

Forecasting COVID-19 and Analyzing the Effect of Government Interventions

Michael Lingzhi Li^a, Hamza Tazi Bouardi^a, Omar Skali Lami^a, Thomas A. Trikalinos^c, Nikolaos K. Trichakis^{a,b}, Dimitris Bertsimas^{1a,b}

^aOperations Research Center, Massachusetts Institute of Technology, Cambridge, MA 02139

^bSloan School of Management, Massachusetts Institute of Technology, Cambridge, MA 02142

^cCenter for Evidence Synthesis in Health, Brown University, Providence, RI 02912

Abstract

Background: During the COVID-19 epidemic, governments around the world have implemented unprecedented non-pharmaceutical measures to control its spread. As these measures carry significant economic and humanitarian cost, it is an important topic to investigate the efficacy of different policies and accurately project the future spread under such said policies.

Methods: We developed a novel epidemiological model, DELPHI, based on the established SEIR model, that explicitly captures government interventions, underdetection, and many other realistic effects. We estimate key biological parameters using a meta-analysis of over 190 COVID-19 research papers and fit DELPHI to over 167 geographical areas since early April. We extract the inferred government intervention effect from DELPHI.

Findings: Our epidemiological model recorded 6% and 11% two-week out-of-sample Median Absolute Percentage Error on cases and deaths, and successfully predicted the severity of epidemics in many areas (including US, UK and Russia) months before it happened. Using the extracted government response, we find mass gathering restrictions and school closings on average reduced infection rates the most, at $29.9 \pm 6.9\%$ and $17.3 \pm 6.7\%$, respectively. The most stringent policy, stay-at-home, on average reduced the infection rate by $74.4 \pm 3.7\%$ from baseline across countries that implemented it. We also further show that a reversal of stay-at-home policies in some countries, such as Brazil, could have disastrous results by end of July.

Interpretation: Our findings highlight that among the widely implemented policies around the world, mass gathering restrictions and school closings appear to be the most effective policies in reducing the infection rate. Given the continued spread of the epidemic in many countries, we recommend these policies to continue to the extent that they can be feasibly implemented. Our results also show that under an assumption of R_0 of 2.5-3 for COVID-19, stay-at-home policies appear to be the only effective policy that was widely implemented in reducing the R_0 below 1. This implies that stay-at-home policies might be necessary, for at least the vulnerable population, if an uncontrolled second wave reemerges.

Funding: None.

Keywords: Epidemiology, COVID-19, Government Intervention, Underdetection, Reopening

¹Corresponding author, dbertsim@mit.edu. M.L.L, H.T.B, O.S.L, T.A.T, N.K.T, D.B designed the study, M.L.L, H.T.B, O.S.L acquired data, carried out analysis and formulated the results. M.L.L, H.T.B, O.S.L, T.A.T, N.K.T, D.B wrote the manuscript.

Research in Context

Evidence before this study

Previous research into COVID-19 has focused on reporting estimates of epidemiological parameters of COVID-19. We conducted an extensive literature search on PubMed and MedRxiv including keywords such as "non-pharmaceutical interventions" and "government interventions". We discovered some studies reporting on the theoretical effect of non-pharmaceutical interventions in a theoretical modeling framework. There have also been a few published studies reporting on the overall effect of government interventions in the very early stages of the epidemics in various regions, such as the United States and Europe. However, there were few studies that tried to quantify the effect of each policy that was implemented, and none that the authors know of that are conducted on the global scale of this paper.

Added value of this study

As governments continue to implement non-pharmaceutical interventions, we aim to understand the effect of different policies that have been implemented in the past. We developed a novel epidemiological model that has been continuously providing high accuracy forecasts since early April. It also provides global estimates for the effects of different policies as they have been implemented across 167 areas. The large number of areas we consider enable us to derive inference for many popular policies that have been implemented, including mass gathering restrictions, school closures, along with travel and work restrictions.

Implications of all the available evidence

The evidence indicates that mass gathering restrictions were the most effective single policy in reducing the spread of COVID-19, followed by school closings. Stay-at-home policies greatly reduced the effective R_0 and most likely enabled the effective control of the epidemics in many regions. Policy simulations suggest that many countries around the world are not yet suitable for a loosening of policy guidance, or there would be potentially severe humanitarian costs.

1. Introduction

2 Currently, the world is facing the deadliest pandemic in recent history - COVID-19. As
3 of June 7th, there have been over 7.0 million confirmed cases of COVID-19 and the disease
4 has taken over 400,000 lives. To stop the further spread of COVID-19, governments around
5 the world have enacted some of the most wide-ranging non-pharmaceutical interventions in
6 history. These interventions, especially the more severe ones, carry significant economic and
7 humanitarian cost. Thus, it is critical to understand the effectiveness of such interventions in
8 limiting disease spread.

9 However, there are many challenges in attempting to understand the effect of government
10 interventions in a specific region or country. Different regions have implemented, often
11 concurrently, a variety of different policies, and worse, even the same interventions could produce
12 largely different effects in different societies, due to differences in factors such as demographics,
13 population density, and culture.

14 Thus, in order to provide a sensible analysis of the effect of policies across different coun-
15 tries, in early April we created a novel epidemiological model, DELPHI, to model the spread

16 of COVID-19. DELPHI (Differential Equations Lead to Predictions of Hospitalizations and In-
17 fections) extends a classic SEIR model to include many realistic effects that are critical in this
18 pandemic, including deaths and underdetection. Specifically, we included an explicit nonlinear
19 multiplicative factor on the infection rate to model the spread as it happened in different regions.
20 Such explicit characterization of government intervention allows us to understand the effect of
21 different non-pharmaceutical interventions as they have been implemented in various regions
22 while accounting for regional population characteristics including baseline infection rate and
23 mortality rate. Furthermore, we formulated DELPHI with data scarcity as a key consideration.

24 The aforementioned innovations have allowed DELPHI to produce relatively accurate pro-
25 jections even during the early stages of the epidemic. A major hospital system in the United
26 States planned its intensive care unit (ICU) capacity based on our forecasts. Our epidemiolog-
27 ical predictions are used by a major pharmaceutical company to design a worldwide vaccine
28 distribution strategy that can contain future phases of the pandemic. They have also been in-
29 corporated into the US Center for Disease Control's core ensemble forecast. [1]

30 DELPHI has been applied to 167 geographic areas (countries/provinces/states) worldwide,
31 covering all 6 populated continents. Its results and insights have also been available since early
32 April on www.covidanalytics.io. In this paper, we document the statistical innova-
33 tions, quantitative results, and insights extracted from the DELPHI model.

34 2. Methods

35 2.1. The DELPHI Model

36 The DELPHI model is a compartment epidemiological model that extends the classic SEIR
37 model into 11 states under the following 8 groups:

- 38 • **Susceptible (S):** People who have not been infected.
- 39 • **Exposed (E):** People currently infected, but not contagious and within the incubation
40 period.
- 41 • **Infected (I):** People currently infected and contagious.
- 42 • **Undetected (U_R) & (U_D):** People infected and self-quarantined due to the effects of the
43 disease, but not confirmed due to lack of testing. Some of these people recover (U_R) and
44 some die (U_D).
- 45 • **Detected, Hospitalized (DH_R) & (DH_D):** People who are infected, confirmed, and hos-
46 pitalized. Some of these people recover (DH_R) and some die (DH_D).
- 47 • **Detected, Quarantine (DQ_R) & (DQ_D):** People who are infected, confirmed, and home-
48 quarantined rather than hospitalized. Some of these people recover (DQ_R) and some die
49 (DQ_D).
- 50 • **Recovered (R):** People who have recovered from the disease (and assumed to be im-
51 mune).
- 52 • **Deceased (D):** People who have died from the disease.

53 In addition to main functional states, we introduce auxiliary states to calculate a few useful
54 quantities: Total Hospitalized (TH), Total Detected deaths (DD) and Total Detected Cases (DT).
55 The full mathematical formulation of the model with all the differential equations can be found
56 in the attached Supplementary Materials.

57 Figure 1a depicts a flow representation of the model, where each arrow represents how
58 individuals can flow between different states. The underlying differential equations are gov-
59 erned by 11 explicit parameters which are shown on the appropriate arrows in Figure 1a and

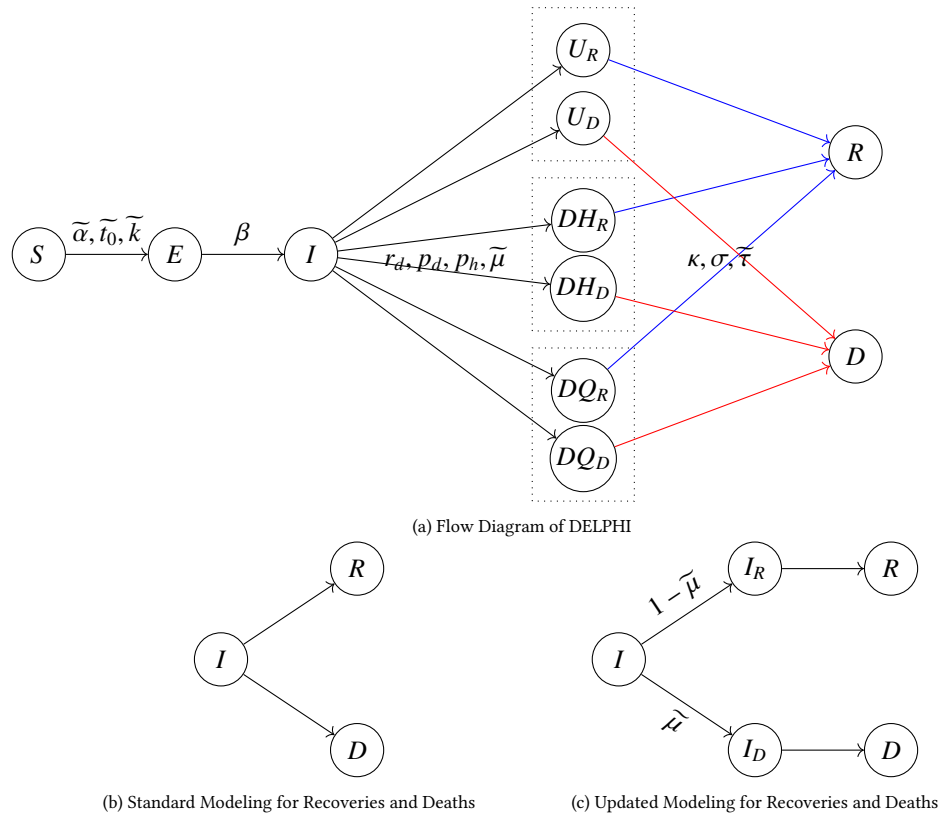


Figure 1: The DELPHI Model

60 defined below. To limit the amount of data needed to train this model, only the parameters denoted with a tilde are being fitted against historical data for each area (country/state/province);
 61 the others are largely biological parameters that are fixed using available clinical data from a
 62 meta-analysis of over 190 papers on COVID-19 available at time of model creation. [2] A small
 63 selection of references for each parameter is given below.

- 64
 65 • $\tilde{\alpha}$ is the baseline infection rate.
 66 • $\gamma(t)$ measures the effect of government response and is defined as:

$$\gamma(t) = \frac{2}{\pi} \arctan\left(\frac{-(t - \tilde{t}_0)}{\tilde{k}}\right) + 1,$$

67 where the parameters \tilde{t}_0 and \tilde{k} capture, respectively, the timing and the strength of the
 68 response. The effective infection rate in the model is $\tilde{\alpha}\gamma(t)$, which is time dependent.

- 69 • r_d is the rate of detection. This equals to $\frac{\log 2}{T_d}$, where T_d is the median time to detection
 70 (fixed to be 2 days). [3]
 71 • β is the rate of infection leaving incubation phase. This equals to $\frac{\log 2}{T_\beta}$, where T_β is the
 72 median time to leave incubation (fixed at 5 days). [4]

- 73 • σ is the rate of recovery of non-hospitalized patients. This equals to $\frac{\log 2}{T_\sigma}$, where T_σ is the
74 median time to recovery of non-hospitalized patients (fixed at 10 days). [5, 6]
- 75 • κ is the rate of recovery under hospitalization. This equals to $\frac{\log 2}{T_\kappa}$, where T_κ is the median
76 time to recovery under hospitalization (fixed at 15 days). [7, 8]
- 77 • $\tilde{\tau}$ is the rate of death. This captures the speed at which a dying patient dies, and thus
78 inversely proportional to how long a dying patient stays alive.
- 79 • $\tilde{\mu}$ is the mortality percentage. This is the percentage of people who die from the disease
80 in a particular region. Note this quantity is independent from the rate of death.
- 81 • p_d is the (constant) percentage of infectious cases detected. This is set to 20%. [3, 9, 10]
- 82 • p_h is the (constant) percentage of detected cases hospitalized. This is set to 15%. [11, 12]

83 Therefore, we fit on 5 parameters from the list above $(\tilde{\alpha}, \tilde{\mu}, \tilde{\tau}, \tilde{t}_0, \tilde{k})$. In addition, we introduce two
84 additional parameters \tilde{k}_1, \tilde{k}_2 to account for the unknown initial population in the infected (I)
85 and exposed (E) states (see Supplementary Materials for details). We thus fit seven parameters
86 per area.

The parameters are fitted by minimizing a weighted Mean Squared Error (MSE) metric with respect to the parameters. Define $DT(t)$ and $DD(t)$ as the number of reported total detected cases and detected deaths, respectively, on day t . Then, the loss function for a training period of T days is defined as:

$$\sum_{t=1}^T t \cdot (\widetilde{DT}(t) - DT(t))^2 + \lambda^2 \cdot \sum_{t=1}^T t \cdot (\widetilde{DD}(t) - DD(t))^2,$$

87 where $\widetilde{DT}(t)$ and $\widetilde{DD}(t)$ are respectively the total detected cases and deaths predicted by DEL-
88 PHI. The factor t gives more prominence to more recent data, as recent errors are more likely
89 to propagate into future errors. The lambda factor $\lambda = \min\left\{\frac{DT(T)}{3 \cdot DD(T)}, 10\right\}$ balances the fitting
90 between detected cases and deaths; this re-scaling coefficient was obtained experimentally. We
91 only include historical data starting when the area recorded more than 100 cases; this allows
92 us to exclude sporadic outbreaks that are not epidemics. Non-convex optimization methods,
93 including trust-region methods [13] and the Nelder-Mead method [14], are utilized to carry out
94 the process of minimization.

95 In the following subsections, we will detail the three key characteristics of the DELPHI
96 model compared to the standard SEIR formulation.

97 2.1.1. Accounting for Under-detection

98 In the COVID-19 crisis, one of the key modeling difficulties is the chronic underdetection
99 of confirmed cases. This is both due to the lack of detection abilities in the early stages of the
100 pandemic and also the similarity between a mild case of COVID-19 and the common flu. Thus,
101 to account for such significant effect, we explicitly included the U_R/U_D states to model people
102 who actually contracted COVID-19 (and are infectious), but were not detected. In particular,
103 we assume that only p_d of the total number of the cases were detected, while $1 - p_d$ of the total
104 cases flow to the U_R/U_D states. There are two methods to gain information on the detection
105 rate: treating p_d as a parameter and fit to the historical data, or recover p_d from serological
106 evidence. However, both methods were impractical during the creation of this model. In an
107 early to mid stage pandemic, a wide range of detection percentages are consistent with the
108 data but leads to vastly different predictions (see e.g. [15]), so historical data could not provide

109 strong evidence. Furthermore, at the time of redaction, the serological data were largely limited
110 to specific sub-areas such as cities and counties (see [16, 17, 18, 19] for examples), while region-
111 wide surveys were largely limited to a few European countries (see [20, 21] for examples and
112 discussion) and only very sparsely available around the world.

113 Thus, we instead fix the detection percentage to be 20% based on various reports trying
114 to understand the extent of underdetection in countries with earlier outbreaks [3, 9, 10]. More
115 recently, an independent study [22] has corroborated our assumption in the United States.

116 2.1.2. Separation of Recovery and Deaths

117 A large focus in many governments' response to the COVID-19 pandemic is to minimize
118 the number of deaths, and thus in DELPHI, we included a death state. In most epidemiological
119 models that extend to include the death state (see e.g. [3, 23] for COVID-19 modeling examples),
120 the death state (D) is shown to flow from the same active infectious state as the recovery state
121 (R), with a schematic shown in Figure 1b. However, this modeling approach would cause the
122 mortality percentage $\tilde{\mu}$ to be dependent on the rates of recovery and death (details are available
123 in the Supplementary Materials). Thus, to resolve such mismatch, we explicitly separated out
124 the $\tilde{\mu}$ fraction of the population infected that would eventually die (I_D) from the $1 - \tilde{\mu}$ fraction
125 that would recover (I_R), as illustrated in Figure 1c.

126 This allows the mortality rate $\tilde{\mu}$ to be independent from the rates of death and recovery. The
127 final DELPHI model further differentiated the I_R states into hospitalized (DH_R), quarantined
128 (DQ_R), and undetected (U_R) states to account for the different treatments people received, and
129 similarly with the I_D states.

130 2.1.3. Modeling Effect of Increasing Government Response

131 One of the key assumptions in the standard SEIR model is that the rate of infection α is
132 constant throughout the epidemic. However, in real epidemics such as the COVID-19 crisis, the
133 rate of infection starts decreasing as governments respond to the spread of epidemic, and induce
134 behavior changes in societies. To account for such effect, we model the effect of government
135 measures with a sigmoid-like function $\gamma(t)$ (specifically the inverse tangent).

136 The concave-convex nature of an arctan curve models three phases: The early, concave part
137 of the arctan models limited changes in behavior in response to early information, while most
138 people continue business-as-usual activities. The transition from the concave to the convex
139 part of the curve quantifies the sharp decline in infection rate as policies go into full force and
140 the society experiences a shock event. The latter convex part of the curve models a flattening
141 out of the response as the government measures reach saturation, representing the diminishing
142 marginal returns in the decline of infection rate. An illustration of such three phases is included
143 in the Supplementary Materials.

144 Parameters \tilde{t}_0 and \tilde{k} control the timing of such measures and the rapidity of their penetra-
145 tion. This formulation allows us to model, under the same framework, a wide variety of policies
146 that different governments impose, including social distancing, stay-at-home policies, quaran-
147 tines, etc. This modeling captures the increasing force of intervention in the early-mid stages
148 of the epidemic. In Section 3.3, we would show how this model can be extended to provide
149 insights on the relaxation of measures.

150 3. Results and Discussion

151 3.1. Forecasting Results

152 DELPHI was created in early April and has been continuously updated to reflect new ob-
153 served data. Figures 2a and 2b show our projections of the number of cases in Russia and the
154 United Kingdom made on three different dates, and compare them against historical observa-
tions. They suggest that DELPHI achieves strong predictive performance, as the model has been

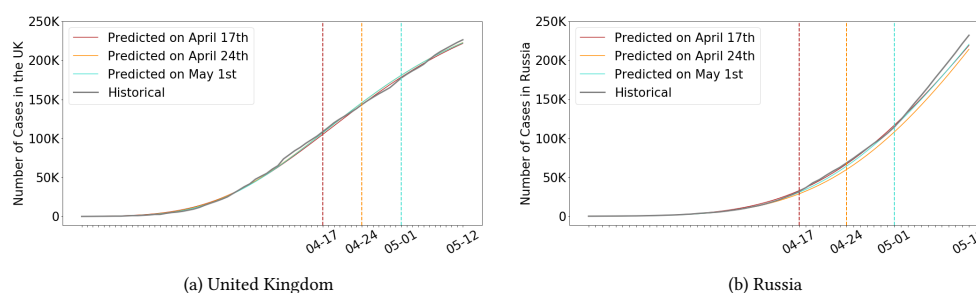


Figure 2: Cumulative number of cases in the UK (a) and Russia (b) according to our projections made at different points in time, against actual observations. Note there predicted curves largely overlap with the actual curve.

155 consistently predicting, with high accuracy, the overall spread of the disease for several weeks.
156 Notably, DELPHI was able to anticipate, as early as April 17th, the dynamics of the pandemic
157 in the United Kingdom (resp. Russia) up to May 12th. At a time when 100-110K (resp. 30-35K)
158 cases were reported, the model was predicting 220-230K (resp. 225-235K) cases by May 12th—a
159 prediction that became accurate a month later.
160

161
162 Furthermore, Table 1 reports the median Mean Absolute Percentage Error (MAPE) on the
163 observed total cases and deaths in each area of the world using parameters obtained on April
164 28th, and evaluated on the 15 days period up until May 12th. Overall, our model seems to predict
165 the epidemic progression relatively well in most countries with $< 10\%$ MAPE on reported cases,
166 and $< 15\%$ MAPE on reported deaths. Additionally, the areas with the highest errors are often
167 those that have the fewest deaths. This stems from the fact that DELPHI—like all SEIR-based
168 models—is not designed to perform well on areas with small populations and interactions. The
169 effect is further exacerbated by the choice of the metric, as MAPE inherently heavily penalizes
170 errors on small numbers. Further detailed results for each country/region that we predict, and
171 examples of areas with high MAPE, can be found in the Supplementary Materials.

172 3.2. Effect of Government Interventions

173 A natural application of the DELPHI model is policy evaluation. For that, we can extract
174 the normalized fitted government response curve $\gamma(t)$ in each area, and utilize it to understand
175 the impact of specific government policies that have been implemented. In particular, we aim
176 to understand the average effect of each policy on $\gamma(t)$ during the period of implementation. To
177 this end, for all countries except US, we collect data from the Oxford Coronavirus Government
178 Response Tracker for historical data on government policies [24], during the period between
179 January 1st 2020, and May 19th 2020. For the US, we collect the policy data from the Institute
180 for Health Metrics and Evaluation [25] during the same period.

| Region | # Areas | Median MAPE Cases (10th, 90th percentile) | Median MAPE Deaths (10th, 90th percentile) |
|---------------|---------|--|---|
| Africa | 19 | 14.7% (3.1, 32.0) | 23.4% (11.8, 60.3) |
| Asia | 32 | 4.8% (2.1, 18.4) | 14.4% (2.9, 65.2) |
| Europe | 42 | 3.4% (0.8, 12.9) | 9.0% (2.3, 24.3) |
| North America | 10 | 7.9% (3.9, 28.3) | 12.6% (2.8, 23.6) |
| Oceania | 2 | 3.2% (2.4, 4.1) | 12.0% (11.0, 13.0) |
| South America | 11 | 14.9% (7.6, 26.7) | 6.1% (3.3, 30.1) |
| United States | 51 | 8.5% (1.9, 16.7) | 7.8% (3.3, 25.1) |
| World | 167 | 5.8% (1.5, 22.6) | 10.6% (2.9, 36.6) |

Table 1: Median country-level Mean Absolute Percentage Error (MAPE) of the predicted number of cases and deaths in each continent (projections made using data up to 04/27 for the period from 04/28 to 05/12).

181 At each point in time, we categorize the government intervention data based on whether
182 they restrict mass gatherings, schools, travel and work activities. We group travel restrictions
183 and work restrictions together due to their tendency to be implemented simultaneously. From
184 January 1st to May 19th, the 167 areas in total implemented 5 combinations of such interven-
185 tions. Specifically, these are: (1) *No measure*; (2) *Restrict travel and work only*; (3) *Restrict mass*
186 *gatherings, travel and work*; (4) *Restrict mass gatherings, schools, travel and work*; and (5) *Stay-at-*
187 *Home*. The detailed correspondence between raw policy data and our categories are contained
188 in the Supplementary Materials. Other potentially feasible combinations were not implemented
189 by the countries. Then for each policy category $i = 1, \dots, 5$, we extract the average value of
190 $\gamma(t)$, $\bar{\gamma}_i$, across all time periods and areas for which policy i was implemented. Then we calculate
191 the residual fraction of infection rate under policy i , p_i , compared to the baseline policy of no
192 measure:

$$p_i = \frac{\bar{\gamma}_i}{\bar{\gamma}_1}$$

| Restrictions | Area-Days | Residual Infection Rate |
|--|-----------|-------------------------|
| None | 2142 | 100% |
| Travel and Work | 2049 | $88.9 \pm 4.5\%$ |
| Mass Gathering, Travel, and Work | 340 | $59.0 \pm 5.2\%$ |
| Mass Gathering, School, Travel, and Work | 1460 | $41.7 \pm 4.3\%$ |
| Stay-at-Home Order | 6585 | $25.6 \pm 3.7\%$ |

Table 2: Implementation Length and Effect of each policy category as implemented across the world.

193 Table 2 shows the number of area-days that each policy was implemented around the world
194 and its effect. We further report the standard deviation of such estimate treating each geograph-
195 ical area as an independent sample. We see that each selected policy was enacted for at least
196 hundreds of Area-Days worldwide, while the stringent stay-at-home policy was cumulatively
197 implemented the most. In particular, we see that mass gathering restrictions generate a large
198 reduction in infection rate, with the incremental reduction between travel and work restrictions
199 compared to mass gathering, travel, and work restrictions is $29.9 \pm 6.9\%$. This is further sup-

ported by the large residual infection rate of $88.9 \pm 4.5\%$ when travel and work restrictions are implemented, but mass gatherings are allowed. Additionally, we observe that closing schools also generate a large reduction in the infection rate, with an incremental effect of $17.3 \pm 6.6\%$ on top of mass gathering and other restrictions. Stay-at-home orders produced the strongest reduction in infection rate across the different countries, with a residual infection rate of just $25.6 \pm 3.7\%$ compared to when no measure was implemented.

If COVID-19 has an average basic reproductive number R_0 of 2.5-3 ([26, 27]), then on average, only the strongest measure (Stay-at-Home orders) are sufficient to control a COVID-19 epidemic in reducing R_0 to be less than 1.

3.3. Extension: Evaluating Reopening Strategies

The DELPHI model provides insights into the effect of government policies through the residual infection rates p_i under each policy.

A natural extension is to utilize the p_i in creating what-if scenarios on the effect of lifting restrictions in different countries by reverting the effect of each policy on $\gamma(t)$ at the time of the hypothetical policy relaxation. Specifically, suppose that we are considering a policy easing from policy i to j at time t_c in some area. Then for all times $t \geq t_c$, we correct the government response as follows:

$$\gamma'(t) = \frac{2}{\pi} \arctan\left(-\frac{t - \tilde{t}_0}{k}\right) + 1 + \underbrace{(p_j - p_i) \cdot \min\left[\frac{2 - \gamma(t_c)}{1 - p_i}, \frac{\gamma(t_c)}{p_i}\right]}_{\text{Differential in policy effect between policy } i \text{ and } j}, \quad \forall t \geq t_c.$$

Essentially, we apply a correction term that is proportional to the fractional difference in policy effect between policy i and j (which is $p_j - p_i > 0$ as it is an easing). The multiplicative factor $\min\left[\frac{2 - \gamma(t_c)}{1 - p_i}, \frac{\gamma(t_c)}{p_i}\right]$ scales the fractional difference so that the resulting $\gamma'(t_c)$ is constrained within the initial range $[0, 2]$. Then, we would replace $\gamma(t)$ with $\gamma'(t)$ in the DELPHI model to forecast the epidemic under the updated policy. Using this correction factor, we predict what would happen in different areas under various future policies. Figure 3 shows results for France and Brazil respectively, under policy change implemented on June 16th (four weeks after the last historical value on May 19th). Further results for other countries are contained in the Supplementary Material.

We observe different levels of risk for the same re-opening strategies across different countries. For example, Figure 3c predicts that loosening measures in Brazil on June 16th would result in a second wave of infections with up to 6.8 million additional cases by July 15th, while even a stay-at-home order would lead to almost 1.9 million additional cases. Such alarming numbers can be understood through Figure 3d where we compute a rolling average of the weekly incidence of cases per 100K people. We can see that Brazil is still on a steep ascending curve, and that any kind of loosening could be catastrophic. Such behaviour stands in sharp contrast with France's situation. Figure 3b demonstrates that the peak has long passed in France and the epidemic has mostly died out. Thus, as we can see in Figure 3a, loosening policies (like France has already started doing) is likely to only minimally affect the number of infections.

To further understand the disparate impact of the policies across countries, we made predictions for the situation around the world assuming a policy that involves mass gathering, travel, and work restrictions was universally implemented on 06/16. Figure 4a shows three clusters of countries for July 15th:

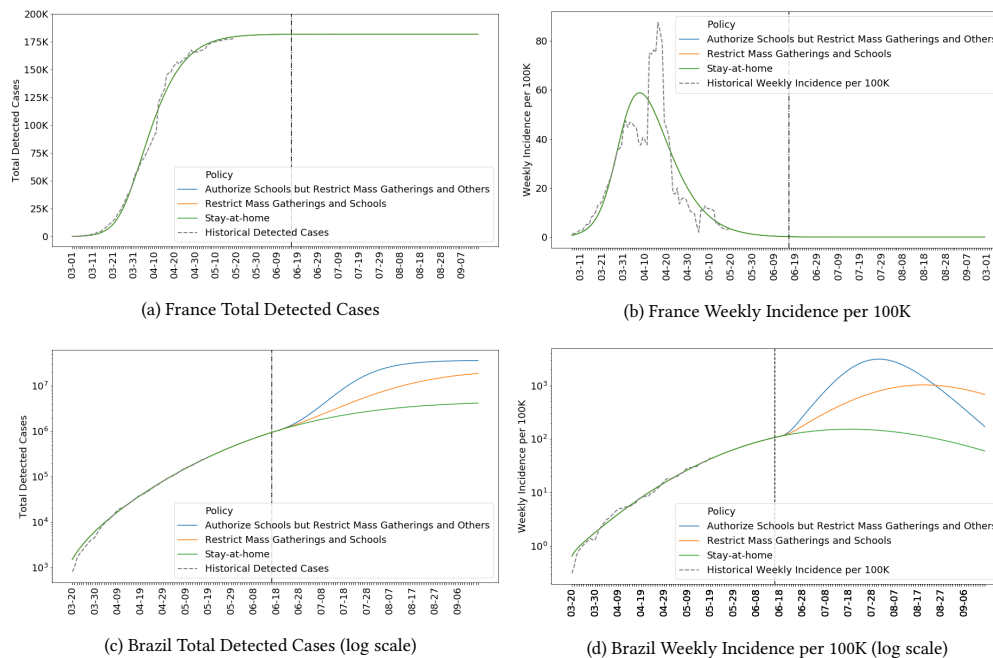
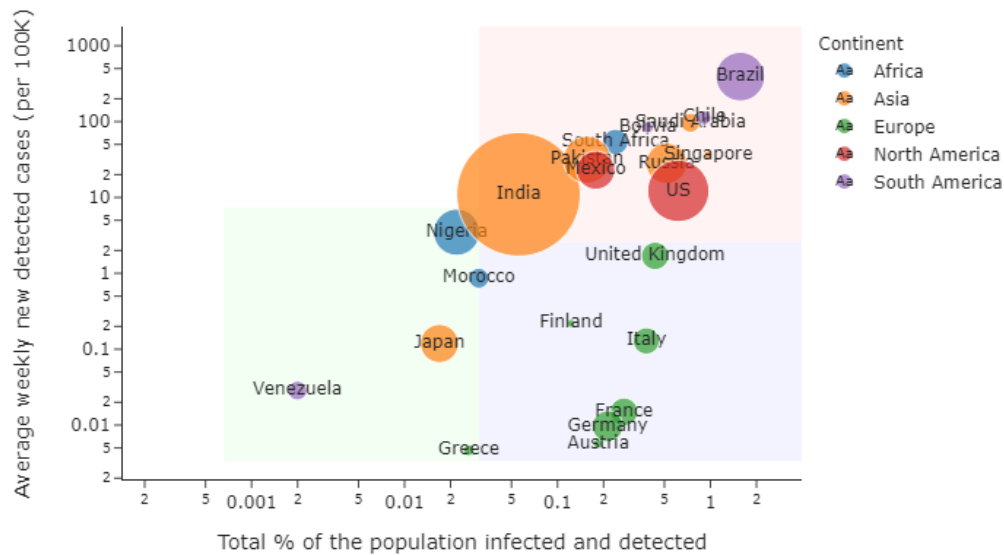


Figure 3: Forecasts of total detected cases and weekly incidence per 100K for France and Brazil under various policies

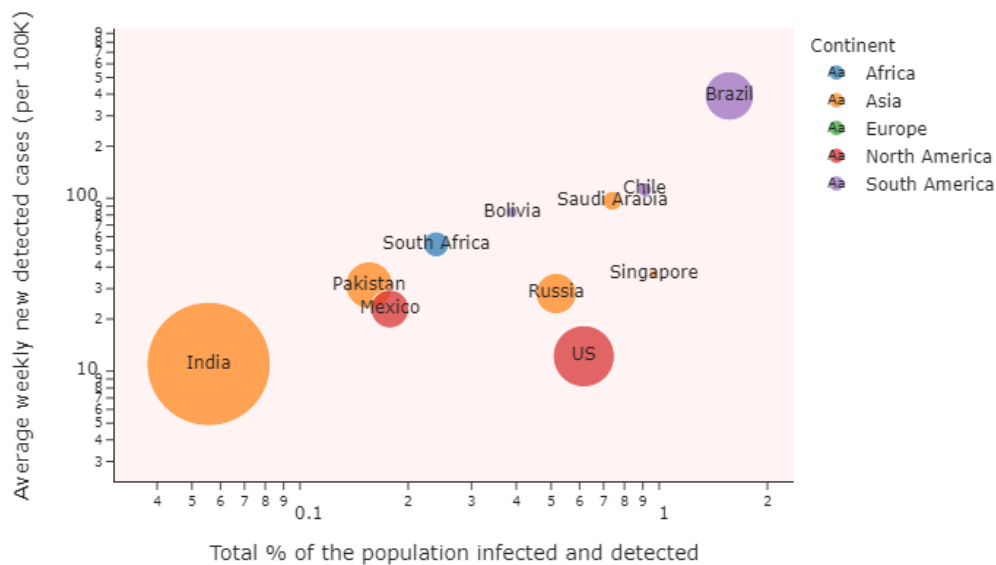
- 235
- 236
- 237
- 238
- 239
- 240
- 241
- 242
- 243
- 244
- 245
- 246
- 247
- Countries with a small number of cumulative cases (relative to the population), and that are in a late stage of the pandemic with relatively few new cases, such as Greece, Japan, Morocco and Venezuela.
 - Countries with a large number of cumulative cases, but that are in a late stage of the pandemic, with relatively few new cases, mainly in Western and Northern Europe (e.g. the United Kingdom, Italy, France and Finland).
 - Countries where the pandemic has had a large impact with a large number of cumulative cases, and where the situation will still be worsening at an alarming rate. These include the United States, India and Brazil. A close-up of these countries is presented in Figure 4b, where we see that DELPHI predicts Brazil would be severely hit by July, with up to 8% of the entire population infected, if the hypothetical policy above is implemented. This suggests that in these countries, such hypothetical policy could be inadequate for controlling the epidemic, and a stronger policy (such as Stay-at-Home orders) is needed.

248 **4. Limitations**

249 One fundamental limitation of this analysis is its observational nature. Thus, despite the
 250 flexible parameters in DELPHI accounting for many state-dependent effects, there are many
 251 other potential confounders and second-order effects that could affect this analysis. For ex-
 252 ample, one effect that is not considered in DELPHI is a time-varying mortality rate caused by
 253 changing treatment procedures designed to best help COVID-19. Including such effect could



(a) Weekly Incidence of Cases (per 100K) in the first half of July against fraction of population infected for multiple countries



(b) Predictions for total cumulative cases (normalized by the population) vs new cases (per 100K) for countries which are predicted to be highly impacted and still worsening at an alarming rate by July 15th

Figure 4: World Predictions for Early July under Mass Gathering, Travel and Work Restrictions

254 sharpen the analysis further, though at the expense of increased fitting difficulty and data re-
255 quirements.

256 This analysis also assumes, in analyzing government interventions, that the same nominal
257 policy (e.g. Mass gathering restrictions) could be compared across countries. In reality, different
258 countries have implemented variants (though largely similar) of restrictions under the same
259 name, and this could further impact the validity of the analysis.

260 In the reopening analysis, we have assumed that the effect of government interventions
261 imposed at the start of the epidemic is indicative of the effect when it is removed. This is
262 potentially affected by a permanent change in social behavior during the epidemic. For example,
263 if a significant portion of the population adapts social distancing measure even after the official
264 restrictions are lifted, this could lead to a smaller resurgence of infections than what is predicted
265 in the analysis.

266 5. Conclusions

267 We introduced DELPHI, a novel epidemiological model that extended SEIR to include many
268 realistic effects critical in this pandemic. DELPHI was able to accurately predict the spread of
269 COVID-19 in many countries, and aid planning for many organizations worldwide. Further-
270 more, the explicit modeling on government intervention allowed us to understand the effect of
271 government interventions, and help inform how societies could reopen.

272 Bibliography

- 273 [1] Covid-19 forecasts: Cumulative deaths, 2020. URL: [https://www.cdc.gov/coronavirus/](https://www.cdc.gov/coronavirus/2019-ncov/covid-data/forecasting-us.html)
274 [2019-ncov/covid-data/forecasting-us.html](https://www.cdc.gov/coronavirus/2019-ncov/covid-data/forecasting-us.html).
- 275 [2] D. Bertsimas, H. Bandi, L. Boussioux, R. Cory-Wright, A. Delarue, V. Digalakis, S. Gilmour, J. Graham, A. Kim,
276 D. Lahlou Kitane, Z. Lin, G. Lukin, M. Li, L. Mingardi, L. Na, A. Orfanoudaki, T. Papalexopoulos, I. Paskov,
277 J. Pauphilet, O. Skali Lami, M. Sobiesk, B. Stellato, K. Carballo, Y. Wang, H. Wiberg, C. Zeng, An aggregated dataset
278 of clinical outcomes for covid-19 patients, 2020. URL: <http://www.covidanalytics.io/dataset>
279 [documentation](http://www.covidanalytics.io/dataset).
- 280 [3] C. Wang, L. Liu, X. Hao, H. Guo, Q. Wang, J. Huang, N. He, H. Yu, X. Lin, A. Pan, et al., Evolving epidemiology
281 and impact of non-pharmaceutical interventions on the outbreak of coronavirus disease 2019 in wuhan, china,
282 *medRxiv* (2020).
- 283 [4] S. A. Lauer, K. H. Grantz, Q. Bi, F. K. Jones, Q. Zheng, H. R. Meredith, A. S. Azman, N. G. Reich, J. Lessler, The
284 incubation period of coronavirus disease 2019 (covid-19) from publicly reported confirmed cases: estimation and
285 application, *Annals of internal medicine* 172 (2020) 577–582.
- 286 [5] Z. Hu, C. Song, C. Xu, G. Jin, Y. Chen, X. Xu, H. Ma, W. Chen, Y. Lin, Y. Zheng, et al., Clinical characteristics of
287 24 asymptomatic infections with covid-19 screened among close contacts in nanjing, china, *Science China Life*
288 *Sciences* (2020) 1–6.
- 289 [6] M. Kluytmans, A. Buiting, S. Pas, R. Bentvelsen, W. van den Bijllaardt, A. van Oudheusden, M. van Rijen, J. Verweij,
290 M. Koopmans, J. Kluytmans, Sars-cov-2 infection in 86 healthcare workers in two dutch hospitals in march 2020,
291 *medRxiv* (2020).
- 292 [7] Y. Liu, W. Sun, L. Chen, Y. Wang, L. Zhang, L. Yu, Clinical characteristics and progression of 2019 novel
293 coronavirus-infected patients concurrent acute respiratory distress syndrome, *medRxiv* (2020).
- 294 [8] J. Grein, N. Ohmagari, D. Shin, G. Diaz, E. Asperges, A. Castagna, T. Feldt, G. Green, M. L. Green, F.-X. Lescure,
295 et al., Compassionate use of remdesivir for patients with severe covid-19, *New England Journal of Medicine*
296 (2020).
- 297 [9] S. G. Krantz, A. S. S. Rao, Level of under-reporting including under-diagnosis before the first peak of covid-19
298 in various countries: Preliminary retrospective results based on wavelets and deterministic modeling, *Infection*
299 *Control & Hospital Epidemiology* (2020) 1–8.
- 300 [10] R. Niehus, P. Martinez de Salazar Munoz, A. Taylor, M. Lipsitch, Quantifying bias of covid-19 prevalence and
301 severity estimates in wuhan, china that depend on reported cases in international travelers, *medRxiv* (2020).

- 302 [11] M. M. Arons, K. M. Hatfield, S. C. Reddy, A. Kimball, A. James, J. R. Jacobs, J. Taylor, K. Spicer, A. C. Bardossy, L. P.
303 Oakley, et al., Presymptomatic sars-cov-2 infections and transmission in a skilled nursing facility, *New England*
304 *Journal of Medicine* (2020).
- 305 [12] H. Xu, S. Huang, S. Liu, J. Deng, B. Jiao, L. Ai, Y. Xiao, L. Yan, S. Li, Evaluation of the clinical characteristics of
306 suspected or confirmed cases of covid-19 during home care with isolation: A new retrospective analysis based on
307 o2o, Available at SSRN 3548746 (2020).
- 308 [13] R. H. Byrd, J. C. Gilbert, J. Nocedal, A trust region method based on interior point techniques for nonlinear
309 programming, *Mathematical programming* 89 (2000) 149–185.
- 310 [14] J. C. Lagarias, J. A. Reeds, M. H. Wright, P. E. Wright, Convergence properties of the nelder–mead simplex method
311 in low dimensions, *SIAM Journal on optimization* 9 (1998) 112–147.
- 312 [15] J. Lourenço, R. Paton, M. Ghafari, M. Kraemer, C. Thompson, P. Simmonds, P. Klenerman, S. Gupta, Fundamental
313 principles of epidemic spread highlight the immediate need for large-scale serological surveys to assess the stage
314 of the sars-cov-2 epidemic, *medRxiv* (2020).
- 315 [16] A. Doi, K. Iwata, H. Kuroda, T. Hasuike, S. Nasu, A. Kanda, T. Nagao, H. Nishioka, K. Tomii, T. Morimoto, et al.,
316 Estimation of seroprevalence of novel coronavirus disease (covid-19) using preserved serum at an outpatient
317 setting in kobe, japan: A cross-sectional study., *medRxiv* (2020).
- 318 [17] H. Streeck, B. Schulte, B. Kuemmerer, E. Richter, T. Höller, C. Fuhrmann, E. Bartok, R. Dolscheid, M. Berger,
319 L. Wessendorf, et al., Infection fatality rate of sars-cov-2 infection in a german community with a super-spreading
320 event, *medRxiv* (2020).
- 321 [18] N. Sood, P. Simon, P. Ebner, D. Eichner, J. Reynolds, E. Bendavid, J. Bhattacharya, Seroprevalence of sars-cov-2–
322 specific antibodies among adults in los angeles county, california, on april 10-11, 2020, *JAMA* (2020).
- 323 [19] E. Bendavid, B. Mulaney, N. Sood, S. Shah, E. Ling, R. Bromley-Dulfano, C. Lai, Z. Weissberg, R. Saavedra,
324 J. Tedrow, et al., Covid-19 antibody seroprevalence in santa clara county, california, *MedRxiv* (2020).
- 325 [20] C. Erikstrup, C. E. Hother, O. B. V. Pedersen, K. Mølbak, R. L. Skov, D. K. Holm, S. Sækmose, A. C. Nilsson, P. T.
326 Brooks, J. K. Boldsen, et al., Estimation of sars-cov-2 infection fatality rate by real-time antibody screening of
327 blood donors, *medRxiv* (2020).
- 328 [21] J. Wise, Covid-19: Surveys indicate low infection level in community, 2020.
- 329 [22] T. Breton, An estimate of unidentified and total us coronavirus cases by state on april 21, 2020, SSRN (2020).
- 330 [23] L. Peng, W. Yang, D. Zhang, C. Zhuge, L. Hong, Epidemic analysis of covid-19 in china by dynamical modeling,
331 *arXiv preprint arXiv:2002.06563* (2020).
- 332 [24] T. Hale, S. Webster, A. Petherick, T. Phillips, B. Kira, Oxford covid-19 government response tracker, blavat-
333 nik school of government, Oxford University. Creative Commons Attribution CC BY standard. Available at:
334 <https://www.bsg.ox.ac.uk/covidtracker>. Accessed on: April 14 (2020) 2020.
- 335 [25] C. Murray, et al., Forecasting the impact of the first wave of the covid-19 pandemic on hospital demand and deaths
336 for the usa and european economic area countries, *medRxiv* (2020).
- 337 [26] S. Zhang, M. Diao, W. Yu, L. Pei, Z. Lin, D. Chen, Estimation of the reproductive number of novel coronavirus
338 (covid-19) and the probable outbreak size on the diamond princess cruise ship: A data-driven analysis, *Internation-
339 al Journal of Infectious Diseases* 93 (2020) 201–204.
- 340 [27] Y. Liu, A. A. Gayle, A. Wilder-Smith, J. Rocklöv, The reproductive number of covid-19 is higher compared to sars
341 coronavirus, *Journal of travel medicine* (2020).