compiled epidemiology dataset into "other" for prediction purposes. Cities grouped into "other" because of mismatches have a total population of 114,000,000, or 8.5% of the total population in China.

France

Département-level populations are obtained from the National Institute of Statistics and Economic database https://www.insee.fr/fr/statistiques/2012713#tableau-TCRD_004_tab1_departements.

We used the most up to date estimation of the population in France as of January 2020.

---

[110] https://www.cdc.gov/coronavirus/2019-ncov/community/large-events/mass-gatherings-ready-for-covid-19.html

[111] https://appropriations.house.gov/sites/democrats.appropriations.house.gov/files/Families%20First%20summary.pdf

South Korea

We downloaded the number of population by provinces from a webpage administered by the Korean Statistical Information Service (KOSIS).[112] The government agency recently updated the population information of February, 2020, which we used for our analysis.

Italy

Region and province level population data come from the Italian National Institute of Statistics (Istat), estimating total population on January 1, 2019. The datasets for all Italian regions and provinces are scraped from Istat's website in *get_adm_info.ipynb*.

Iran

Province level population data for Iran comes from the 2016 Census, as listed on the *City Population* website.[113] It is scraped in *get_adm_info.ipynb*.

United States

State and county level population data come from the 2017 American Community Surveys dataset, and is downloaded via the *census* Python package[114] in *get_adm_info.ipynb.*

---

112 http://kosis.kr/statHtml/statHtml.do?orgId=101&tblId=DT_1B040A3&vw_cd=MT_ZTITLE&list_id=A6&seqNo=&lang_mode=ko&language=kor&obj_var_id=&itm_id=&conn_path=MT_ZTITLE

113 https://www.citypopulation.de/en/iran/admin/

114 https://github.com/datamade/census