Estimating the Cumulative Incidence of COVID-19 in the United States Using Four Complementary Approaches

Fred S. Lu[*,1,2] Andre T. Nguyen[*,1,3,4] Nicholas B. Link[*,1] Jessica T. Davis[7]
Matteo Chinazzi[7] Xinyue Xiong[7] Alessandro Vespignani[7] Marc Lipsitch[5]
Mauricio Santillana[1,5,6,†]

[1]Computational Health Informatics Program, Boston Children's Hospital, Boston, MA
[2]Department of Statistics, Stanford University, Stanford, CA
[3]University of Maryland, Baltimore County, Baltimore, MD
[4]Booz Allen Hamilton, Columbia, MD
[5]Department of Epidemiology, Harvard T.H. Chan School of Public Health
[6]Department of Pediatrics, Harvard Medical School, Boston, MA
[7]Laboratory for the Modeling of Biological and Socio-technical Systems, Northeastern University, Boston, MA USA
[*]These authors contributed equally to this study.
[†]Correspondance to: Mauricio Santillana (msantill@fas.harvard.edu)

August 6, 2020

Abstract

Effectively designing and evaluating public health responses to the ongoing COVID-19 pandemic requires accurate estimation of the prevalence of COVID-19 across the United States (US). Equipment shortages and varying testing capabilities have however hindered the usefulness of the official reported positive COVID-19 case counts. We introduce four complementary approaches to estimate the cumulative incidence of symptomatic COVID-19 in each state in the US as well as Puerto Rico and the District of Columbia, using a combination of excess influenza-like illness reports, COVID-19 test statistics, COVID-19 mortality reports, and a spatially structured epidemic model. Instead of relying on the estimate from a single data source or method that may be biased, we provide multiple estimates, each relying on different assumptions and data sources. Across our four approaches emerges the consistent conclusion that on April 4, 2020, the estimated case count was 5 to 50 times higher than the official positive test counts across the different states. Nationally, our estimates of COVID-19 symptomatic cases as of April 4 have a likely range of 2.2 to 4.9 million, with possibly as many as 8.1 million cases, up to 26 times greater than the cumulative confirmed cases of about 311,000. Extending our method to May 16, 2020, we estimate that cumulative symptomatic incidence ranges from 6.0 to 10.3 million, as opposed to 1.5 million positive test counts. The proposed combination of

approaches may prove useful in assessing the burden of COVID-19 during resurgences in the US and other countries with comparable surveillance systems.

1 Introduction

COVID-19 (SARS-CoV-2), is a coronavirus that was first identified in Hubei, China, in December of 2019. On March 11, due to its extensive spread, the World Health Organization (WHO) declared it a pandemic [1]. As of July 24, 2020, COVID-19 had infected people in nearly every country globally with an official case count surpassing 15 million cases worldwide and 4 million in the United States (US) [2]. It is however accepted that the official case count is capturing only a fraction of the actual infections, and reliable estimates of COVID-19 infections are critical for appropriate resource allocation, effective public health responses, and improved forecasting of disease burden [3].

A lack of widespread testing due to equipment shortages, varying levels of testing by region over time, and uncertainty around test sensitivity make estimating the point prevalence of COVID-19 difficult [4, 5]. In addition, it has been estimated that 18% [6] to 50% [7, 8] of people infected with COVID-19 are asymptomatic or paucisymptomatic. Even in symptomatic infections, under-reporting can further complicate the accurate characterization of the COVID-19 burden. For example, one study estimated that in China, 86% of cases had not been captured by lab-confirmed tests [9], and it is possible that this percentage is even higher in the US [5]. Finally, it has been suggested that the available information on confirmed COVID-19 cases across geographies may be an indicator of the local testing capacity over time, as opposed to an indicator of the epidemic trajectory. Thus, solely relying on positive test counts to infer the COVID-19 epidemic trajectory may not be sensible [10].

The aim of this study is to show how alternative methodologies, each with different sets of inputs and assumptions, can provide a consensus estimate of weekly cumulative symptomatic incidence of COVID-19 in each state in the US. One such approach is to analyze region-specific changes in the number of individuals seeking medical attention with influenza-like illness (ILI), defined as having a fever in addition to a cough or sore throat. The significant overlap in symptoms common to both ILI and COVID-19 suggests that leveraging existing disease monitoring systems, such as ILINet, a sentinel system created and maintained by the United States Centers of Disease Control and Prevention (CDC) [11, 12], may offer a way to estimate the ILI-symptomatic incidence of COVID-19 without needing to rely on COVID-19 testing results. Importantly, regional increases in ILI observed from February to April 2020 in conjunction with stable or decreasing influenza case numbers present a discrepancy (i.e., an increase in ILI not explained by an increase in influenza) that can be used to impute COVID-19 ILI-symptomatic cases.

A second and related approach uses ILI data to deconfound COVID-19 testing results from state-level testing capabilities. These two approaches show that existing ILI surveillance systems are a useful signal for measuring COVID-19 ILI-symptomatic incidence in the US, especially during the early stages of the outbreak. However, they are dependent on reporting from the ILINet system, and thus become less reliable outside of peak flu season and when COVID-19 precautions disrupt routine health care use.

Our third approach uses reported COVID-19-attributed deaths to estimate COVID-19 symptomatic incidence (broader than the ILI-symptomatic incidence of the first two methods) and improves upon previously introduced methodologies [13, 14, 15, 16, 17]. COVID-19 deaths may represent a lower-noise estimate of cases than surveillance testing given that patients who have died are sicker, more likely to be hospitalized, and thus more likely to be tested than the general infected population.

The fourth approach is based on the use of the Global Epidemic and Mobility model (*GLEAM*), a fully stochastic epidemic modeling platform that uses real-world data to perform in silico simulations of the spatial spread of COVID-19 in the US [18]. The mechanistic modeling stage explores the parameter space defined by the basic reproduction number, generation time, seasonality scaling factor, social distancing policies, and generates a corpus of simulated epidemic profiles. The simulation results can be aggregated at the level of each US state and the entire country. The model selection stage is performed by measuring the information loss with respect to the ground truth surveillance data of the weekly death incidence in each state.

While previous work has attempted to quantify COVID-19 incidence in the United States using discrepancies in ILI trends [19, 20], to the best of our knowledge this study is the first to offer a range of estimates at the state level, leveraging a suite of complementary methods based on different assumptions. We believe that this provides a more balanced picture of the uncertainty over COVID-19 (ILI-)symptomatic incidence in each state. While our results are approximations and depend on a variety of (likely time-dependent) estimated factors, we believe that our presented case counts better represent (ILI-)symptomatic incidence than simply relying on laboratory-confirmed COVID-19 tests. Providing such estimates for each state enables the design and implementation of more effective and efficient public health measures to mitigate the effects of the ongoing COVID-19 epidemic outbreak. While the scope of this paper is focused on the United States, the methods introduced here are general enough that they may prove useful to estimate COVID-19 burden in other locations with comparable disease (and death) monitoring systems.

2 Results

We implement five methods based on four approaches to estimate the symptomatic incidence of COVID-19 within the US from March 1 to May 16, 2020. These dates correspond to the early stages of the outbreak (with fewer than 50 confirmed cases in the US), up to the date of the the CDC reports as of May 28th, 2020. The first two methods, labeled *div-IDEA* and *div-Vir*, fall under the *Divergence* approach, which first estimates what the level of ILI activity across the US would have been if the COVID-19 outbreak had not occurred. Each method develops a control time series and uses the unexpected increase in the ILI rate over the control to infer the burden of COVID-19. *div-IDEA* is based on a parametric epidemiological model, the IDEA model [21], fitted to the observed 2019-2020 ILI (prior to the introduction of COVID-19 to the US), while *div-Vir* is based on the time-evolution of empirical observations of positive virological influenza test statistics. The third method, using the *COVID Scaling* approach, leverages healthcare ILI visits and COVID-19 test statistics to directly infer the proportion of ILI due to COVID-19 in the full population. These three methods estimate ILI symptomatic incidence and may miss symptomatic patients not matching the ILI symptoms (for the remainder of the paper, we use 'ILI-symptomatic' to denote COVID-19 patients with ILI symptoms and 'symptomatic' to denote COVID-19 patients with any symptoms). In addition, these methods are accurate only while ILI surveillance systems are operating normally (usually only during the flu season) and only while the outbreak has not yet overwhelmed hospitals. We use the ILI based methods to estimate ILI-symptomatic case counts until April 4th, 2020.

The fourth method, using the *mortality MAP* (*mMAP*) approach, uses the time series of reported COVID-19-attributed deaths in combination with the observed epidemiological characteristics of COVID-19 in hospitalized individuals to infer the latent disease onset time series. This is then scaled up to yield estimates of symptomatic case counts using previously reported fatality ratios and asymptomatic rates. The Methods section provides extensive details on the assumptions and data sources for each of these approaches. Finally we use a fifth method based on the explicit modeling of the epidemic using the *GLEAM* model, calibrated on reported deaths. The model provides the number of individuals that have been infected, the number of individuals that are currently infectious, and the number of daily new infections in US states and at the national level. *GLEAM* estimates the cumulative number of both symptomatic and asymptomatic infections, so it is scaled down by 40%, the current best point estimate for the number of infections that are asymptomatic [22, 23], to produce estimates of symptomatic cases.

2.1 Adjusted Assumptions Represent Most Likely Scenarios

Each method from the first three approaches has an adjusted version, which represents our best guess taking into account all information available to us, and an unadjusted version, which uses

pre-COVID-19 baseline information. Specifically, the adjusted divergences (*div-IDEA* and *div-Vir*) and *COVID Scaling* methods incorporate an increased probability that an individual with ILI symptoms will seek medical attention after the start of the COVID-19 outbreak based on recent survey data [24, 25]. The adjusted *mMAP* incorporates newer information from serological testing, indicating a lower fatality ratio (and thus higher estimated symptomatic case count) than expected. In addition, it supplements the confirmed COVID-19 deaths with unusual increases in pneumonia-related deaths across the country that may represent untested COVID-19 cases. Since there is no unadjusted version for *GLEAM* and because its infection fatality ratio is more similar to the adjusted *mMAP*, we group *GLEAM* in with the adjusted methods. In most states, as seen in Fig. 1, the adjusted estimates from each method are more closely clustered than their unadjusted counterparts, increasing our confidence in the adjusted range estimates of COVID-19 cumulative symptomatic incidence (ILI-symptomatic specifically for the ILI based methods).

2.2 Estimated Case Counts Far Surpass Reported Positive Cases

We first computed estimates for the national and state levels (including the District of Columbia and Puerto Rico) using these four approaches for the time period between March 1, 2020 and April 4, 2020. The adjusted methods estimate that there had been 2.2 to 4.9 million symptomatic or ILI-symptomatic COVID-19 cases in the US; including unadjusted estimates raises the upper limit to 8.1 million cases. In comparison, around 311,000 positive cases had been officially recorded during that time period. Fig. 1(a) displays the COVID-19 symptomatic case count estimates from our methods (ILI-symptomatic in particular for the ILI based methods) at the national and state levels compared with the reported case numbers. The results suggest that the estimated true numbers of infected cases are nearly uniformly much higher than those reported. Next, we extended our methods to produce estimates through May 16, 2020 using recent data, displayed in Fig. 1(b). Because of a strong decline in ILINet statistics due to the end of the flu season and unusually low numbers of reporting providers, our *Divergence* and *COVID Scaling* approaches report few or no cases after April 4, 2020. Therefore, our recent estimates are computed using the *mMAP* method and the *GLEAM* model, which estimate between 6.0 and 10.3 million symptomatic cases had occurred as of May 16. In contrast, 1.5 million positive test counts had been reported. This highlights that models using only confirmed test cases may significantly underestimate the actual COVID-19 cumulative incidence in the United States, which is consistent with what previous studies have shown [9, 20].

As a naive baseline, if one only adjusts the number of reported cases by the (likely) percentage of asymptomatic cases (18% [6] to 50% [7, 8]) and symptomatic cases not seeking medical attention (up to 73% [26]), one would conclude that the actual number of cases were about four to eight times the number of reported cases; this ratio would also be constant across states. In contrast, our

methods frequently estimate 5-fold to 50-fold more symptomatic (for *mMAP*) or ILI-symptomatic (for *Divergence* and *COVID Scaling*) cases than those reported and show significant state-level variability (see Fig 2). The median estimates for the ratios of estimated cases to reported cases from March 1 to April 4, 2020 for the adjusted *div-IDEA* method is 18 (with a 90% interval from 1 to 101), for adjusted *div-Vir* is 21 (2, 67), for adjusted *COVID-Scaling* is 17 (3, 76), for adjusted *mMAP* is 11 (4, 20), and for *GLEAM* is 10 (2, 29).

Using our methods, we also compute cumulative case estimates for each week within the studied period. Fig. 3 highlights the rapid increase in estimated COVID-19 cases over the United States as well as in New York, Washington, and Louisiana, three locations which experienced early outbreaks. These methods suggest that states under-reported COVID-19 case counts even early in March, likely due to limited testing availability. In New York and Louisiana, the estimates were more similar across methods than in Washington. Since Washington had already experienced an outbreak by February 28 [27], testing shortages may have been more pronounced than in the other states. Our divergence analysis approach does not rely on any COVID-19 test-dependent data (including deaths) and therefore may provide more accurate estimates in Washington.

2.3 State-level Comparisons

Over the period of March 1, 2020 to April 4, 2020, the adjusted *div-IDEA*, *div-Vir*, *COVID Scaling*, *mMAP*, and *GLEAM* approaches estimated that between 24 and 35 (24, 24, 35, 33, and 25, respectively) locations had actual (ILI-)symptomatic case counts above 10 times the reported counts (Figs. 1 and 2). Up to 12 locations had at least one adjusted estimate above 50 times the reported counts, with three of them above 100 times the reported counts (Nebraska, Oregon, Missouri). Places with low official case counts, such as Alaska and North Dakota, may have experienced significantly more COVID-19 cases than reported. Even places with high official case counts, such as Georgia, Pennsylvania, and Texas, appeared to be significantly under-reporting. As expected, our methods computed high estimates in New York and New Jersey, locations with especially high numbers of confirmed cases. Over the period leading up to May 16, 2020, *mMAP* and *GLEAM* estimates indicate that up to 30 locations have estimated case counts above five times the reported counts, with two locations over 10 times (Connecticut and Michigan).

Using the unadjusted methods, the ILI-based methods yield significantly higher estimates than *mMAP* (median estimates of 84k, 155k, 62k, 11k for *div-IDEA*, *div-Vir*, *COVID Scaling*, and *mMAP*, respectively, for the locations that have estimates for all methods). However, the adjusted versions of the methods (including *GLEAM*) are more similar (median estimates of 35k, 73k, 85k, 23k, and 17k for *div-IDEA*, *div-Vir*, *COVID Scaling*, *mMAP*, and *GLEAM*), providing support that the adjusted methods are more accurate than the unadjusted ones.

All five methods generally agree on the ordering of states by (ILI-)symptomatic case count (Table 1), with rank correlations of the adjusted methods ranging from 0.47 to 0.97. *mMAP* and *GLEAM* have 0.97 and 0.91 correlations with the reported case counts, which is likely because official COVID-19 deaths and positive COVID-19 cases represent overlapping pools of patients and are therefore subject to similar biases. *COVID Scaling* also shows a relatively high correlation with the reported cases, 0.88, which may reflect the use of COVID-19 test statistics in its model. *div-IDEA* and *div-Vir*, however, solely rely on aggregate data from ILINet, which may cover a different set of patients.

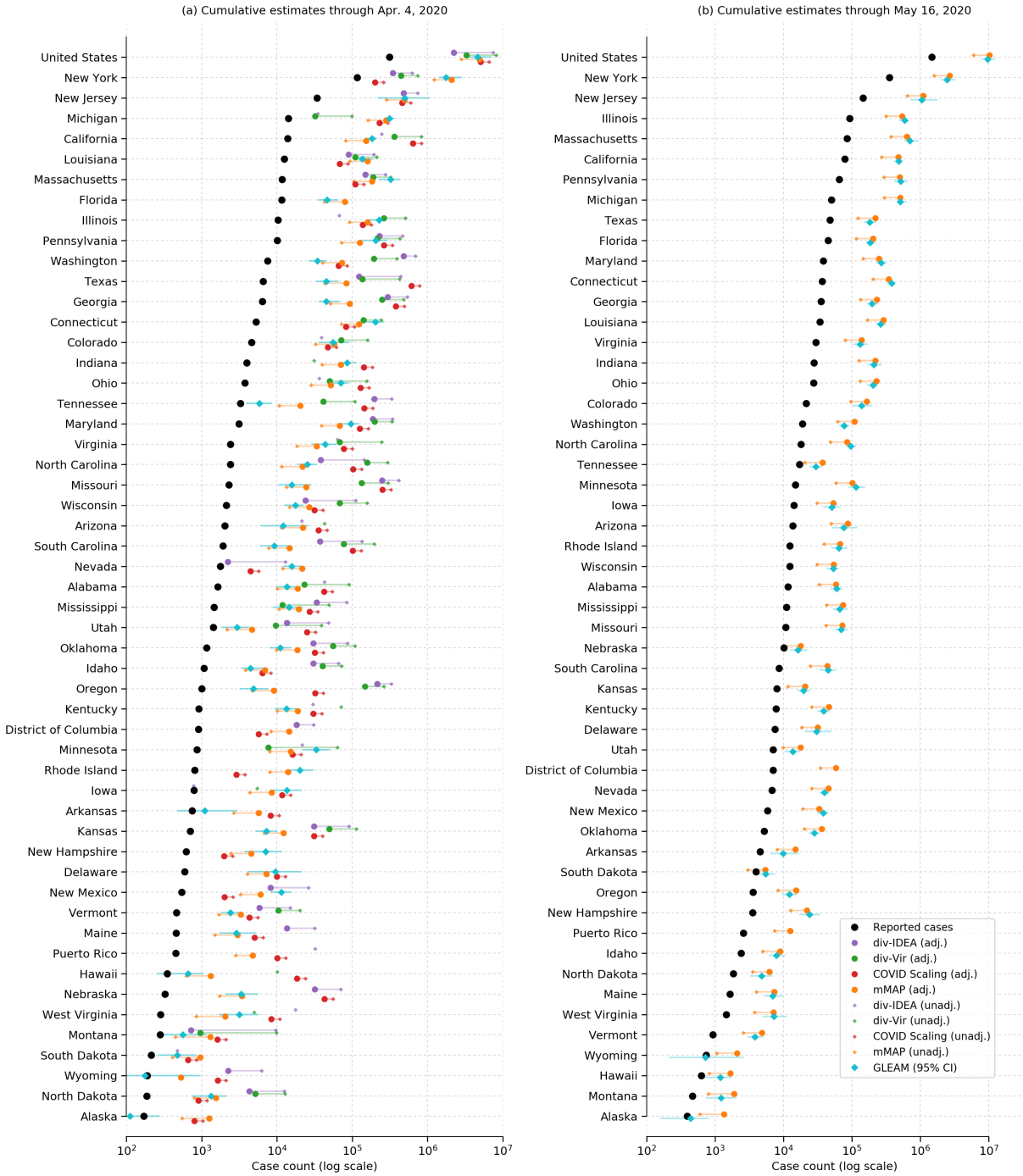


Figure 1: COVID-19 (ILI-)symptomatic case count estimates compared with reported case counts at the national and state levels from March 1, 2020 to (a) April 4, 2020 and (b) May 16, 2020. Cases are presented on a log scale. Adjusted methods take into account increased visit propensity (*div-IDEA*, *div-Vir*, *COVID Scaling*) and pneumonia-recorded deaths along with a lower estimated case fatality rate (*mMAP*). In places where the ILI-based methods show no divergence in observed and predicted ILI visits, the estimates of COVID-19 cannot be calculated and are not shown. Note that Florida does not provide ILI data, so only *mMAP* could be estimated there.

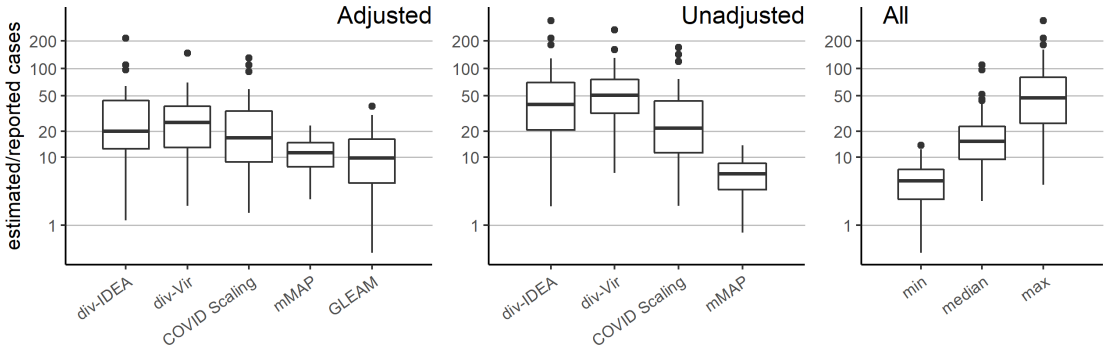


Figure 2: Distribution of the state-level ratios of estimated to reported case counts from March 1, 2020 to April 4, 2020. The right-hand plot shows the results of using all methods together: taking the *min*, *median*, and *max* of the state-level estimates across methods.

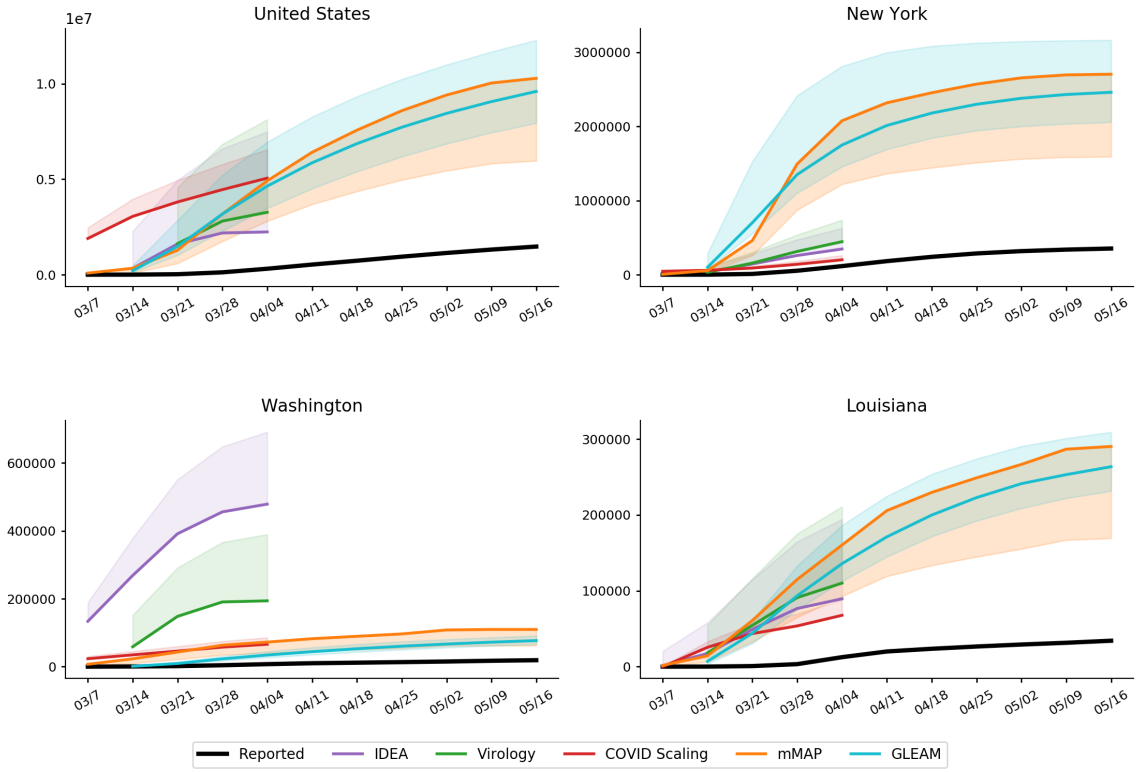


Figure 3: Cumulative weekly case counts from March 1 to May 16, 2020 for the United States, New York, Washington, and Louisiana, as estimated by each method and the reported cases. The estimate for each week indicates total cases up to the denoted date. Solid and dotted lines indicate the adjusted and unadjusted estimates, respectively. Refer to the Supplementary Materials for results over all locations.

	Reported	div-IDEA	div-Virology	COVID Scaling	mMAP	GLEAM
Reported	1.00	0.58	0.71	0.88	0.97	0.91
div-IDEA	–	1.00	0.74	0.64	0.58	0.47
div-Virology	–	–	1.00	0.69	0.70	0.66
COVID Scaling	–	–	–	1.00	0.86	0.78
mMAP	–	–	–	–	1.00	0.96
GLEAM	–	–	–	–	–	1.00

Table 1: Pairwise Spearman correlations between adjusted methods and reported case counts from March 1, 2020 to April 4, 2020 across the state level.

3 Discussion

We present five methods based on four distinct approaches to estimate the COVID-19 cumulative symptomatic incidence across the United States. The methods are complementary, in that they rely on different methods, assumptions and use diverse datasets. Despite their clear differences, these methods estimate that the likely COVID-19 cumulative symptomatic incidence varies from 5 to 50 times higher, at the state level, than what has been reported so far in the U.S. By providing ranges of estimates, both within and across models, our suite of methods offers a robust picture of the under-ascertainment of state-level COVID-19 case counts. When making public health decisions to respond to COVID-19, it is important to account for the uncertainty in estimates of symptomatic incidence; the multiple estimates presented here provide a consistent picture of the number of infected individuals.

Our estimates are specifically for symptomatic cases, while a high proportion of COVID-19 cases are believed to be asymptomatic [6, 7, 23]. To estimate total cases, our counts can be adjusted by the proportion of symptomatic cases. For example, if 40% of cases are asymptomatic, this could indicate a total cumulative incidence of up to 17.1 million as of May 16, 2020.

Our approaches could be expanded to include other data sources and methods to estimate incidence, such as Google searches [28, 29, 30], electronic health record data [31], clinician's searches [32], and/or mobile health data [33]. Accurate and appropriately sampled serological testing would provide the most accurate estimate of incidence and would be useful for public health measures, especially when attempting to relax or re-institute shelter-in-place recommendations. In addition, serological testing could be used to evaluate the reliability of the methods presented in this study. This could inform prevalence estimation methods for COVID-19 in other countries as well as for future pandemics. The ILI-based methods presented in this study demonstrate the potential of existing and well-established ILI surveillance systems to monitor future pandemics that, like COVID-19, present similar symptoms to ILI. This is especially promising given the WHO

initiative launched in 2019 to expand influenza surveillance globally [34]. Incorporating estimates from influenza and COVID-19 forecasting and participatory surveillance systems may prove useful in future studies as well [18, 35, 36, 37, 38, 39].

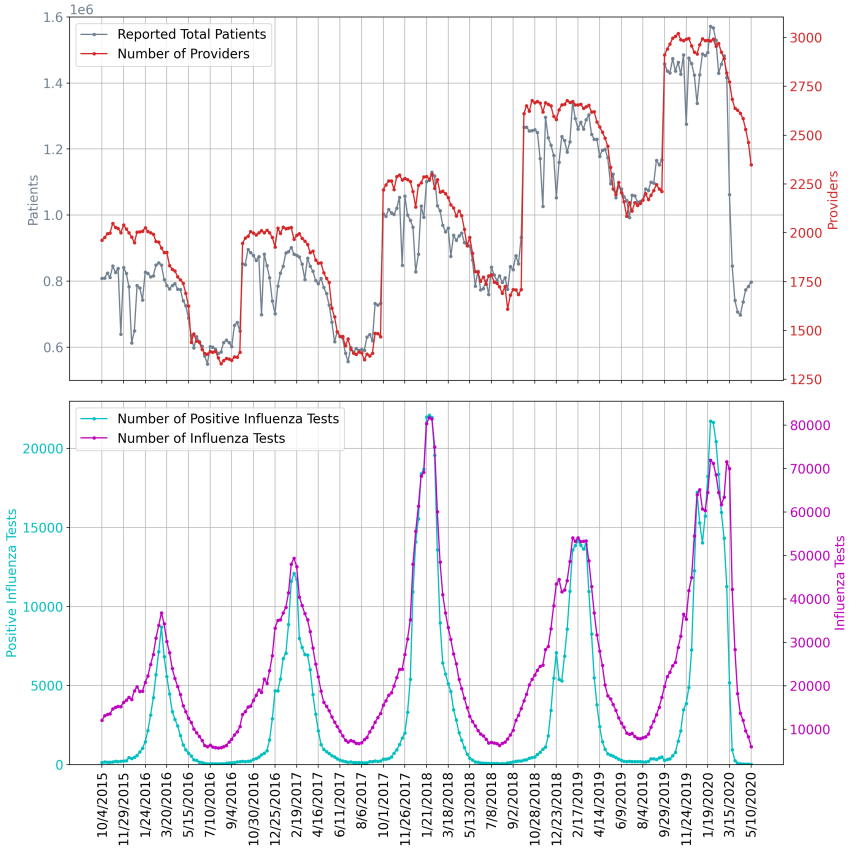
Limitations. Since the *Divergence* and *COVID Scaling* approaches are estimated using ILINet statistics, their symptomatic incidence estimates are dependent on the ILI definition of a fever and cough or sore throat. Thus, they may miss a percentage of COVID-19 patients that are symptomatic without meeting the ILI definition. With this limitation, the reported estimates may serve as an approximate lower bound. Given a clearer understanding of COVID-19 symptoms, our *Divergence* and *COVID Scaling* estimates could be adjusted upward by the proportion of symptomatic to ILI-symptomatic patients.

The uncertainty and bias of each individual method should be considered carefully. The *Divergence* methods suffer from the same challenges faced when attempting to scale CDC-measured ILI activity to the entire population [40]. In particular, scaling to case counts in a population requires estimates for $p(\text{visit})$, the probability that a person seeks medical attention for any reason, and $p(\text{visit} \mid \text{ILI})$ which captures health care seeking behavior given that a person is experiencing ILI; these estimates are likely to change over time, especially during the course of a pandemic. Moreover, the weekly symptomatic incidence estimates from this method decrease towards the end of March, perhaps caused by a change in health care seeking behavior after the declaration of a national emergency on March 13, 2020 and the widespread implementation of shelter-in-place mitigation strategies. It is also important to note that ILI based methods are expected to be accurate only while ILI surveillance systems are operating normally (reporting tends to decrease outside of the flu season) and only while the outbreak has not yet overwhelmed hospitals and doctors. Figure 4 shows the underlying influenza surveillance data for the last five seasons. We note a sharp decrease in the total number of reported patients in late March 2020 even though the number of providers did not decrease more than is usually expected. This suggests that the ILINet signal may no longer be reliable until regular reporting patterns return. As a result, we only use ILI based methods to estimate COVID-19 symptomatic incidence early in the outbreak.

COVID Scaling relies on the assumption that COVID-19 positive test proportions uniformly represent the pool of all ILI patients and that shortages in testing do not bias the positive proportion upward or downward. *mMAP* is limited by assumptions of the the distribution of time from case onset to death. Furthermore, *mMAP* and *GLEAM* rely on assumptions about the fatality rate (symptomatic case fatality rate for *mMAP* and infection fatality rate for *GLEAM*) and accurate reporting of COVID-19 deaths (or in the case of adjusted *mMAP*, that excess pneumonia deaths capture all unreported COVID-19 deaths). It is possible that many deaths caused by COVID-19 are not being officially counted as COVID-19 deaths because of a lack of testing (and

that accounting for increased pneumonia deaths does not fully capture this) [41]. In New York City, for example, probable COVID-19 deaths (as in, not needing a test result) are being reported as COVID-19 deaths and accounted for a 42% increase in cumulative COVID-19 death counts as of April 29, 2020 [42], indicating that other locations not counting probable deaths could be missing a significant portion of deaths. Under-reporting of deaths may explain why *mMAP* and *GLEAM* sometimes yield lower case estimates than *Divergence* and *COVID Scaling* even though its symptomatic case definition is more inclusive. A high-level summary of the three methods, their estimation strategy, and their assumptions are provided in Table 5.5.

Figure 4: The underlying influenza surveillance data for the last five seasons. The top subplot shows the ILINet total number of patients and participating providers. The bottom subplot shows the total reported numbers of influenza tests conducted and positive influenza tests.



4 Conclusions

We have presented four complementary approaches for estimating the true COVID-19 cumulative (ILI-)symptomatic incidence in the United States from March 1 to May 16, 2020 at the national

and state levels. The approaches rely on different datasets and modeling assumptions in order to balance the inherent biases of each individual method. While the case count estimates from these methods vary, there is general agreement among them that the actual state-level symptomatic case counts up to April 4, 2020 were likely 5 to 50 times greater than what was reported. Up to May 16, 2020, most states likely had 5 to 10 times more cases than reported, with a total estimated range of 6.0 million to 10.3 million cases over the United States.

A more accurate picture of the burden of COVID-19 is actionable knowledge that will help guide and focus public health responses. As social distancing measures are being (or have been) relaxed, some locations are experiencing a resurgence in cases. If the true case counts are near the upper bound of our estimated symptomatic case count, then a fair proportion (up to 4% as of May 16) of the US population may have already been infected. Factoring in asymptomatic cases this could increase the proportion up to 8%. On the other hand, it is evident that the large majority of the population has not yet been exposed to COVID-19, and therefore effective, informed public health responses to future upsurges in cases will be essential in the upcoming months.

5 Data and Methods

CDC ILI and Virology: The CDC US Outpatient Influenza-like Illness Surveillance Network (ILINet) monitors the level of ILI circulating in the US at any given time by gathering information from physicians' reports about patients seeking medical attention for ILI symptoms. ILI is defined as having a fever (temperature of 37.8+ Celsius) and a cough or a sore throat. ILINet provides public health officials with an estimate of ILI activity in the population but has a known availability delay of 7 to 14 days. National level ILI activity is obtained by combining state-specific data weighted by state population [12]. Additionally, the CDC reports information from the WHO and the National Respiratory and Enteric Virus Surveillance System (NREVSS) on laboratory test results for influenza types A and B. The data is available from the CDC FluView dashboard [11]. We omit Florida from our analysis as ILINet data is not available for Florida.

COVID-19 Case and Death Counts: The US case and death counts are taken from the New York Times repository, which compiles daily reports of counts at the state and county levels across the US [43]. For the *mMAP* validation in the supplementary materials, the case and death counts from other countries are taken from the John's Hopkins University COVID-19 dashboard [44]. Counts are taken up until May 28, 2020.

COVID-19 Testing Counts: In addition, daily time series containing positive and negative COVID-19 test results within each state were obtained from the COVID Tracking Project [45].

US Demographic Data: The age-stratified, state-level population numbers are taken from 2018 estimates from the US census [46].

5.1 Approach 1: Divergence

Viewing COVID-19 as an intervention, this approach aims to construct control time series representing the counterfactual 2019-2020 influenza season without the effect of COVID-19. While inspired by the synthetic control literature [47, 48], we are forced to construct our own controls since COVID-19 has had an effect in every state. We formulate a control as having the following two properties:

1. The control produces a reliable estimate of ILI activity, where ILI refers to the symptomatic definition of having a fever in addition to a cough or sore throat.
2. The control is not affected by the COVID-19 intervention (that is, the model of ILI conditional on any relevant predictors is independent of COVID-19).

We construct two such controls, one model-based and one data-driven.

5.1.1 Method 1: Incidence Decay and Exponential Adjustment Model

The Incidence Decay and Exponential Adjustment (IDEA) model [21] is a single equation epidemiological model that estimates disease incidence over time early in an outbreak while accounting for control activities and behaviours. The model is as follows:

$$I(t) = \left(\frac{R_0}{(1+d)^t} \right)^t$$

where $I(t)$ is the incident case count at serial interval time step t. R_0 is the basic reproduction number, and d is a discount factor modeling reductions in the effective reproduction number with time due to public health interventions, changes in public behavior, environmental factors, and the depletion of susceptible hosts. The IDEA model has been shown to be identical to Farr's law for epidemic forecasting and can be expressed in terms of a susceptible-infectious-removed (SIR) compartmental model with improving control [49].

We fit the IDEA model to ILI case counts from the start of the 2019-2020 influenza season to the last week of February 2020. The start of the 2019-2020 influenza season is defined in a location specific manner as the first occurrence of two consecutive weeks with ILI activity above 2%. Model fitting is done using non-linear least squares with the Trust Region Reflective algorithm as the optimizer. Next, the model is used to predict what ILI would have been had the COVID-19 pandemic not occurred. In other words, we use the IDEA model ILI estimates as the counterfactual

when assessing the impact of the COVID-19 intervention. When fitting the IDEA model, we use a serial interval of half a week, consistent with the serial interval estimates from [50] for influenza. We note that serial interval estimates from [51] for COVID-19 as well as from [52] for SARS-CoV-1 are longer than that of influenza, but that is not an issue as we use IDEA to model ILI.

5.1.2 Method 2: Virology

As an alternative control to the IDEA model, we also present an estimator of ILI activity using influenza virology results. As suggested by [19], there has been a divergence in March between CDC measured ILI activity and the fraction of ILI specimens that are influenza positive. Clinical virology time series were obtained from the CDC virologic surveillance system consisting of over 300 laboratories participating in virologic surveillance for influenza through either the US WHO Collaborating Laboratories System or NREVSS [12]. Total number of tests, total influenza positive tests, and percent positive tests are our variables of interest.

None of the three time series satisfy both properties of a valid control, as defined in 5.1, since total number of tests is directly susceptible to increase when ILI caused by COVID-19 is added. Similarly, percent positive flu tests may decrease when COVID-19 is present. On the other hand, total positive flu tests satisfies property 2, but is not a reliable indicator of ILI activity (property 1) on its own because it is highly dependent on the quantity of tests administered.

We propose a modification that satisfies the properties. Let F_t^+, N_t, I_t denote positive flu tests, total specimens tested, and ILI visit counts respectively. In addition, let F_t be the true underlying flu counts. For any week t we assume the following relation:

$$F_t = \frac{F_t^+ \cdot I_t}{N_t}$$

There are two interpretations of this quantity: 1) It extrapolates the positive test percentage (F_t^+/N_t) to all ILI patients (I_t), a quantity known in the mechanistic modeling literature as ILI+ [53]. 2) Test frequency is a confounder in the relationship between the number of positive tests (F^+) and total flu (F_t). Adjusting for test frequency closes the indirect pathway between F_t and F_t^+ [54]. In the Supplementary Material, we demonstrate over a series of examples that this estimator behaves as desired. Each estimate of F_t is then scaled to population ILI cases using least squares regression over pre-COVID-19 ILI counts.

In other words, we first use virology data to estimate F_t (actual flu cases causing ILI) as percent ILI visits times percent positive for flu. Then, modeling ILI visits (I_t) as an affine function of F_t in a normal (without COVID-19) situation, we use 2019 pre-COVID-19 data to map F_t to I_t. This allows us to estimate the divergence after the COVID-19 intervention occurs.

5.1.3 ILI Case Count Estimation

In order to fit the IDEA and virology models, we estimate the ILI case count in the population from the CDC's reported percent ILI activity, which measures the fraction of medical visits that were ILI related.

In a similar fashion to the approach of [40], we can use Bayes' rule to map percent ILI activity to an estimate of the actual population-wide ILI case count. Let $p(\text{ILI})$ be the probability of any person having an influenza-like illness during a given week, $p(\text{ILI} \mid \text{visit})$ be the probability that a person seeking medical attention has an influenza-like illness, $p(\text{visit})$ be the probability that a person seeks medical attention for any reason, and $p(\text{visit} \mid \text{ILI})$ the probability that a person with an influenza-like illness seeks medical attention. Bayes' rule gives us

$$p(\text{ILI}) = \frac{p(\text{visit})}{p(\text{visit} \mid \text{ILI})} \cdot p(\text{ILI} \mid \text{visit})$$

$p(\text{ILI} \mid \text{visit})$ is the CDC's reported percent ILI activity, for $p(\text{visit})$ we use the estimate from [40] of a weekly doctor visitation rate of 7.8% of the US population, and for $p(\text{visit} \mid \text{ILI})$ we use a base estimate of 27%, consistent with the findings from [26]. Once $p(\text{ILI})$ is calculated, we multiply $p(\text{ILI})$ by the population size to get a case count estimate within the population.

5.1.4 Visit Propensity Adjustment

We note that health care seeking behavior varies by region of the United States as shown in [26]. To better model these regional behavior differences, we adjust $p(\text{visit} \mid \text{ILI})$, the probability that a person with an influenza-like illness seeks medical attention, using regional baselines for the 2019-2020 influenza season [12].

Additionally, because our method estimates the increase in ILI visits due to the impact of COVID-19, we must distinguish an increase due to COVID-19 cases from an underlying increase in medical visit propensity in people with ILI symptoms. Due to the widespread alarm over the spread of COVID-19, it would not be unreasonable to expect a potential increase in ILI medical visits even in the hypothetical absence of true COVID-19 cases.

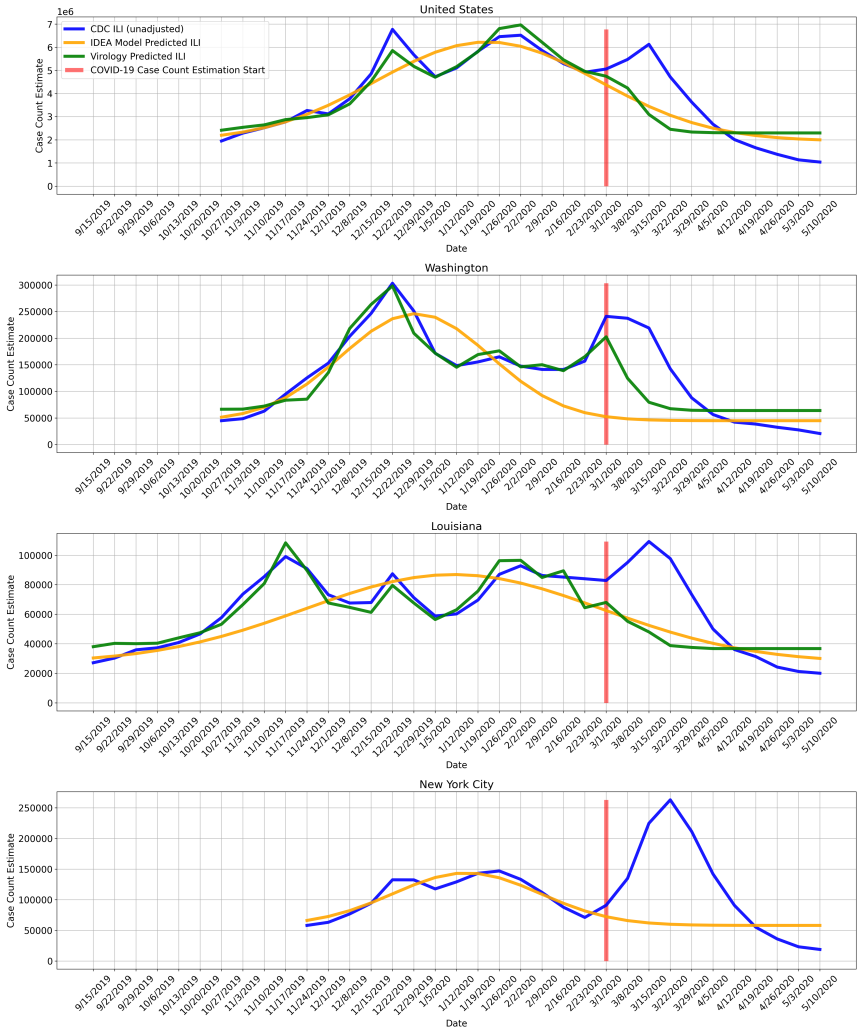
For this reason, we also explore increasing $p(\text{visit} \mid \text{ILI})$ from 27% to 35% to measure the possible effect of a change in health care seeking behavior due to COVID-19 media attention and panic. The increase of $p(\text{visit} \mid \text{ILI})$ to 35% is consistent with health care seeking behavior surveys done after the start of COVID-19 [24, 25]. The *Divergence* and *COVID Scaling* methods have *adjusted* versions which incorporate this shift as well as *unadjusted* versions that keep the baseline 27% propensity.

5.1.5 Estimating COVID-19 Case Counts

The ultimate goal is to estimate the true burden of COVID-19. The IDEA and virology predicted ILI case counts can be used to estimate CDC ILI had COVID-19 not occurred. In other words, the IDEA and virology predicted ILI can be used as counterfactuals when measuring the impact of COVID-19 on CDC measured ILI. The difference between the observed CDC measured ILI and the counterfactual (IDEA predicted ILI or virology predicted ILI) for a given week is then the estimate of COVID-19 ILI-symptomatic case counts for that week. Fig. 5 shows example observed CDC measured ILI, IDEA model predicted ILI, and virology predicted ILI. The supplementary materials contain similar plots to Fig. 5 for all locations. For this method as well as the following two, we start estimating COVID-19 case counts the week starting on March 1, 2020. We note that while the IDEA and virology ILI predictions tend to track CDC ILI well earlier in the flu season, after COVID-19 started to impact the United States there is a clear divergence between predictions and observed CDC ILI, with CDC ILI increasing while the counterfactual estimates decrease.

This method is expected to be accurate only while ILI surveillance systems are operating normally (reporting tends to decrease outside of the flu season) and only while the outbreak has not yet overwhelmed hospitals and doctors. As a result, we use ILI based methods to estimate COVID-19 symptomatic incidence only early in the outbreak, until April 4th. The disappearance of the divergence does not mean that the outbreak is over, but rather that the ILI signal is no longer reliable.

Figure 5: COVID-19 is treated as an intervention, and we measure COVID-19 impact on observed CDC ILI, using IDEA model predicted ILI and virology predicted ILI as counterfactuals. The difference between the higher observed CDC ILI and the lower IDEA model predicted ILI and virology predicted ILI is the measured impact of COVID-19. The impact directly maps to an estimate of COVID-19 ILI-symptomatic case counts. Virology predicted ILI is omitted when virology data is not available. We note that this method is meaningful only at the beginning of the outbreak (March 2020), while ILI surveillance systems are still fully operational and before they are impacted by COVID-19. The disappearance of the divergence does not mean that the outbreak is over, but rather that the ILI signal is no longer reliable.



5.2 Approach 2: COVID Scaling

This approach infers the COVID-19 fraction of the total ILI by extrapolating testing results obtained from the COVID Tracking Project [45], following the same reasoning as the Virology Divergence method. That is,

$$C_t = \frac{C_t^+ \cdot I_t}{N_t^c}$$

where C_t^+, N_t^c, I_t denote positive COVID-19 tests, total COVID-19 specimens, and ILI visit counts respectively.

State-level testing results were aggregated to the weekly level and positive test percentages were computed using the positive and negative counts, disregarding pending tests. Positive test counts were adjusted for potential false negatives. There are varying estimates for the false negative rate for the RT-PCR used in COVID-19 tests, with some reports suggesting rates as high as 25-30% [55, 56]. We apply a 15% false negative rate in our analysis; repeating our analysis using a range of values from 5% to 25% yielded little difference in our estimates. On the other hand, COVID-19 testing is highly specific, so we assume no false positives. Then, the number of false negatives (FN) can be computed from the recorded (true) positives (TP) and the false negative rate (fnr) as

$$FN = TP \cdot \frac{fnr}{1 - fnr}$$

Because COVID-19 testing is sparse in many states, there are issues with zero or low sample sizes, as well as testing backlogs. Rather than taking the empirical positive test percentage (C_t^+/N_t^c), we first smoothed the test statistics over time by aggregating results over a 2-week sliding window. This has a Bayesian interpretation of combining each week's observed statistics with the prior of the previous week, weighted by relative specimen count. For convenience, C_t^+ and N_t^c henceforth refer to these respective quantities. We also applied the same process to the ILI information to reduce noise and so that the data are comparable. This helped but did not address all issues with case backlog, so we further smoothed the COVID-19 estimates using a Bayesian spatial model:

Denote p_{jt} as the probability that a given ILI patient in state j and week t has COVID-19. Under the condition that testing is applied uniformly, the COVID-19 status of patient i from state j in week t is

$$X_{jt}^{(i)} \sim Bernoulli(p_{jt})$$

Assuming COVID-19 status is independent in each ILI patient, conditional on the state prevalence, the state testing results follow a Binomial distribution. We apply a spatial prior based on first-order

conditional dependence:

$$p_{jt} \sim Beta(\alpha_{jt} N_{0t}, (1-\alpha_{jt}) N_{0t})$$
$$\alpha_{jt} = \frac{1}{|\mathcal{N}_j|} \sum_{k \in \mathcal{N}_j} p_{kt}$$

where $\mathcal{N}_j$ are the neighbors of state j. The strength of the prior was specified by setting N_{0t} to be the number of total tests at the 5th quantile among all states in each week. Finally, we compute α_{jt} by replacing each p_{kt} by their empirical estimates. Using the Beta-Binomial conjugacy we derive closed-form posterior mean estimates for p_{jt}:

$$\hat{p}_{jt} = \frac{C^{+}_{jt} + \alpha_{jt} N_{0t}}{N^{c}_{jt} + N_{0t}}$$

As described previously, the weekly, state-level reported percent ILI were then multiplied by $\hat{p}_{jt}$ to get an estimate of the percent of medical visits that could be attributed to COVID-19. These values were subsequently scaled to the whole population using the same Bayes' rule method as described in *ILI Case Count Estimation* (4.2.3).

5.3 Approach 3: Mapping Mortality to COVID-19 Cases

Other studies have introduced methods to infer COVID-19 cases from COVID-19 deaths using (semi-)mechanistic disease models [15] or statistical curve-fitting based on assumptions of epidemic progression [16], but, to the best of our knowledge, no methods have been proposed to directly infer COVID-19 cases without either of these assumptions.

Mortality Map (*mMAP*) uses, under a Bayesian framework, reported deaths to predict previous true case counts, similar to prior work on influenza [17]. *mMAP* accounts for right-censoring (i.e. COVID-19 cases that are not resolved yet) by adapting previously used methods [13]. A study of clinical cases in Wuhan found that the time in days from symptom onset to death roughly follows a log-normal distribution with mean 20.2 and standard deviation 11.6 [57]. It also found the mean time from hospitalization to death to be 13.2 days, similar to the estimate of 13.7 from a large cohort study in California [58], suggesting that the timing of disease progression is similar in the United States. Using this distribution, a time series of reported deaths, D, and the age-adjusted symptomatic case fatality rate (sCFR), we estimate the distribution of symptomatic cases C, defined at the usual time of symptom onset, using an iterative Bayesian approach. We use Bayes' rule to define the probability that there was a case on day t given a death on day τ

$$p(\text{case on } t \mid \text{death on } \tau) = \frac{p(\text{death on } \tau \mid \text{case on } t) \cdot p(\text{case on } t)}{p(\text{death on } \tau)} \tag{1}$$

Let C_{d^*} denote the predicted distribution of when D are classified as cases (i.e. are hospitalized), C_d denote the predicted distribution of when D and future deaths are classified as cases (so

adjusted for right-censoring), and t_{max} denote the most recent date with deaths reported. Let $p(\text{death on } \tau \mid \text{case on } t) = p(T = (\tau - t))$ denote the log-normal probability. *mMAP* performs the following steps:

1. Initialize the prior probability of a case on day t, $p_0(\text{case on } t)$, as uniform.

2. Repeat the following for each iteration i:

   - Calculate $C_{d^*}^{(i)}$.

$$\begin{aligned} C_{d^*}^{(i)}(t) &= \sum_{\tau=t+1}^{t_{max}} D(\tau) \cdot p_{i-1}(\text{case on } t \mid \text{death on } \tau) \\ &= \sum_{\tau=t+1}^{t_{max}} D(\tau) \cdot \frac{p(T = (\tau - t)) \cdot p_{i-1}(\text{case on } t)}{\sum_{s=1}^{\tau-1} p(T = (\tau - s)) \cdot p_{i-1}(\text{case on } s)} \end{aligned} \tag{2}$$

   where the denominator is equivalent to $p(\text{death on } \tau)$ in (1).

   - We estimate that the proportion $p(T \leq (t_{max} - t))$ of $C_d^{(i)}(t)$ have died by t_{max} and use this to adjust for right censoring.

$$C_d^{(i)}(t) = \frac{C_{d^*}^{(i)}(t)}{p(T \leq (t_{max} - t))} \tag{3}$$

   - Update prior probabilities

$$p_i(\text{case on } t) = \frac{C_d^{(i)}(t)}{\sum C_d^{(i)}(t)} \tag{4}$$

   - Repeat until $\frac{||C_d^{(i)} - C_d^{(i-1)}||}{||C_d^{(i)}||} \leq \epsilon$, where ϵ is a pre-specified tolerance level.

3. $C_d(t)$ represents the number of cases on day t that will lead to death. We scale this to estimate the number of all symptomatic cases by dividing by the sCFR.

$$C(t) = \frac{C_d(t)}{sCFR} \tag{5}$$

Interestingly, the update step for $C_d^{(i)}(t)$ in each iteration is the same as the Richardson-Lucy deconvolution step proposed for influenza [17] and the expectation-maximization step to find the maximum likelihood estimate for emission tomography (without right-censoring in the latter) [59], albeit with different notation in each study. Supplementary section 4.3 demonstrates that *mMAP* successfully predicts cases in simulated and true scenarios using data from six countries. As well, supplementary section 4.1 demonstrates that if *mMAP* converges, which it does for every US state, C_d fully explains deaths under the assumed probability distribution (6), and supplementary section 4.2 demonstrates that this satisfies the calculation of the fatality rate as presented in [13].

$$D(t) = \sum_{\tau=1}^{t-1} p(T = t - \tau) \cdot C_d(\tau) \ , \ \forall \ t \ \in 1..t_{max} \tag{6}$$

Alternatively, if one were interested in estimating the incidence of all cases – symptomatic and asymptomatic – $C_d(t)$ would need to be divided by the infection fatality ratio (IFR) in step 3 (equation 5). For the sake of comparison with the ILI-based methods in this study, we chose to use sCFR in the denominator in equation 5 to estimate the incidence of just the symptomatic cases. The national sCFR values used are 1.8% and 1.1% for the adjusted method. These values were found by adjusting the IFR estimates (1.1% and 0.65%) with an assumed 40% asymptomatic rate (estimates of the percentage of asymptomatic cases range from 18% [6] to 50% [7, 8] and the CDC puts 40% as the best point estimate of this number [22, 23]). The first IFR value comes from an analysis of individual case data in China and repatriated Chinese citizens in January and February to estimate the fatality ratio for all — symptomatic and asymptomatic — infections [60]. The second value comes from a meta-analysis of published IFR values and is the CDC best point estimate of the IFR [22, 23]. The sCFR estimates for each state are adjusted using the age-stratified fatality rate [61] and the population age structure provided by the US census [46].

5.3.1 Accounting for Unreported COVID-19 Deaths

While *mMAP* assumes all COVID-19 deaths are reported, some deaths will be unreported because of limited testing and false negative results [62, 63]. Previous research on the H1N1 epidemic estimated that the ratio of lab-confirmed deaths to actual deaths caused by the disease was 1:7 nationally [64] and 1:15 globally [65]. While the actual rate of under-reporting is unknown, we include an adjustment, mMAP^{adj}, that uses an estimate of unreported COVID-19 deaths based on reports of excess pneumonia deaths, as has been done in previous studies [62]. mMAP^{adj} assumes that excess pneumonia deaths in March 2020 were due to COVID-19.

The CDC reports weekly pneumonia deaths, $D_P(w)$, expected weekly pneumonia deaths based on a model of historical trends, $\mathbb{E}[D_P(w)]$, and deaths that are classified as pneumonia and COVID-19, $D_{P \cap COV}(w)$ [66, 67]. We estimate that the number of un-classified COVID-19 deaths each week, $D_U(w)$, is the following:

$$D_U(w) = D_P(w) - \mathbb{E}[D_P(w)] - D_{P \cap COV(w)} \tag{7}$$

This results in 355, 438, 605, and 540 nationwide excess deaths for the four weeks from March 1 to March 28, which is the most recent data at the time of writing this paper. To account for missing data in recent weeks, We assume that the weekly number of excess deaths remains constant after March 21, i.e. that there were 540 excess deaths the weeks of March 29 - April 4 and April 5 - April 11. Since the expected pneumonia deaths are not available at the state level, excess

deaths are calculated nationally and then attributed to each state s in proportion to the number of pneumonia deaths in that state. Then, the weekly excess deaths are evenly distributed across each day of the week.

$$\begin{aligned} D_U^s(w) &= D_U(w) \cdot \frac{D_P^s(w)}{\sum_{s \in S} D_P^s(w)} \\ D^{s,adj}(t) &= D^s(t) + \frac{1}{7} D_U^s(w), \text{ where } t \in w \end{aligned} \tag{8}$$

5.4 Approach 4: Global Epidemic and Mobility Model

The Global Epidemic and Mobility model is an individual-based, stochastic, and spatial epidemic model. *GLEAM* uses real-world data to perform *in silico* simulations of the spatial spread of infectious diseases at the global level. In the model, the world is divided into over 3,200 geographic subpopulations constructed using a Voronoi tessellation of the Earth's surface. Subpopulations are centered around major transportation hubs and integrate data on the population such as age specific contact patterns [68], short-range (i.e. commuting) and long-range (i.e. flights) mobility data from the Offices of Statistics for 30 countries on 5 continents as well as the Official Aviation Guide (OAG) and IATA databases (updated in 2019) [69, 70]. The model has been used extensively to analyze previous epidemic such as the H1N1 2009 pandemic and the Zika epidemic in the Americas [71, 72, 73], and to simulate the early spreading of COVID-19 in mainland China [18].

We use the model to analyze the spatiotemporal spread and magnitude of the COVID-19 epidemic in the continental US. For COVID-19 the model adopts a classic $SLIR$ disease characterization in which individuals can be classified into four compartments: susceptible, latent, infectious, or removed. Susceptible individuals become latent through interactions with infectious individuals. During both the latent and infectious stages we assume that individuals are able to travel. Following the infectious period, individuals then progress into the removed compartment where they are no longer able to infect others, meaning they have either recovered, been hospitalized, isolated, or have died. *GLEAM* is able to simulate explicitly the disease dynamic at the individual level.

Approximate Bayesian Computation is used to estimate the posterior distribution of the basic parameters of the model. The calibration of the global model for COVID-19 is reported in [18]. Within the US, we have implemented domestic airline traffic reductions and local commuting pattern reductions. The magnitude of these reductions is based on the analysis of data from millions of (anonymized, aggregated, privacy-enhanced) devices [74] and official airline data from OAG. We consider two major social distancing periods in the US. The first period includes mitigation policies widely adopted on March 16, 2020 [20], including system-wide school closures, work from home policies (smart work), and reduction in casual social interactions in the community. The second period refers to the issuing in more than 41 states of "stay at home" or "shelter in place" orders

starting on April 1, 2020. The impact of these mitigation policies is reflected in specific contact patterns calculated in the model's synthetic populations on the different layers where individuals interact: households, schools, workplaces, and in the general community. We also consider in each state the progression into reopening phases after April 30th, 2020. We assume varying levels of effectiveness of the mitigation policies and generate an ensemble of models that provides the weekly number number of new deaths by using estimates of COVID-19 severity from available data [23, 75, 60]. We then perform model selection based on the information loss with respect to the reported weekly deaths. The selected models allow us to compute the median and 95% CI for cumulative infections in each state (Fig. 6). The estimated total number of infections can be adjusted to provide an estimate of COVID-19 symptomatic cases by reducing the predictions by an estimated asymptomatic rate of 40% [22, 23]. In Fig. 6, we report the model estimates of the cumulative number of infections on May 16, 2020 compared to the number of cases reported through that date within each state. We see a strong correlation between the reported cases and our model's estimated number of infections, (Pearson's correlation coefficient on log-values 0.98, $p < 0.001$). If we assume that the number of reported cases and simulated infections are related through a simple binomial stochastic sampling process, we find that the median ascertainment rate of detecting an infected individual by May 16, 2020 is 11.2% (95%CI: [6.4%, 40.5%]). The detailed model's results are publicly available at https://covid19.gleamproject.org/.

5.5 Aggregation of Estimates

The divergence-based methods predict national COVID-19 symptomatic incidence directly using national ILI data. *mMAP* and *GLEAM* predict national symptomatic incidence using national death data, while *COVID Scaling* estimates national symptomatic incidence by aggregating the case estimates from each state.

The *Divergence* and *COVID Scaling* methods provide separate case estimates for each week within the studied period, which are summed to the total cumulative case estimates. *mMAP* and *GLEAM* provide daily estimates which are further aggregated by week.

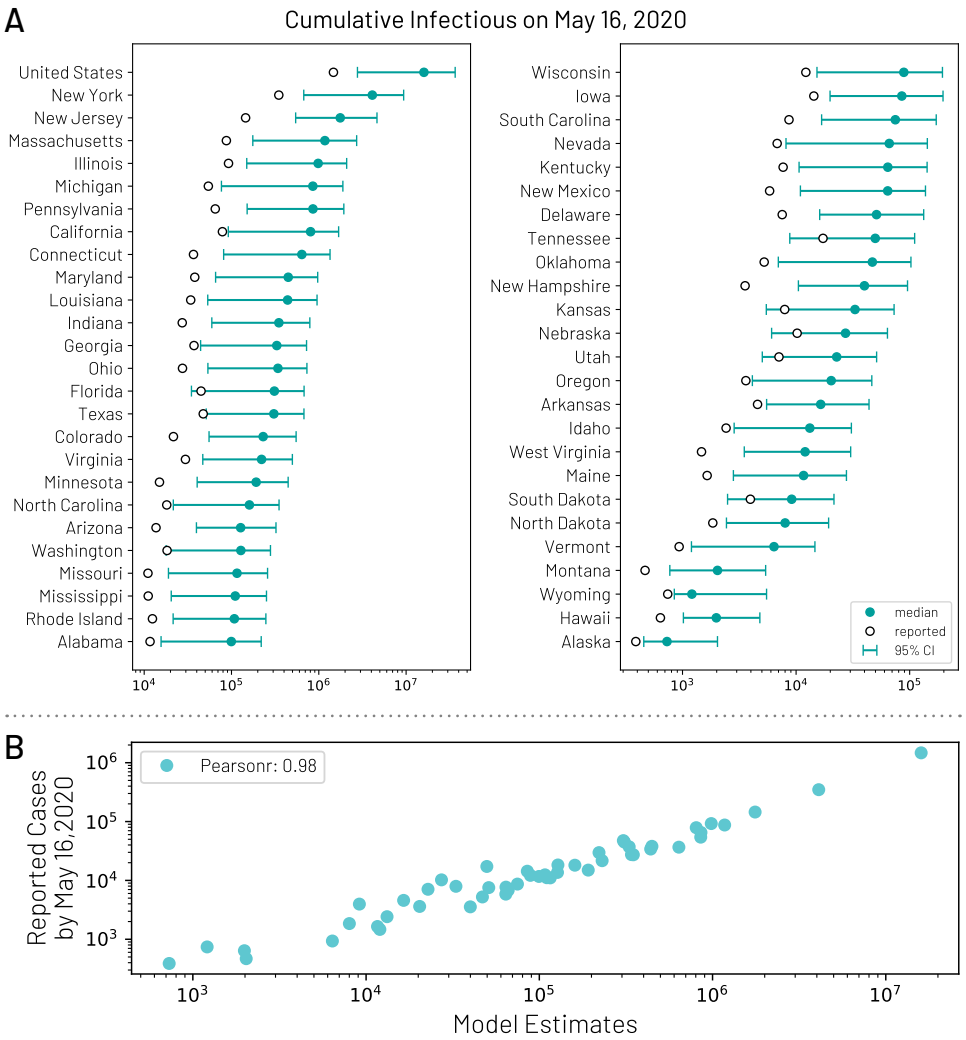


Figure 6: (A) Model estimates of the cumulative number of infections using the *GLEAM* model by May 16, 2020 for each state. (B) Correlation between the number of reported cases of COVID-19 for each state and the model estimates of the total number of infections by May 16, 2020.

Approach	*Divergence*	*COVID Scaling*	*mMAP*	*GLEAM*
Brief Description	Treat COVID-19 ILI-symptomatic case count estimation as a causal inference problem. COVID's impact on ILI activity is measured using as controls an Incidence Decay with Exponential Adjustment model as well as influenza testing statistics.	Extrapolate state-level positive test percentages for COVID-19 to the weekly ILI data to estimate COVID-19 proportion in medical visits, then scale to the whole population.	Using reported COVID-19 deaths, the sCFR, and a distribution of time from cases to deaths, predict the latent case distribution.	Estimates cumulative infections using a stochastic spatially structured epidemic model, calibrated on weekly incident deaths.
Data Input	ILI activity and influenza test results.	ILI activity and COVID-19 test results.	COVID-19 deaths.	COVID-19 deaths.
Model Assumptions	1. The divergence between predicted ILI activity for the 2019-2020 season and measured ILI activity after the start of the COVID-19 pandemic can be attributed to COVID-19. 2. Scaling from ILI to population is reliable.	1. COVID-19 test reports accurately represent the pool of weekly ILI visits. 2. Delayed test reporting does not significantly affect positive test proportions after applying smoothing 3. Scaling from ILI to population is reliable.	1. All COVID-19 deaths are reported (mMAP) or explained by excess pneumonia deaths (mMAP^{adj}). 2. The distribution of time from cases to death is log-normal. 3. The age-stratified IFR is the same as reported in [61].	1. Modeling estimates for the effect of school closures, smart working, and social distancing effects on the transmissibility of SARS-CoV-2. 2. Spatial variation of the IFR are not considered. 3. Differential transmissibility across age brackets is not considered.
Expected Bias	This method can be sensitive to model fit and changes in healthcare seeking behavior among symptomatic individuals. This method will work only while ILI surveillance is reliable.	ILI visits and COVID-19 tests may capture different segments of the sick population.	May underestimate cases as many COVID-19 related deaths may go unreported or untested.	Revision to the current estimate of the IFR affects the model estimated of the total number of infections/cases.

Table 1: (On previous page) Comparing the four approaches to estimate COVID-19 cases in the US.

References

[1] World Health Organization. Report of the who-china joint mission on coronavirus disease 2019.

[2] Centers for Disease Control and Prevention. Locations with confirmed covid-19 cases.

[3] Marc Lipsitch and Mauricio Santillana. Enhancing situational awareness to prevent infectious disease outbreaks from becoming catastrophic. *Global Catastrophic Biological Risks*, pages 59–74, 2019.

[4] Sheila Kaplan Sheri Fink Katie Thomas Michael D. Shear, Abby Goodnough and Noah Weiland. The lost month: How a failure to test blinded the u.s. to covid-19.

[5] Arjun K. Manrai and Kenneth D. Mandl. Covid-19 testing: overcoming challenges in the next phase of the epidemic.

[6] Kenji Mizumoto, Katsushi Kagaya, Alexander Zarebski, and Gerardo Chowell. Estimating the asymptomatic proportion of coronavirus disease 2019 (covid-19) cases on board the diamond princess cruise ship, yokohama, japan, 2020. *Eurosurveillance*, 25(10), 2020.

[7] Tara John. Iceland lab's testing suggests 50% of coronavirus cases have no symptoms. *CNN*, Apr 2020.

[8] Michael Day. Covid-19: identifying and isolating asymptomatic people helped eliminate virus in italian village, 2020.

[9] Ruiyun Li, Sen Pei, Bin Chen, Yimeng Song, Tao Zhang, Wan Yang, and Jeffrey Shaman. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (sars-cov2). *Science*, 2020.

[10] Justin Kaashoek and Mauricio Santillana. Covid-19 positive cases, evidence on the time evolution of the epidemic or an indicator of local testing capabilities? a case study in the united states. *Available at SSRN: https://ssrn.com/abstract=3574849, April*, 2020.

[11] Centers for Disease Control and Prevention. Fluview.

[12] Centers for Disease Control and Prevention. U.s. influenza surveillance system: Purpose and methods.

[13] Hiroshi Nishiura, Don Klinkenberg, Mick Roberts, and Johan AP Heesterbeek. Early epidemiological assessment of the virulence of emerging infectious diseases: a case study of an influenza pandemic. *PLoS One*, 4(8), 2009.

[14] TW Russell, J Hellewell, S Abbott, CI Jarvis, K van Zandvoort, et al. Using a delay-adjusted case fatality ratio to estimate under-reporting. *Centre for Mathematical Modeling of Infectious Diseases Repository*, 2020.

[15] Seth Flaxman, Swapnil Mishra, Axel Gandy, H Juliette T Unwin, Helen Coupland, et al. Report 13: Estimating the number of infections and the impact of non-pharmaceutical interventions on covid-19 in 11 european countries. 2020.

[16] IHME COVID, Christopher JL Murray, et al. Forecasting covid-19 impact on hospital bed-days, icu-days, ventilator-days and deaths by us state in the next 4 months. *medRxiv*, 2020.

[17] Edward Goldstein, Jonathan Dushoff, Junling Ma, Joshua B Plotkin, David JD Earn, and Marc Lipsitch. Reconstructing influenza incidence by deconvolution of daily mortality time series. *Proceedings of the National Academy of Sciences*, 106(51):21825–21829, 2009.

[18] Matteo Chinazzi, Jessica T Davis, Marco Ajelli, Corrado Gioannini, Maria Litvinova, Stefano Merler, Ana Pastore y Piontti, Kunpeng Mu, Luca Rossi, and Kaiyuan Sun. The effect of travel restrictions on the spread of the 2019 novel coronavirus (covid-19) outbreak. *Science*, 2020.

[19] Graham C. Gibson Estee Cramer Caitlin M. Rivers Nicholas G. Reich, Evan L. Ray. Looking for evidence of a high burden of covid-19 in the united states from influenza-like illness data.

[20] Justin D Silverman and Alex D Washburne. Using ili surveillance to estimate state-specific case detection rates and forecast sars-cov-2 spread in the united states. *medRxiv*, 2020.

[21] David N Fisman, Tanya S Hauck, Ashleigh R Tuite, and Amy L Greer. An idea for short term outbreak projection: nearcasting using the basic reproduction number. *PloS one*, 8(12), 2013.

[22] Gideon Meyerowitz-Katz and Lea Merone. A systematic review and meta-analysis of published research data on covid-19 infection-fatality rates. *medRxiv*, 2020.

[23] Centers for Disease Control and Prevention. Covid-19 pandemic planning scenarios.

[24] Jonathan Rothwell. Estimating covid-19 prevalence in symptomatic americans.

[25] Pascal Geldsetzer. Knowledge and perceptions of covid-19 among the general public in the united states and the united kingdom: A cross-sectional online survey.

[26] Kristin Baltrusaitis, Alessandro Vespignani, Roni Rosenfeld, Josh Gray, Dorrie Raymond, and Mauricio Santillana. Differences in regional patterns of influenza activity across surveillance systems in the united states: Comparative evaluation. *JMIR Public Health and Surveillance*, 5(4):e13403, 2019.

[27] Temet M McMichael. Covid-19 in a long-term care facility—king county, washington, february 27–march 9, 2020. *MMWR. Morbidity and Mortality Weekly Report*, 69, 2020.

[28] Shihao Yang, Mauricio Santillana, and Samuel C Kou. Accurate estimation of influenza epidemics using google search data via argo. *Proceedings of the National Academy of Sciences*, 112(47):14473–14478, 2015.

[29] Fred S Lu, Mohammad W Hattab, Cesar Leonardo Clemente, Matthew Biggerstaff, and Mauricio Santillana. Improved state-level influenza nowcasting in the united states leveraging internet-based data and network approaches. *Nature communications*, 10(1):1–10, 2019.

[30] Shihao Yang, Samuel C Kou, Fred Lu, John S Brownstein, Nicholas Brooke, and Mauricio Santillana. Advances in using internet searches to track dengue. *PLoS computational biology*, 13(7), 2017.

[31] Mauricio Santillana, AT Nguyen, Tamara Louie, Anna Zink, Josh Gray, Iyue Sung, and John S Brownstein. Cloud-based electronic health records for real-time, region-specific influenza surveillance. *Scientific reports*, 6:25732, 2016.

[32] Mauricio Santillana, Elaine O Nsoesie, Sumiko R Mekaru, David Scales, and John S Brownstein. Using clinicians' search query data to monitor influenza epidemics. *Clinical Infectious Diseases*, 59(10):1446, 2014.

[33] Aaron C Miller, Inder Singh, Erin Koehler, and Philip M Polgreen. A smartphone-driven thermometer application for real-time population-and individual-level influenza surveillance. *Clinical Infectious Diseases*, 67(3):388–397, 2018.

[34] World Health Organization. Global influenza strategy 2019–2030.

[35] Shengjie Lai, Nick W Ruktanonchai, Liangcai Zhou, Olivia Prosper, Wei Luo, Jessica R Floyd, Amy Wesolowski, Chi Zhang, Xiangjun Du, Hongjie Yu, et al. Effect of non-pharmaceutical interventions for containing the covid-19 outbreak: an observational and modelling study. *medRxiv*, 2020.

[36] Mauricio Santillana, André T Nguyen, Mark Dredze, Michael J Paul, Elaine O Nsoesie, and John S Brownstein. Combining search, social media, and traditional data sources to improve influenza surveillance. *PLoS computational biology*, 11(10), 2015.

[37] Mark S Smolinski, Adam W Crawley, Kristin Baltrusaitis, Rumi Chunara, Jennifer M Olsen, Oktawia Wójcik, Mauricio Santillana, Andre Nguyen, and John S Brownstein. Flu near you: crowdsourced symptom reporting spanning 2 influenza seasons. *American journal of public health*, 105(10):2124–2130, 2015.

[38] John S Brownstein, Shuyu Chu, Achla Marathe, Madhav V Marathe, Andre T Nguyen, Daniela Paolotti, Nicola Perra, Daniela Perrotta, Mauricio Santillana, Samarth Swarup, et al. Combining participatory influenza surveillance with modeling and forecasting: Three alternative approaches. *JMIR public health and surveillance*, 3(4):e83, 2017.

[39] Dianbo Liu, Leonardo Clemente, Canelle Poirier, Xiyu Ding, Matteo Chinazzi, Jessica T Davis, Alessandro Vespignani, and Mauricio Santillana. A machine learning methodology for real-time forecasting of the 2019-2020 covid-19 outbreak using internet searches, news alerts, and estimates from mechanistic models. *arXiv preprint arXiv:2004.04019*, 2020.

[40] Wan Yang, Marc Lipsitch, and Jeffrey Shaman. Inference of seasonal and pandemic influenza transmission dynamics. *Proceedings of the National Academy of Sciences*, 112(9):2723–2728, 2015.

[41] Sarah Kliff and Julie Bosman. Official counts understate the u.s. coronavirus death toll. *The New York Times*, Apr 2020.

[42] NYC Health. Confirmed and probably covid-19 deaths daily report, Apr 2020. `https://www1.nyc.gov/assets/doh/downloads/pdf/imm/covid-19-deaths-confirmed-probable-daily-04302020.pdf`.

[43] New York Times. Data from the new york times, based on reports from state and local health agencies, 2020. `https://github.com/nytimes/covid-19-data`.

[44] Ensheng Dong, Hongru Du, and Lauren Gardner. An interactive web-based dashboard to track covid-19 in real time. *The Lancet infectious diseases*, 2020. Data obtained from `https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data/csse_covid_19_time_series`.

[45] The covid tracking project. `https://covidtracking.com/`.

[46] U.S. Census Bureau. Annual estimates of the civilian population by single year of age and sex for the united states and states: April 1, 2010 to july 1, 2018, 2019. `https://www.census.gov/data/tables/time-series/demo/popest/2010s-state-detail.html`.

[47] Kay H. Brodersen, Fabian Gallusser, Jim Koehler, Nicolas Remy, and Steven L. Scott. Inferring causal impact using bayesian structural time-series models. *Annals of Applied Statistics*, 9:247–274, 2015.

[48] Alberto Abadie, Alexis Diamond, and Jens Hainmueller. Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American statistical Association*, 105(490):493–505, 2010.

[49] Mauricio Santillana, Ashleigh Tuite, Tahmina Nasserie, Paul Fine, David Champredon, Leonid Chindelevitch, Jonathan Dushoff, and David Fisman. Relatedness of the incidence decay with exponential adjustment (idea) model,"farr's law" and sir compartmental difference equation models. *Infectious disease modelling*, 3:1–12, 2018.

[50] Margaretha Annelie Vink, Martinus Christoffel Jozef Bootsma, and Jacco Wallinga. Serial Intervals of Respiratory Infectious Diseases: A Systematic Review and Analysis. *American Journal of Epidemiology*, 180(9):865–875, 10 2014.

[51] Qun Li, Xuhua Guan, Peng Wu, Xiaoye Wang, Lei Zhou, Yeqing Tong, Ruiqi Ren, Kathy SM Leung, Eric HY Lau, Jessica Y Wong, et al. Early transmission dynamics in wuhan, china, of novel coronavirus–infected pneumonia. *New England Journal of Medicine*, 2020.

[52] Marc Lipsitch, Ted Cohen, Ben Cooper, James M Robins, Stefan Ma, Lyn James, Gowri Gopalakrishna, Suok Kai Chew, Chorh Chuan Tan, Matthew H Samore, et al. Transmission dynamics and control of severe acute respiratory syndrome. *Science*, 300(5627):1966–1970, 2003.

[53] Sen Pei, Sasikiran Kandula, Wan Yang, and Jeffrey Shaman. Forecasting the spatial transmission of influenza in the united states. *Proceedings of the National Academy of Sciences*, 115(11):2752–2757, 2018.

[54] Andrew Forney Carlos Cinelli and Judea Pearl. A crash course in good and bad control.

[55] Chunxia Qin, Fang Liu, Tzu-Chen Yen, and Xiaoli Lan. 18 f-fdg pet/ct findings of covid-19: a series of four highly suspected cases. *European Journal of Nuclear Medicine and Molecular Imaging*, pages 1–6, 2020.

[56] Yang Yang, Minghui Yang, Chenguang Shen, Fuxiang Wang, Jing Yuan, Jinxiu Li, Mingxia Zhang, Zhaoqin Wang, Li Xing, Jinli Wei, et al. Laboratory diagnosis and monitoring the viral shedding of 2019-ncov infections. *medRxiv*, 2020.

[57] Natalie M Linton, Tetsuro Kobayashi, Yichi Yang, Katsuma Hayashi, Andrei R Akhmetzhanov, Sung-mok Jung, Baoyin Yuan, Ryo Kinoshita, and Hiroshi Nishiura. Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: a statistical analysis of publicly available case data. *Journal of clinical medicine*, 9(2):538, 2020.

[58] Joseph A Lewnard, Vincent X Liu, Michael L Jackson, Mark A Schmidt, Britta L Jewell, Jean P Flores, Chris Jentz, Graham R Northrup, Ayesha Mahmud, Arthur L Reingold, et al. Incidence, clinical outcomes, and transmission dynamics of hospitalized 2019 coronavirus disease among 9,596,321 individuals residing in california and washington, united states: a prospective cohort study. *medRxiv*, 2020.

[59] Lawrence A Shepp and Yehuda Vardi. Maximum likelihood reconstruction for emission tomography. *IEEE transactions on medical imaging*, 1(2):113–122, 1982.

[60] Robert Verity, Lucy C Okell, Ilaria Dorigatti, Peter Winskill, Charles Whittaker, Natsuko Imai, Gina Cuomo-Dannenburg, Hayley Thompson, Patrick GT Walker, Han Fu, et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *The Lancet Infectious Diseases*, 2020.

[61] Neil Ferguson, Daniel Laydon, Gemma Nedjati Gilani, Natsuko Imai, Kylie Ainslie, Marc Baguelin, Sangeeta Bhatia, Adhiratha Boonyasiri, ZULMA Cucunuba Perez, Gina Cuomo-Dannenburg, et al. Report 9: Impact of non-pharmaceutical interventions (npis) to reduce covid19 mortality and healthcare demand. 2020.

[62] Daniel Weinberger, Ted Cohen, Forrest Crawford, Farzad Mostashari, Don Olson, Virginia E Pitzer, Nicholas G Reich, Marcus Russi, Lone Simonsen, Annie Watkins, and Cecile Viboud. Estimating the early death toll of covid-19 in the united states. *medRxiv*, 2020.

[63] Emma Brown, Beth Reinhard, and Aaron Davis. Coronavirus death toll: Americans are almost certainly dying of covid-19 but being left out of the official count. *the Washington Post*, Apr 2020.

[64] Vivek Charu, Lone Simonsen, Roger Lustig, Claudia Steiner, and Cécile Viboud. Mortality burden of the 2009-10 influenza pandemic in the united states: improving the timeliness of influenza severity estimates using inpatient mortality records. *Influenza and other respiratory viruses*, 7(5):863–871, 2013.

[65] Fatimah S Dawood, A Danielle Iuliano, Carrie Reed, Martin I Meltzer, David K Shay, Po-Yung Cheng, Don Bandaranayake, Robert F Breiman, W Abdullah Brooks, Philippe Buchy, et al. Estimated global mortality associated with the first 12 months of 2009 pandemic influenza a h1n1 virus circulation: a modelling study. *The Lancet infectious diseases*, 12(9):687–695, 2012.

[66] Centers for Disease Control and Prevention. Pneumonia and influenza mortality surveillance from the national center for health statistics mortality surveillance system. `https://gis.cdc.gov/grasp/fluview/mortality.html` accessed April 12, 2020.

[67] Centers for Disease Control and Prevention. Provisional death counts for coronavirus disease (covid-19). `https://www.cdc.gov/nchs/nvss/vsrr/covid19/index.htm` accessed April 12, 2020.

[68] Dina Mistry, Maria Litvinova, Matteo Chinazzi, Laura Fumanelli, Marcelo FC Gomes, Syed A Haque, Quan-Hui Liu, Kunpeng Mu, Xinyue Xiong, M Elizabeth Halloran, et al. Inferring high-resolution human mixing patterns for disease modeling. *arXiv preprint arXiv:2003.01214*, 2020.

[69] International Air Transportation Association https://www.iata.org/.

[70] Official Aviation Guide https://www.oag.com/.

[71] Duygu Balcan, Vittoria Colizza, Bruno Gonçalves, Hao Hu, José J Ramasco, and Alessandro Vespignani. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the National Academy of Sciences*, 106(51):21484–21489, 2009.

[72] Duygu Balcan, Bruno Gonçalves, Hao Hu, José J Ramasco, Vittoria Colizza, and Alessandro Vespignani. Modeling the spatial spread of infectious diseases: The GLobal Epidemic and Mobility computational model. *Journal of computational science*, 1(3):132–145, 2010.

[73] Qian Zhang, Kaiyuan Sun, Matteo Chinazzi, Ana Pastore y Piontti, Natalie E Dean, Diana Patricia Rojas, Stefano Merler, Dina Mistry, Piero Poletti, Luca Rossi, et al. Spread of Zika virus in the Americas. *Proceedings of the National Academy of Sciences*, 114(22):E4334–E4343, 2017.

[74] Brennan Klein, Tim LaRock, Stefan McCabe, and et al. Assessing changes in commuting and individual mobility in major metropolitan areas in the united states during covid-19 outbreak.

[75] Joseph T Wu, Kathy Leung, Mary Bushman, Nishant Kishore, Rene Niehus, Pablo M de Salazar, Benjamin J Cowling, Marc Lipsitch, and Gabriel M Leung. Estimating clinical severity of covid-19 from the transmission dynamics in wuhan, china. *Nature Medicine*, 26(4):506–510, 2020.

Contributions, Conflicts of Interest, and Funding.

FL, AN, and NL contributed equally to a literature search, figures, study design, data collection, data analysis, data interpretation, and writing. MS contributed to the literature search, study design, data interpretation, and writing. ML contributed to the literature search, data interpretation, and writing.

No conflicts of interest.

MS is partially supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number R01GM130668. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.