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]

[a]*Operations Research Center, Massachusetts Institute of Technology, Cambridge, MA 02139*
[b]*Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA 02142*
[c]*Center 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 ± 6.9% and 17.3 ± 6.7%, respectively. The most stringent policy, stay-at-home, on average reduced the infection rate by 74.4 ± 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 R0 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 R0 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

[1]Corresponding 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 R0 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

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

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

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

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

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

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

2. Methods

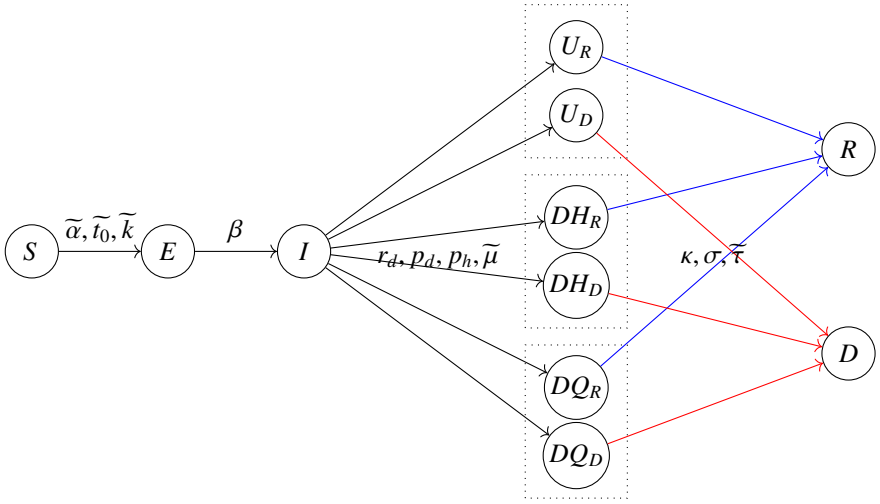
2.1. The DELPHI Model

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

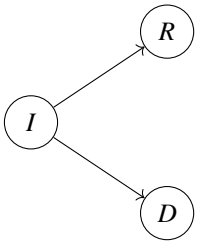
- **Susceptible (S)**: People who have not been infected.
- **Exposed (E)**: People currently infected, but not contagious and within the incubation period.
- **Infected (I)**: People currently infected and contagious.
- **Undetected (U_R) & (U_D)**: People infected and self-quarantined due to the effects of the disease, but not confirmed due to lack of testing. Some of these people recover (U_R) and some die (U_D).
- **Detected, Hospitalized (DH_R) & (DH_D)**: People who are infected, confirmed, and hospitalized. Some of these people recover (DH_R) and some die (DH_D).
- **Detected, Quarantine (DQ_R) & (DQ_D)**: People who are infected, confirmed, and home-quarantined rather than hospitalized. Some of these people recover (DQ_R) and some die (DQ_D).
- **Recovered (R)**: People who have recovered from the disease (and assumed to be immune).
- **Deceased (D)**: People who have died from the disease.

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

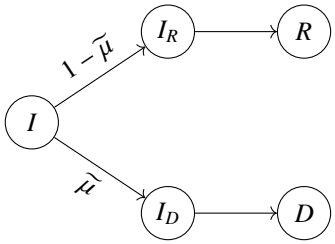
Figure 1a depicts a flow representation of the model, where each arrow represents how individuals can flow between different states. The underlying differential equations are governed by 11 explicit parameters which are shown on the appropriate arrows in Figure 1a and



(a) Flow Diagram of DELPHI



(b) Standard Modeling for Recoveries and Deaths



(c) Updated Modeling for Recoveries and Deaths

Figure 1: The DELPHI Model

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); the others are largely biological parameters that are fixed using available clinical data from a meta-analysis of over 190 papers on COVID-19 available at time of model creation. [2] A small selection of references for each parameter is given below.

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

$$\gamma(t) = \frac{2}{\pi} \arctan\left(\frac{-(t-\widetilde{t_0})}{\widetilde{k}}\right) + 1,$$

  where the parameters $\widetilde{t_0}$ and $\widetilde{k}$ capture, respectively, the timing and the strength of the response. The effective infection rate in the model is $\widetilde{\alpha}\gamma(t)$, which is time dependent.
- r_d is the rate of detection. This equals to $\frac{\log 2}{T_d}$, where T_d is the median time to detection (fixed to be 2 days). [3]
- β is the rate of infection leaving incubation phase. This equals to $\frac{\log 2}{T_\beta}$, where T_β is the median time to leave incubation (fixed at 5 days). [4]

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

Therefore, we fit on 5 parameters from the list above $(\widetilde{\alpha}, \widetilde{\mu}, \widetilde{\tau}, \widetilde{t_0}, \widetilde{k})$. In addition, we introduce two additional parameters $\widetilde{k_1}$, $\widetilde{k_2}$ to account for the unknown initial population in the infected (I) and exposed (E) states (see Supplementary Materials for details). We thus fit seven parameters 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:

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

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

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

2.1.1. Accounting for Under-detection

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

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

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

2.1.2. Separation of Recovery and Deaths

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

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

2.1.3. Modeling Effect of Increasing Government Response

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

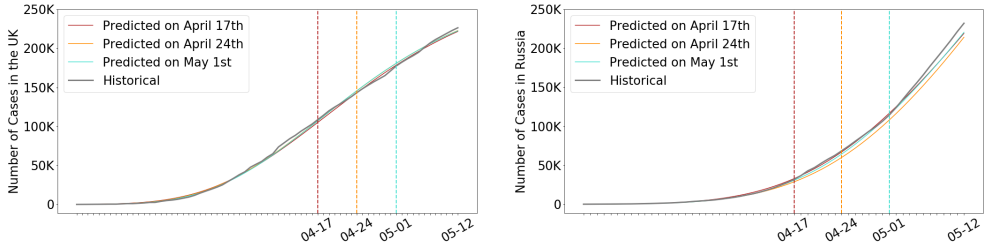
The concave-convex nature of an arctan curve models three phases: The early, concave part of the arctan models limited changes in behavior in response to early information, while most people continue business-as-usual activities. The transition from the concave to the convex part of the curve quantifies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

Parameters $\widetilde{t_0}$ and $\widetilde{k}$ control the timing of such measures and the rapidity of their penetration. This formulation allows us to model, under the same framework, a wide variety of policies that different governments impose, including social distancing, stay-at-home policies, quarantines, etc. This modeling captures the increasing force of intervention in the early-mid stages of the epidemic. In Section 3.3, we would show how this model can be extended to provide insights on the relaxation of measures.

3. Results and Discussion

3.1. Forecasting Results

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



(a) United Kingdom (b) Russia

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.

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

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

3.2. Effect of Government Interventions

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

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

Table 2 shows the number of area-days that each policy was implemented around the world and its effect. We further report the standard deviation of such estimate treating each geographical area as an independent sample. We see that each selected policy was enacted for at least hundreds of Area-Days worldwide, while the stringent stay-at-home policy was cumulatively implemented the most. In particular, we see that mass gathering restrictions generate a large reduction in infection rate, with the incremental reduction between travel and work restrictions 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.

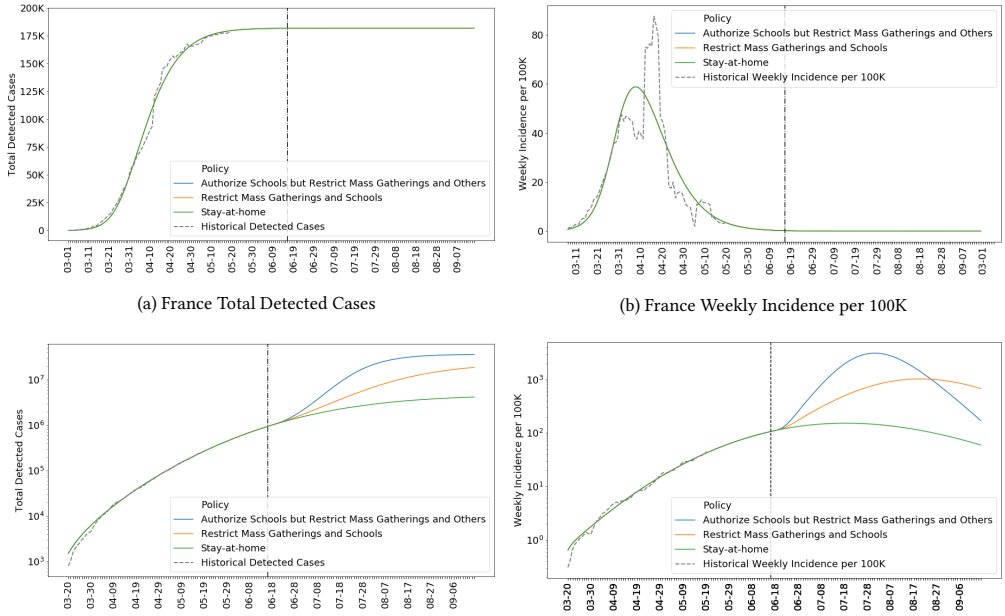
A natural extension is to utilize the p_i in creating what-if scenarios on the effect of lifting restrictions in different countries by reverting the effect of each policy on $\gamma(t)$ at the time of the hypothetical policy relaxation. Specifically, suppose that we are considering a policy easing from 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:

$$\gamma'(t) = \frac{2}{\pi} \arctan\left(-\frac{t - \widetilde{t_0}}{\widetilde{k}}\right) + 1 + \underbrace{(p_j - p_i) \cdot \min\left[\frac{2 - \gamma(t_c)}{1 - p_i}, \frac{\gamma(t_c)}{p_i}\right]}_{\text{Differential in policy effect between policy } i \text{ and } j}, \quad \forall t \geq t_c.$$

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:



(a) France Total Detected Cases

(b) France Weekly Incidence per 100K

(c) Brazil Total Detected Cases (log scale)

(d) Brazil Weekly Incidence per 100K (log scale)

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

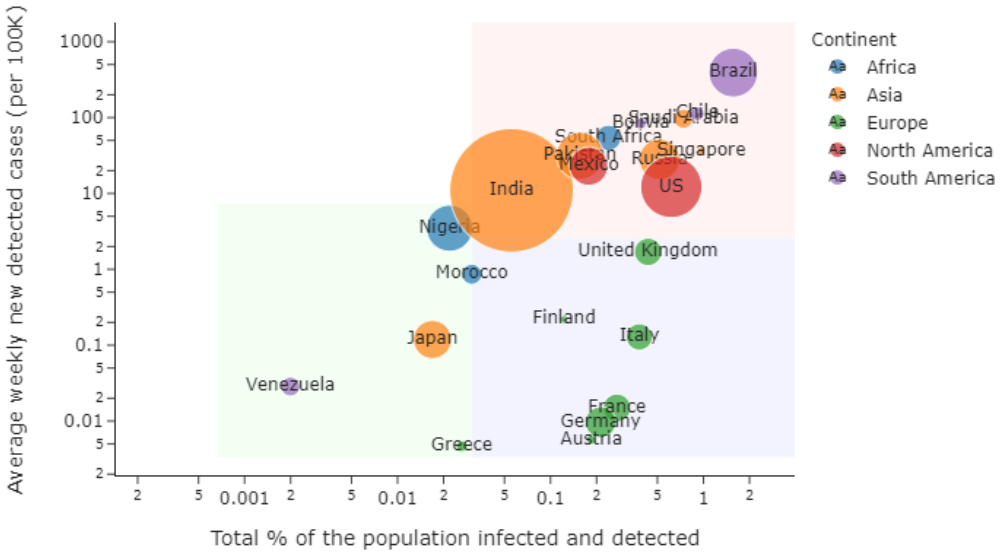
- Countries with a small number of cumulative cases (relative to the population), and that are in a late stage of the pandemic with relatively few new cases, such as Greece, Japan, Morocco and Venezuela.

- 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).

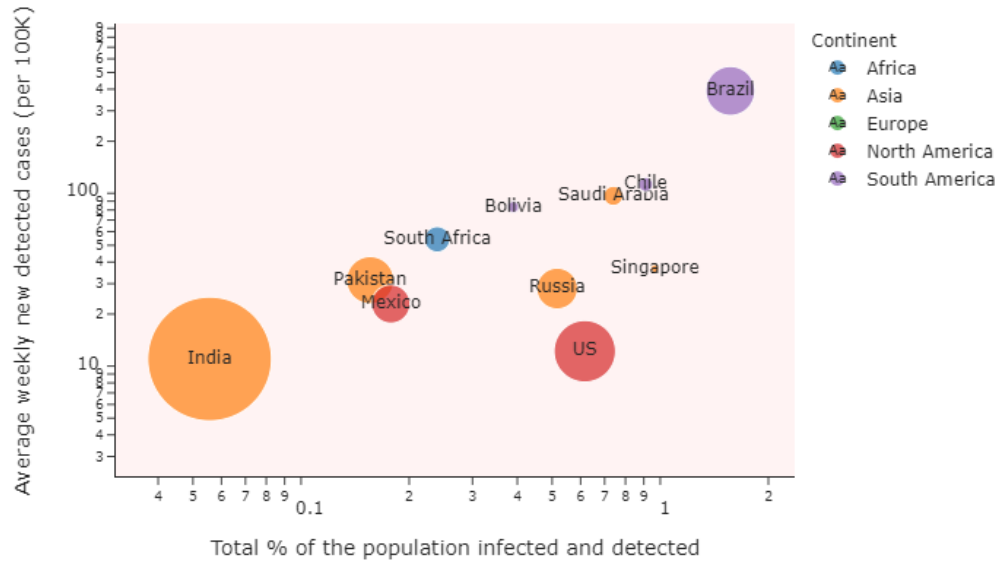
- Countries where the pandemic has had a large impact with a large number of cumulative cases, and where the situation will still be worsening at an alarming rate. These include the United States, India and Brazil. A close-up of these countries is presented in Figure 4b, where we see that DELPHI predicts 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.

4. Limitations

One fundamental limitation of this analysis is its observational nature. Thus, despite the flexible parameters in DELPHI accounting for many state-dependent effects, there are many other potential confounders and second-order effects that could affect this analysis. For example, one effect that is not considered in DELPHI is a time-varying mortality rate caused by 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

sharpen the analysis further, though at the expense of increased fitting difficulty and data requirements.

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

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

5. Conclusions

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

Bibliography

[1] Covid-19 forecasts: Cumulative deaths, 2020. URL: https://www.cdc.gov/coronavirus/2019-ncov/covid-data/forecasting-us.html.

[2] D. Bertsimas, H. Bandi, L. Boussioux, R. Cory-Wright, A. Delarue, V. Digalakis, S. Gilmour, J. Graham, A. Kim, D. Lahlou Kitane, Z. Lin, G. Lukin, M. Li, L. Mingardi, L. Na, A. Orfanoudaki, T. Papalexopoulos, I. Paskov, J. Pauphilet, O. Skali Lami, M. Sobiesk, B. Stellato, K. Carballo, Y. Wang, H. Wiberg, C. Zeng, An aggregated dataset of clinical outcomes for covid-19 patients, 2020. URL: http://www.covidanalytics.io/datasetdocumentation.

[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 and impact of non-pharmaceutical interventions on the outbreak of coronavirus disease 2019 in wuhan, china, medRxiv (2020).

[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 incubation period of coronavirus disease 2019 (covid-19) from publicly reported confirmed cases: estimation and application, Annals of internal medicine 172 (2020) 577–582.

[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 24 asymptomatic infections with covid-19 screened among close contacts in nanjing, china, Science China Life Sciences (2020) 1–6.

[6] M. Kluytmans, A. Buiting, S. Pas, R. Bentvelsen, W. van den Bijllaardt, A. van Oudheusden, M. van Rijen, J. Verweij, M. Koopmans, J. Kluytmans, Sars-cov-2 infection in 86 healthcare workers in two dutch hospitals in march 2020, medRxiv (2020).

[7] Y. Liu, W. Sun, L. Chen, Y. Wang, L. Zhang, L. Yu, Clinical characteristics and progression of 2019 novel coronavirus-infected patients concurrent acute respiratory distress syndrome, medRxiv (2020).

[8] J. Grein, N. Ohmagari, D. Shin, G. Diaz, E. Asperges, A. Castagna, T. Feldt, G. Green, M. L. Green, F.-X. Lescure, et al., Compassionate use of remdesivir for patients with severe covid-19, New England Journal of Medicine (2020).

[9] S. G. Krantz, A. S. S. Rao, Level of under-reporting including under-diagnosis before the first peak of covid-19 in various countries: Preliminary retrospective results based on wavelets and deterministic modeling, Infection Control & Hospital Epidemiology (2020) 1–8.

[10] R. Niehus, P. Martinez de Salazar Munoz, A. Taylor, M. Lipsitch, Quantifying bias of covid-19 prevalence and severity estimates in wuhan, china that depend on reported cases in international travelers, medRxiv (2020).

[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. Oakley, et al., Presymptomatic sars-cov-2 infections and transmission in a skilled nursing facility, New England Journal of Medicine (2020).

[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 suspected or confirmed cases of covid-19 during home care with isolation: A new retrospective analysis based on o2o, Available at SSRN 3548746 (2020).

[13] R. H. Byrd, J. C. Gilbert, J. Nocedal, A trust region method based on interior point techniques for nonlinear programming, Mathematical programming 89 (2000) 149–185.

[14] J. C. Lagarias, J. A. Reeds, M. H. Wright, P. E. Wright, Convergence properties of the nelder–mead simplex method in low dimensions, SIAM Journal on optimization 9 (1998) 112–147.

[15] J. Lourenço, R. Paton, M. Ghafari, M. Kraemer, C. Thompson, P. Simmonds, P. Klenerman, S. Gupta, Fundamental principles of epidemic spread highlight the immediate need for large-scale serological surveys to assess the stage of the sars-cov-2 epidemic, medRxiv (2020).

[16] A. Doi, K. Iwata, H. Kuroda, T. Hasuike, S. Nasu, A. Kanda, T. Nagao, H. Nishioka, K. Tomii, T. Morimoto, et al., Estimation of seroprevalence of novel coronavirus disease (covid-19) using preserved serum at an outpatient setting in kobe, japan: A cross-sectional study., medRxiv (2020).

[17] H. Streeck, B. Schulte, B. Kuemmerer, E. Richter, T. Höller, C. Fuhrmann, E. Bartok, R. Dolscheid, M. Berger, L. Wessendorf, et al., Infection fatality rate of sars-cov-2 infection in a german community with a super-spreading event, medRxiv (2020).

[18] N. Sood, P. Simon, P. Ebner, D. Eichner, J. Reynolds, E. Bendavid, J. Bhattacharya, Seroprevalence of sars-cov-2–specific antibodies among adults in los angeles county, california, on april 10-11, 2020, JAMA (2020).

[19] E. Bendavid, B. Mulaney, N. Sood, S. Shah, E. Ling, R. Bromley-Dulfano, C. Lai, Z. Weissberg, R. Saavedra, J. Tedrow, et al., Covid-19 antibody seroprevalence in santa clara county, california, MedRxiv (2020).

[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. Brooks, J. K. Boldsen, et al., Estimation of sars-cov-2 infection fatality rate by real-time antibody screening of blood donors, medRxiv (2020).

[21] J. Wise, Covid-19: Surveys indicate low infection level in community, 2020.

[22] T. Breton, An estimate of unidentified and total us coronavirus cases by state on april 21, 2020, SSRN (2020).

[23] L. Peng, W. Yang, D. Zhang, C. Zhuge, L. Hong, Epidemic analysis of covid-19 in china by dynamical modeling, arXiv preprint arXiv:2002.06563 (2020).

[24] T. Hale, S. Webster, A. Petherick, T. Phillips, B. Kira, Oxford covid-19 government response tracker, blavatnik school of government, Oxford University. Creative Commons Attribution CC BY standard. Available at: https://www. bsg. ox. ac. uk/covidtracker. Accessed on: April 14 (2020) 2020.

[25] C. Murray, et al., Forecasting the impact of the first wave of the covid-19 pandemic on hospital demand and deaths for the usa and european economic area countries, medRxiv (2020).

[26] S. Zhang, M. Diao, W. Yu, L. Pei, Z. Lin, D. Chen, Estimation of the reproductive number of novel coronavirus (covid-19) and the probable outbreak size on the diamond princess cruise ship: A data-driven analysis, International Journal of Infectious Diseases 93 (2020) 201–204.

[27] Y. Liu, A. A. Gayle, A. Wilder-Smith, J. Rocklöv, The reproductive number of covid-19 is higher compared to sars coronavirus, Journal of travel medicine (2020).