CORONAVIRUS

Using influenza surveillance networks to estimate state-specific prevalence of SARS-CoV-2 in the United States

Justin D. Silverman,[1,2]*† Nathaniel Hupert,[3,4] Alex D. Washburne[5]*†

[1]College of Information Science and Technology, Penn State University, University Park, Pennsylvania 16802, USA. [2]Department of Medicine, Penn State University, Hershey, Pennsylvania 17033, USA. [3]Weill Cornell Medicine, Cornell University, New York, New York 10065, USA. [4]New York-Presbyterian Hospital, New York, New York 10065, USA. [5]Department of Microbiology and Immunology, Montana State University, Bozeman, Montana 59717, USA.

*Denotes Equal Contribution

*Corresponding author. Email: JustinSilverman@psu.edu, alex.d.washburne@gmail.com

Detection of SARS-CoV-2 infections to date has relied heavily on RT-PCR testing. However, limited test availability, high false-negative rates, and the existence of asymptomatic or sub-clinical infections have resulted in an under-counting of the true prevalence of SARS-CoV-2. Here, we show how influenza-like illness (ILI) outpatient surveillance data can be used to estimate the prevalence of SARS-CoV-2. We found a surge of non-influenza ILI above the seasonal average in March 2020 and showed that this surge correlated with COVID-19 case counts across states. If 1/3 of patients infected with SARS-CoV-2 in the US sought care, this ILI surge would have corresponded to more than 8.7 million new SARS-CoV-2 infections across the US during the three-week period from March 8 to March 28, 2020. Combining excess ILI counts with the date of onset of community transmission in the US, we also show that the early epidemic in the US was unlikely to have been doubling slower than every 4 days. Together these results suggest a conceptual model for the COVID-19 epidemic in the US characterized by rapid spread across the US with over 80% infected patients remaining undetected. We emphasize the importance of testing these findings with seroprevalence data and discuss the broader potential to use syndromic surveillance for early detection and understanding of emerging infectious diseases.

INTRODUCTION

The ongoing severe acute respiratory syndrome–coronavirus 2 (SARS-CoV-2) pandemic continues to cause substantial morbidity and mortality around the world [1, 2]. Regional preparation for the pandemic requires estimating the growth rate of the epidemic, the timing of the epidemic peak, the demand for hospital resources, and the degree to which current policies may curtail the epidemic, all of which benefit from accurate estimates of the true prevalence of the virus within a population [3]. Confirmed cases are thought to be underestimates of true prevalence due to some unknown combination of patients not reporting for testing, testing not being conducted, and false-negative test results. Estimating the true prevalence of SARS-CoV-2 would inform the scale of upcoming surges in hospital demand, the proportion of individuals who remain susceptible to contracting the disease, and estimates of key epidemiological parameters such as the epidemic growth rate and the fraction of infections that are subclinical.

The current literature suggests that the predominant symptoms associated with COVID-19 are fever, cough, and sore throat; that is, patients often present with an influenza-like illness (ILI) yet test negative for influenza [4, 5]. As COVID-19 often presents with similar symptoms to influenza, existing surveillance networks in place for tracking influenza could be used to help track COVID-19. Outpatient ILI surveillance has proven to be a useful tool for assessing the impact of influenza [6, 7]. When combined with the number of providers and patients in a given region, ILI surveillance allows estimation of influenza prevalence and severity [8, 9, 10, 11, 12, 13, 14]. Studies of outpatient ILI have repeatedly demonstrated that confirmed influenza case rates underestimate disease burden, likely due to preferential testing of more severe cases [8, 14, 9, 13]. Together these features suggest that ILI surveillance could provide a crucial tool for estimating COVID-19 prevalence within the US.

Here, we quantified the baseline prevalence of non-influenza ILI in the US over the past 10 years and identified a recent surge of non-influenza ILI starting the first week of March, 2020. This surge of excess ILI correlated with known patterns of SARS-CoV-2 spread across states within the US yet was orders of magnitude larger than the number of confirmed COVID-19 cases reported by the end of March.

RESULTS

Influenza-like illness surge

We identified excess ILI cases by first subtracting cases due to influenza and then subtracting the seasonal signal of non-influenza ILI (Fig. 1). Our approach identified known outbreaks of respiratory disease, including the recent outbreak of Respiratory Syncytial Virus that occurred in Washington state in December 2019 [15]. Starting in March of 2020, many states, including Washington, New York, Oregon, Pennsylvania, Maryland, Colorado, New Jersey, and Louisiana, showed a surge in number of non-influenza ILI cases in excess of seasonal norms. For example, in the fourth week of March, 2020, New York State saw approximately 2 times higher non-influenza ILI than it had ever seen since the inception of the ILINet surveillance system within the US. We found that 10.2% of all outpatient visits in New York State during this time were for ILI that could not be explained by either influenza or the normal seasonal variation of respiratory pathogens (8.0% to 11.2% credible set). As the seasonal surge of endemic non-influenza respiratory pathogens declined toward the later weeks in March, this excess ILI correlated more strongly with state-level patterns of newly confirmed COVID-19 cases, suggesting that this surge is a reflection of ILI due to SARS-CoV-2 (Pearson $\rho > 0.35$ and $p < 0.05$ for the last three weeks; fig. S2). The US-wide ILI surge appeared to peak during the week starting on March 15 and subsequently decreased in numerous states the following week; notable exceptions are New York and New Jersey, two of the states that were the hardest hit by the epidemic, which had not started a decline by the week ending March 28.

Changes in care-seeking behavior

We estimated the ILI surge in each US state as an increase in the proportion of outpatients with ILI in that state compared to all outpatient visits in that state. Consequently, changes in the care-seeking behavior of individuals with ILI or of non-ILI during this time period could each affect our estimates of disease prevalence during the ILI surge. If patients with mild ILI were more likely seek medical care during the month of March 2020 than in prior years, then our estimates of COVID-19 prevalence based on the ILI surge would be falsely elevated. Additionally, if non-ILI patients were less likely to seek medical care during the month of March compared to prior years, then this too could falsely elevate our estimates of COVID-19 prevalence based on the ILI surge. Although ILINet does not provide information to ascertain care seeking behavior, we were able to obtain syndromic surveillance data from New York City's emergency departments, which provided up to date information on care seeking behavior of both ILI and non-ILI conditions [16, 17].

If the ILI surge reflected higher rates of detection of typically mild ILI, then we would expect emergency department ILI rates to increase yet the proportion of those ILI cases admitted to the hospital to decrease. However, although the daily number of ILI visits to emergency departments across New York City increased in March 2020, the proportion of those patients who went on to be admitted also increased by as much as 3-fold compared to the baseline rate prior to March (fig. S3A). This observation suggests that patients with mild ILI presented less often to hospital emergency departments. Such a decrease in care-seeking behavior for mild ILI, if similar across the US, could deflate the estimated size of the ILI surge in the later weeks of March by a factor of approximately 3.

If non-ILI patients were less likely to seek medical care, then we would expect that the number of patients complaining of other symptoms not typically associated with COVID-19 (for example vomiting) would also decrease compared to prior years. In the month of March, the daily number of patients presenting with vomiting decreased by as much as a factor of 3 compared to the baseline rate in prior years (fig. S3B). Assuming that all non-ILI conditions were similarly decreased during March, this would suggest that our estimates of the ILI surge could be inflated by as much as a factor of 3. This assumption is conservative as it assumes that even individuals with severe conditions (such as severe trauma) would avoid seeking health-care in response to COVID-19 at the same rate as those with more mild conditions such as vomiting. However, the potential 3-fold decreased care-seeking behavior for non-ILI conditions cancels out the potential 3-fold decreased care-seeking behavior of mild ILI, suggesting that our estimates of prevalence based on the ILI surge may be insensitive to recent changes in care-seeking behavior (fig. S3C). Overall these estimates suggest a conceptual model in which health care utilization for both mild ILI and non-ILI conditions declined at similar rates as COVID-19 increased in the US.

COVID-19 prevalence in the US

To estimate the proportion and magnitude of the March 2020 US ILI surge attributable to SARS-CoV-2 infections, we made the following three assumptions: (1) that the patient population reported by sentinel providers is representative of their state each week; (2) that changes in care-seeking behavior of ILI patients is occurring at a similar rate as that of other non-ILI patients; and (3) that the total number of patients in the US who require medical care over the course of a year has not substantially changed since 2018. Our first assumption is common and underlies prior studies which have used ILI to estimate influenza prevalence [8, 14]. Our second assumption is supported by our New York City analysis which suggests that both mild ILI and non-ILI conditions have seen similar changes in healthcare seeking behavior. Our third assumption is based on the observation that the increasing need for health-care between March 8 and March 28, 2020 due to COVID-19 is likely small compared to the approximately 1

billion outpatient encounters that occur annually [18, 19]. These assumptions together with surveys describing the average number of patients seen by providers [19], the number of providers in each state [20], and the total number of outpatient visits per year [21, 18], allowed us to estimate that, if outpatient clinics remained open during the COVID-19 epidemic, we would expect that there would have been approximately 2.8 million patient encounters with ILI due to COVID-19 between March 8 to March 28, 2020 (95% credible set 2.6 million to 3.0 million).

Not all patients infected with SARS-CoV-2 will present to a health-care provider with ILI. Although we cannot directly measure the rate of such sub-clinical cases, a number of prior studies on asymptomatic rates of COVID-19 and the care-seeking behavior of ILI patients in the US suggest a lower-bound on the subclinical rate of patients with ILI. A recent study of passengers on the Diamond Princess cruise-ship accounted for a right-censoring of patients sampled and estimated that 18% of patients infected with SARS-CoV-2 are asymptomatic for the course of their infection (95% credible set 16% to 20%). This estimate likely represents an underestimate given that the majority of passengers were over 60 years old, a demographic thought to have a lower asymptomatic rate than younger individuals [22]. Beyond asymptomatic individuals, a large study of adult health-care seeking behavior in the United States found that, of a random sample of over 17,000 individuals with ILI, 40% of those went on to seek health care [23]. Together these additional contributions from sub-clinical cases correspond to a mean clinical rate of 32% (the overall rate at which SARS-CoV-2 cases seek medical care) and a lower bound of 8.7 million SARS-CoV-2 infections between March 8th and March 28th (95% credible set 8.0 million to 9.4 million). Prevalence estimates for each state within this time-period are shown in fig. S4.

Syndromic case detection rates

We define the syndromic case detection rate as the number of confirmed COVID-19 cases in a week divided by the size of the ILI surge that week. The syndromic case detection rate varied by state and over time (fig. S5). Our estimated syndromic case detection rates increased over the month of March; this was expected given increases in testing capacity across the US since the February 28 detection of community transmission in Washington State. For the week ending March 14, COVID-19 cases in the states with the highest estimated syndromic case detection rate (Washington, Nevada, and Michigan) only captured approximately 1% of ILI surges in those states. In the last week of the month ending on March 28, the syndromic case detection rate across the US increased to 12.5% (95% credible interval 9.5%-18.3%).

Epidemic growth rates, clinical rates, and infection fatality rates implied by the ILI surge

The true prevalence of SARS-CoV-2 is unknown at the time of this writing. However, if we assume the excess non-influenza ILI is almost entirely due to SARS-CoV-2, an assumption that becomes more valid as SARS-CoV-2 becomes more prevalent, we can use the excess non-influenza ILI to define lower bounds on the exponential growth rate of the US SARS-CoV-2 epidemic. By estimating the number of patients visiting clinics for COVID-19 in the US in March, we can also identify the mutual dependence of exponential growth rates, the rate of sub-clinical infections, and the time between the onset of infectiousness and a patient reporting as ILI (Fig. 2). Using stochastic Susceptible, Exposed, Infectious and Recovered (SEIR) simulations of US COVID-19 epidemics with a January 15 start date [24], we find that an initial epidemic doubling time longer than 4 days is unlikely to explain the ILI surge. Doubling times longer than 4 days fail to produce enough infected individuals to match the observed excess ILI. Doubling time faster than 4 days can explain the observed excess ILI with a clinical rate that depends on the growth rate. Here, we define the clinical rate as the proportion of infected individuals who present to a health care provider.

In keeping with our sub-4 day doubling times, we found that across the entire US, new deaths due to COVID-19 doubled every 3.01 days over the month of March (±0.001, p-value of test that doubling rate is less than 4 days approximately 0). If there was only a 1-day lag from onset of infectiousness to presentation with ILI and the entirety of the first week of the US ILI surge is comprised of patients with COVID-19, then an epidemic starting January 15th and growing at the rate of deaths in the US would imply a 12% clinical rate (Fig. 2A). A four-day lag between the onset of infectiousness and presentation with ILI yields a clinical rate of 25% among the 87% of simulations which could account for the ILI surge. The 25% overall clinical rate estimated from a January 15 start date and the doubling time of US COVID-19 deaths is in close agreement with the 32% clinical rate we estimated independently based on a 18% asymptomatic rate and 40% symptomatic clinical rate. Although our epidemic model suggests the first week of the ILI surge is consistent with the US epidemic start date and growth rate, the ILI surge across the US peaked the week ending March 21, much earlier than our epidemic models, suggesting the epidemic in the US differed from the SEIR model through some combination of factors. Such factors could include successful interventions, even faster decreases in care-seeking than observed in New York, heterogeneity in susceptibility [25], or an early epidemic doubling faster than every 3 days.

Faster growth rates require lower clinical rates to explain the ILI surge. Epidemic curves growing at the rate of deaths in Italy, doubling every 2.65 days, could better match the curvature of the ILI surge by peaking around mid to late March, but would imply a clinical rate of 4.7% the second week of March with a 4-day lag between onset and recorded as ILI

stm.sciencemag.org

(Fig. 2B and C). If the entirety of the ILI surge was attributable to COVID-19, the slowest-possible doubling time for the US epidemic which can explain the ILI surge would be a doubling time of of 4 days. Any evidence of significant secondary introductions, super-spreading, or rapid transmission events in early transmission chains will decrease these estimated clinical rates [26]. Evidence of slow initial spread would increase the estimated clinical rates.

Last, estimating the infection fatality rate from the ILI surge requires knowing the clinical rate and the delay from clinical presentation with ILI to death. If patients present with ILI at the onset of their illness, exhibit a 16 day median lag between onset and death [27], and have a 32% clinical rate as estimated from the 18% asymptomatic rate and 40% clinical rate of symptomatic COVID-19 cases, then the observed ILI surge corresponds to an infection fatality rate of 0.29%. We stress that estimating the infection fatality rate from this ILI surge is highly sensitive to both the lag from presentation with ILI to death and the clinical rate (fig. S6). Consequently, the ILI surge is compatible with fatality rates ranging from 0.07% to 1.4% depending on the unknown sub-clinical rate and lag from presentation with ILI to death. Under the CDC planning scenarios specifying a 4-day lag from onset of symptoms to presentation to the doctor with ILI [28] and a 15 day lag from onset to death, the resulting 11-day lag from ILI to death produces IFR estimates of 0.57% (0.51-0.68% 95% credible set) for the unadjusted ILI surge and 0.19% (0.17-0.22% 95% credible set) for the ILI surge adjusted to account for asymptomatic and subclinical cases.

DISCUSSION

We use outpatient ILI surveillance data from around the US to estimate the prevalence of SARS-CoV-2. We found a clear, anomalous surge in ILI outpatients during the COVID-19 epidemic that correlated with the progression of the epidemic in multiple states across the US. The surge of non-influenza ILI outpatients was much larger than the number of confirmed case in each state, providing evidence of large numbers of probable symptomatic COVID-19 cases that remained undetected. This result is also consistent with ILI excess observed in France in late-February/early-March [29]. Additionally, this finding predicts that the slowest epidemic doubling time that could explain the ILI surge would be 4 days, and that this rate could only be achieved with unusually fast early transmission or super-spreading events and a clinical rate near 100%. Consistent with this prediction, we found that deaths due to COVID-19 within the US doubled every 3.0 days and note that this empirical growth rate for the US epidemic can account for the ILI surge with a 25% clinical rate assuming a 4 day lag from the onset of infectiousness to presentation as an outpatient with ILI. Together, these results suggest that SARS-CoV-2 spread rapidly throughout the

US since its January 15th start date and was likely accompanied by a large undiagnosed population of potential COVID-19 outpatients with presumably milder distribution of clinical symptoms than estimated from prior studies of SARS-CoV-2+ inpatients.

Excess ILI appears to have peaked during the week starting on March 15th, leading the observed ILI dynamics to diverge from the overall epidemic dynamics implied by the growth rate of COVID-19 deaths in the US. If the ILI dynamics were proportional to the epidemic curve then the two could be related via a constant subclinical rate. However, the changing ratio between COVID-19 prevalence estimated by the ILI surge and the epidemic curves parameterized by the growth rate of US deaths suggests additional mechanisms may be behind the ILI slowdown. Mechanisms which can explain the difference between our simulated epidemic curves and the ILI surge include effective social distancing, disproportionate reductions in ILI care-seeking behavior relative to non-ILI care-seeking behavior, or heterogeneity in susceptibility or contact structure not captured in our SEIR model [25].

Our empirical estimate of the size of the ILI surge has several potential limitations. First, the observed ILI surge may represent more than just SARS-CoV-2 infected patients. A second epidemic of a non-seasonal pathogen that presents with ILI could confound our estimates of ILI due to SARS-CoV-2. However, this seems unlikely as additional viral surveillance through the US Centers for Disease Control and Prevention (CDC) suggests that between March 8 to March 28 other monitored respiratory viruses were at low prevalence [30]. Nonetheless, were our approach to be used during winter months, additional steps would be needed to account for concomitant non-influenza seasonal pathogens. Additionally, our assumption that changes in health-care seeking behavior are similar between mild ILI and non-ILI condition may be incorrect. Although this assumption was supported by New York City emergency department surveillance data, it is possible that differential health-care seeking would be present in other locations or in the outpatient setting. Last, it is also possible that our use of ILI data has underestimated the prevalence of SARS-CoV-2 within the US. Although early clinical reports focused on cough and fever as the dominant features of COVID-19 [5], other reports have documented digestive symptoms as the complaint affecting up to half of patients with laboratory-confirmed COVID-19 [31], and alternative presentations, including asymptomatic or unnoticeable infections, could result in underestimation of SARS-CoV-2 prevalence.

Additionally, our models have several limitations. First, we assumed that ILI prevalence within states can be scaled to case counts at the state level. This is based on the assumption that the average number of cases seen by sentinel

providers in a given week is representative of the average number of patients seen by all providers within that state in a given week. Errors in this assumption would cause proportional errors in our estimated case counts and syndromic case detection rate. Second, our US-wide SEIR models vary by growth rate alone and as such may not capture important heterogeneity in susceptibility or transmission as well as regional variation, intervention-induced changes in transmission, or clustering of infection outbreaks. Our models were used to illustrate that the ILI surge is consistent with an estimated growth rate and start date for the US epidemic and to specify the mutual dependency of growth rate, the lag between the onset of infection and presentation to a doctor, and clinical rates. Finer models with regional demographic and case-severity compartments are needed to translate our range of estimated prevalence, growth rate, and clinical rates into actionable models for public health managers. Last, our method of calculating the infection fatality rate relied on assumptions about the clinical rate and the delay from patients recorded as ILI to death. Our clinical rate required using patterns of care-seeking for typical seasonal causes of ILI as did our delay from ILI to death; consequently, neither should be relied on as a definitive source for COVDI-19 and estimating the clinical rate and delay from ILI to death for COVID-19 specifically will reduce the large uncertainty around our ILI-estimated infection fatality rates.

Despite these potential limitations, the ILI surge identified in syndromic surveillance time-series allowed early estimates of COVID-19 prevalence, estimates that were not possible from confirmed case data due to early logistical delays in SARS-CoV-2 testing in the US. Our prevalence estimates are supported by a serosurvey conducted in New York State. We estimated that over 8.3% of New York State residents were infected by SARS-CoV-2 by March 28; on April 23, 2020, New York State announced that 14% of residents had evidence of past infection by SARS-CoV-2 by March 29 at which time the cumulative PCR-confirmed case counts totaled only 0.3% of New York's population [32].

Although an ILI surge tightly correlated with COVID-19 case counts across the US and consistent with the New York State serology strongly suggests that SARS-CoV-2 has potentially infected millions in the US, further laboratory confirmation of our hypotheses are still needed to guide public health decisions. Our findings make testable predictions that one would find relatively high seroprevalence in other states that have already seen an ILI surge and that seroprevalence of individuals infected in March across states is proportional to relative sizes of the states' ILI surges. A study of ILI patients from mid-March who were never diagnosed with COVID-19 could produce a focused test of our predictions about the number and regional prevalence of undetected COVID-19 cases presenting with ILI during that time. If

seroprevalence estimates beyond New York State continue to corroborate our prevalence estimates from syndromic surveillance, this would strongly suggest lower case severity rates for COVID-19 than were assumed in late March by comparing PCR-confirmed case counts to deaths. Further corroboration of our estimates of the magnitude of the ILI surge would suggest ILI and other public time-series of outpatient illness allow early and reliable estimates of crucial epidemiological parameters for rapidly unfolding, novel pandemic diseases. As not all novel pandemic diseases are expected to present with influenza-like symptoms, surveillance of other illnesses that commonly present in the outpatient setting could provide a vital tool for rapidly understanding and responding to novel infectious diseases.

MATERIALS AND METHODS

Study design

The goal of our study was to use publicly available data to estimate the number of patients seeking care for non-influenza ILI in excess of seasonal trends during the three weeks spanning March 8 to March 28, 2020 and then use this ILI surge to estimate COVID-19 incidence in March and parameterize epidemiological model growth rates and clinical rates.

The ILI surge detection above produced an excess proportion of patients visiting outpatient providers for non-influenza ILI in each week and each state. To scale up the proportion of patients to a national number of COVID-19 cases, we estimated the number of patients per sentinel provider in the CDC dataset, normalized that number of patients per provider to a number of patients per doctor, and scaled that up by an estimated number of practicing doctors in the US. The result was an estimated number of COVID-19 patients visiting doctors in each state for each week – we called this our "unadjusted" ILI surge.

The unadjusted ILI surge is an under-estimate of COVID-19 prevalence due to only clinical infections, those that seek medical care. We accounted for both asymptomatic infections and symptomatic but sub-clinical infections to produce an "adjusted" ILI surge as our final estimate of COVID-19 incidence in each state and each week. We then used the unadjusted and adjusted ILI surges to estimate syndromic case detection and fatality rates. We also used the unadjusted ILI surge as an empirical observation to evaluate epidemiological modelling of COVID-19 growth rates and clinical rates in the US.

Throughout our methods, we use i to denote the index state i and let t index week t (with $t=0$ referring to October 3, 2010; the start of state-specific ILINet surveillance).

Data sources

Since 2010 the CDC has maintained ILINet for weekly influenza surveillance. Each week approximately 2,600 enrolled providers distributed throughout all 50 states as well

as Puerto Rico, the District of Columbia, and the US Virgin Islands, report the total number of patient encounters n_{it} and the total number of which met criteria for influenza-like illness (ILI – dined as a temperature 100F [37.8C] or greater, and a cough or sore-throat without a known cause other than influenza; y_{it}.) [37]. For scale, in the 2018-2019 season ILINet reported approximately 60 million outpatient visits. Coupled to these data are weekly state-level reports from clinical and public health labs detailing the number of patient samples tested for influenza n_{it}^{flu} as well as the number of these samples which are positive for influenza y_{it}^{flu}. Therefore ILINet data can be thought of as a weekly state-level time-series representing the superimposed prevalence of various viruses which can cause ILI. ILINet data was obtained through the CDC FluView Interactive portal [33] [38].

In addition to ILINet data, we downloaded US State population data for the 2020 year was downloaded from https://worldpopulationreview.com/states/. The number of primary care providers in each state per 100,000 residents b_i was obtained from the United Health Foundation [20][24]. COVID-19 confirmed case counts were obtained from The New York Times' database maintained at https://github.com/nytimes/covid-19-data. This dataset contains the daily cumulative confirmed case count for COVID-19 for each state z_{il} for day l. The dataset of deaths in Italy was downloaded from https://github.com/pcm-dpc/COVID-19 on April 6, 2020.

Data processing

Within the ILINet dataset, New York City and New York were summed into a combined New York variable representing both New York City and the surrounding state. Due to incomplete data in one or more of the data-sources described above the Virgin Islands, Puerto Rico, The Commonwealth of the Northern Mariana Islands, and Florida were excluded from subsequent analysis. In addition, to match the weekly reporting of ILI from ILINet, daily cumulative confirmed COVID-19 cases were converted to weekly counts of new cases by

$$\tilde{z}_{it} = \sum_{l \in t} z_{il} - z_{i(l-1)}$$

Extracting non-influenza ILI signal

To subtract influenza signal from y_{it} we assumed that the population of patients with ILI within a state are the same population that are potentially tested for influenza. This assumption allows us to calculate the number of non-influenza ILI cases as

$$\tilde{y}_{it} = \left(1 - \frac{y_{it}^{\text{flu}}}{n_{it}^{\text{flu}}}\right) y_{it}$$

Mean imputation based on neighboring states was used to address missing values in laboratory influenza quantification. To assess the impact of this model for extracting non-influenza ILI signal, we calculated COVID-19 prevalence without first removing signal from influenza, we found little change in our prevalence estimates (fig. S7). This likely reflects that influenza also demonstrates strong seasonal patterns that can be addressed as discussed below.

Identifying ILI Surges

We identified ILI surges in $\tilde{y}_{it}$ by training a model on $\tilde{y}_{it}$ for all data prior to July 21, 2019. We then used this model to predict the prevalence of non-influenza ILI $\left(\hat{\pi}_{it}\right)$ for dates after and including July 21, 2019. We calculated the ILI surge as the difference between the observed proportion of non-influenza ILI $\tilde{y}_{it} / n_{it}$ and $\hat{\pi}_{it}$.

To account for variation in the number of total patients, we modeled y_{it} as binomial distributed. To account for correlation in non-influenza ILI over time, we use a Gaussian Process model which assumes that weeks that are closer together will have more similar levels of non-influenza ILI. The following model reflects these modeling choices:

$$\tilde{y}_{it} = \text{Binomial}\left(\pi_{it}, n_{it}\right)$$

$$\pi_{it} = \frac{\exp(\eta_{it})}{1 + \exp(\eta_{it})}$$

$$\eta_{it} \sim N\left(\lambda_i(t), \sigma^2\right)$$

$$\lambda_i(t) =\sim \mathcal{GP}\left(\theta(t), \sigma^2\Gamma\right)$$

$$\sigma^2 \sim \text{InverseGamma}\left(\upsilon, \xi\right)$$

$$\theta(t) = \theta$$

$$\Gamma(t, t+s) = \alpha\exp\left(\frac{-s^2}{2\rho^2}\right)$$

Where $\mathcal{GP}$ refers to a Gaussian process. We made the following prior specifications: We set the bandwidth parameter for the squared exponential kernel as ρ=3 representing a strong local correlation in time that died off sharply beyond 3 weeks, α=1 representing a signal to noise ratio of approximately 1, υ=1 and ξ=1 representing weak prior knowledge regarding the overall scale of variation in the latent space. Finally, we set θ=−2.197 representing an off-season prevalence of 0.1% non-influenza ILI. Samples from the posterior predictive density $p\left(\pi_{it} \mid \tilde{y}_{i1}, \ldots, \tilde{y}_{iT}, n_{i1}, \ldots, n_{iT}\right)$ were

collected using the function *basset* from the R package *stray* [34]; a total of 4000 such samples were collected, for each state, in this analysis. We defined the prevalence of non-influenza ILI in excess of normal seasonal variation as

$$y_{it}^* = \tilde{y}_{it} / \eta_{it} - \hat{\pi}_{it}.$$

To investigate whether our results were sensitive to the above model specification, we alternatively used the sample mean and variance from years 2010-2018 as an estimate of typical seasonal non-influenza ILI. Despite not accounting for the binomial count structure of ILI data or correlations in the proportion ILI between weeks, this simpler model resulted in nearly identical prevalence estimates (fig. S8). Still, we used the GP-derived estimates throughout this paper due to their better accounting for the known binomial count and week-to-week correlation structure of ILI-causing pathogen prevalence.

To exclude variation attributable to unseasonably high rates of other ILI causing viruses (such as the outbreak of RSV in Washington state in November-December 2019) we only investigated y_{it}^* for weeks after March 7th 2020 as only these later weeks had high correlation to the COVID-19 confirmed case rate (fig. S2).

Calculating scaling factors to relate ILINet data to COVID-19 cases

As new COVID-19 case counts $\tilde{z}_{it}$ represent the number of confirmed cases in an entire state and ILINet data represents the number of cases seen by a select number of enrolled providers, we had to estimate scaling factors w_i to enable comparison of ILINet data to confirmed case counts at the state level. Let π_{it}^* denote the probability that a patient with ILI in state i has COVID-19 as estimated from ILINet data. Let p_i denote the population of state i and let b_i denote the number of primary care providers per 100,000 people in state i. We translated the inferred proportion of individuals with ILI due to COVID-19 to the state level by considering the average number of patients seen across all providers in the state in a 5-day work-week. In addition, we added a discount factor λ=0.55 to calibrate these estimates with prior reports regarding the total number of outpatient visits per year [18]. This yielded our estimated number of COVID-19 cases (excess ILI at the state level) as

$$w_i = \frac{5 b_i p_i}{10^5} m \lambda$$

$$y_{it}^\dagger = w_i \pi_{it}$$

where m=20.2 is the mean number of patients seen by physicians per day [19].

Accounting for sub-clinical infections

To account for the contribution of sub-clinical SARS-CoV-2 infections we used a recent analysis of cohort surveillance from the Diamond Princess [35]. Monte-Carlo simulations were used to propagate error from our uncertainty regarding potential asymptomatic infections affecting the clinical rate δ^b into our calculation of posteriors for epidemic trajectories. To match posterior estimates, we used quantile matching to parameterize $\delta^c \sim \text{Beta}(\alpha, \beta)$ to achieve a mean of 0.179 and a 95% probability set of (0.155, 0.202). In addition, we took $\delta^c = 0.4$ based on a large study of adult health-care seeking behavior in the United States [23]. To account for these sub-clinical contributions we used adjusted scaling factors

$$w_i^* = \frac{w_i}{\left(1 - \delta^c\right)\left(1 - \delta^b\right)}$$

Estimating syndromic case detection rates

Assuming that the majority of SARS-CoV-2 testing within the US has been directed by patient symptoms [36], the pool of newly diagnosed SARS-CoV-2+ patients is a subset of the pool of SARS-CoV-2+ patients who are identified as having ILI. Therefore, we calculated the probability that a SARS-CoV-2+ patient with ILI who seeks medical care will be identified as having SARS-CoV-2 as $\delta^s = \tilde{z}_{ij} / y_{it}^\dagger$ (fig. S5).

Estimating infection fatality rates

The exact lag from an outpatient being recorded as ILI to death is unknown, but estimated lag times from onset to death and from hospitalization to death [27] can be used to understand the range of implied infection fatality rates from the ILI surge. We calculated the infection fatality rate implied by the ILI surge as a function of the unknown lag from patients being recorded as ILI and death, and we repeat this calculation for both the raw and subclinical rate adjusted ILI estimates. For a lag of l days from ILI reporting to death, the infection fatality rate was estimated by dividing the magnitude of the adjusted or raw ILI surge by all new deaths occurring within the dates (2020/03/08 + l, ..., 2020/03/28 + l). A plot of the fatality rate by lag for raw and unadjusted ILI surges revealed a large range of fatality rates compatible with the ILI surge and highly sensitive to the estimate of lag and clinical rates. One study [27] estimated a median 11.2 days from hospitalization to death and 16.1 days from symptom onset to death. For the raw ILI surge estimate, 11 day and 16 day lag times would produce median infection fatality rate estimates of 0.57% and 0.89%, respectively, without adjusting for any subclinical infections; for the subclinical-adjusted ILI surge estimate, these lag times would produce median infection fatality rate estimates of 0.19% and 0.29%, respectively.

Growth rate estimation

As of April 6, 2020, deaths from SARS-CoV-2 epidemic were still growing nearly exponentially as evidenced by nearly linear growth on a log y axis. Early in the epidemic, estimating exponential growth rates by Poisson regression with a log link function produces accurate estimates of the true growth rate [37], and so we estimated growth rates for the US and Italy by Poisson generalized linear models predicting new deaths using date as a quantitative explanatory variable. US COVID-19 deaths from March 5, 2020 to April 1, 2020, were summed by date to calculate national-level statistics. Initially, April 2-5 were included but were found to have anomalously high leverage and were hence excluded from our analysis. We applied the same procedure to COVID-19 deaths in Italy, focusing on deaths from February 24 until March 12. We used the slope from Poisson regression as the estimated exponential growth rate, which yielded a US growth rate of $r_{\text{US}} = 0.23$ or a 3.01 day doubling time of the infected population over time and an Italian growth rate of $r_{\text{IT}} = 0.26$ or a 2.65-day doubling time of the infected population over time.

Epidemic simulations and clinical rates

The following SEIR models

$$\dot{S} = \zeta - \beta SI - \omega_b S$$
$$\dot{E} = \beta SI - \gamma E - \omega_b E$$
$$\dot{I} = \gamma E - \upsilon I - \omega_i I$$
$$\dot{R} = \upsilon I - \omega_b R$$

were parameterized for the US to a timescale of units days by setting $\zeta = 3.23 \times 10^{-5}$ corresponding to a crude birth rate of 11.8 per 1000 per year, a baseline mortality rate $\omega_b = 2.38 \times 10^{-7}$ corresponding to 8.685 per 1000 per year, and an infectious mortality rate $\omega_i = 4.96 \times 10^{-4}$ corresponding to a infection fatality rate of 0.5% required to fit US deaths under a 20-day lag from onset to death. Further, we drew a random incubation period γ^{-1} ~LogNormal (1.087,0.153) reflecting empirical estimates of a median 5 days from exposure to symptom onset with 4.2-6 day 95% credible interval [38, 39] which is then offset by 2 days of pre-symptomatic transmission as documented across carefully studied clusters in Singapore [40] resulting in a 2.2-4 day 95% credible interval for a log-normally distributed incubation period. Similarly, in each simulation we also drew a random infectious period $\upsilon^{-1} \sim \text{LogNormal}(2.193,0.105)$ based on 2 days of pre-symptomatic infectiousness and high viral loads in nasopharyngeal samples [41, 42] combined with persistence of high loads of SARS-CoV-2 that can be cultured up to 7 days after symptom onset [43], resulting in our use of a 7.3-11 day

95% credible interval for the infectious period. Finally, we parameterized β to ensure $I(t)$ grew with a specified exponential growth rate early in the epidemic. We ran a total of 2,000 simulations for each of the two growth rate distributions (US and Italy) analyzed. Growth rates were drawn at random from a normal distribution with standard deviation of 0.1 and centered on r_{US} and r_{IT}, respectively. To illustrate the mutual dependence between estimates of growth rate, clinical rate, and the lag between the onset of infectiousness to presentation to a doctor with ILI, we ran 2,000 simulations with uniform growth rates in the interval [0.173,0.365] corresponding to a range of doubling times between 1.9 days and 4 days.

Each simulation was initialized with $(S, E, I, R, t) = (3.27 \times 10^8, 0, 1, 0, 0)$ where time 0 was January 15 and simulations were run until August 5, 2020. The SEIR model was simulated with a Gillespie algorithm through the R package adaptivetau [44] on the assumption that a large amount of variation in the epidemic trajectory stems from uncertainty in trajectory of early transmission chains. The number of infected individuals on a given day was the last observed $I(t)$ for that day, and a weekly pool of infected patients was computed by a moving sum over the number of infected individuals every day for the past week,

$$I_w(t) = \sum_{k=0}^{6} I_{t-k}$$

Defining $Y_t = \sum_i y_{it}^\dagger$ as the national excess ILI, the clinical rate implied by a given simulation was estimated as

$$\delta^c(t_d) = \frac{Y_t}{I_w(t - t_d)}$$

for a given time delay t_d it takes from the onset of infectiousness to a patient reporting to the doctor with ILI.

REFERENCES AND NOTES

1. N. Zhu, D. Zhang, W. Wang, X. Li, B. Yang, J. Song, X. Zhao, B. Huang, W. Shi, R. Lu, P. Niu, F. Zhan, X. Ma, D. Wang, W. Xu, G. Wu, G. F. Gao, W. Tan; China Novel Coronavirus Investigating and Research Team, A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N. Engl. J. Med.* **382**, 727–733 (2020). doi:10.1056/NEJMoa2001017 Medline

2. World Health Organization, "Coronavirus disease 2019 (COVID-19): situation report, 59," 2020.

3. J. Lourenco, R. Paton, M. Ghafari, M. Kraemer, C. Thompson, P. Simmonds and S. Klenerman, "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.

4. N. Chen, M. Zhou, X. Dong, J. Qu, F. Gong, Y. Han, Y. Qiu, J. Wang, Y. Liu, Y. Wei, J. Xia, T. Yu, X. Zhang, L. Zhang, Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: A descriptive study. *Lancet* **395**, 507–513 (2020). doi:10.1016/S0140-6736(20)30211-7 Medline

5. D. Wang, B. Hu, C. Hu, F. Zhu, X. Liu, J. Zhang, B. Wang, H. Xiang, Z. Cheng, Y. Xiong, Y. Zhao, Y. Li, X. Wang, Z. Peng, Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. *JAMA* **323**, 1061–1069 (2020). doi:10.1001/jama.2020.1585 Medline

6. T. L. Brammer, E. L. Murray, K. Fukuda, H. E. Hall, A. Klimov, N. J. Cox, Surveillance for influenza—United States, 1997-98, 1998-99, and 1999-00 seasons. *MMWR Surveill. Summ.* **51**, 1–10 (2002). Medline

7. R. Garten, L. Blanton, A. I. A. Elal, N. Alabi, J. Barnes, M. Biggerstaff, L. Brammer, A. P. Budd, E. Burns, C. N. Cummings, T. Davis, S. Garg, L. Gubareva, Y. Jang, K. Kniss, N. Kramer, S. Lindstrom, D. Mustaquim, A. O'Halloran, W. Sessions, C. Taylor, X. Xu, V. G. Dugan, A. M. Fry, D. E. Wentworth, J. Katz, D. Jernigan, Update: Influenza activity in the United States during the 2017–18 season and composition of the 2018–19 influenza vaccine. *Morbidity and Mortality Weekly Report* **67**, 634–642 (2018). doi:10.15585/mmwr.mm6722a4 Medline

8. J. Buffington, L. E. Chapman, L. M. Schmeltz, A. P. Kendal, Do family physicians make good sentinels for influenza? *Arch. Fam. Med.* **2**, 859–864, discussion 865 (1993). doi:10.1001/archfami.2.8.859 Medline

9. J. Oliva, C. Delgado-Sanz, A. Larrauri; Spanish Influenza Surveillance System, Estimating the burden of seasonal influenza in Spain from surveillance of mild and severe influenza disease, 2010-2016. *Influenza Other Respir. Viruses* **12**, 161–170 (2018). doi:10.1111/irv.12499 Medline

10. V. Sotomayor, R. A. Fasce, N. Vergara, F. De la Fuente, S. Loayza, R. Palekar, Estimating the burden of influenza-associated hospitalizations and deaths in Chile during 2012-2014. *Influenza Other Respir. Viruses* **12**, 138–145 (2018). doi:10.1111/irv.12502 Medline

11. L. Feng, S. Feng, T. Chen, J. Yang, Y. C. Lau, Z. Peng, L. Li, X. Wang, J. Y. T. Wong, Y. Qin, H. S. Bond, J. Zhang, V. J. Fang, J. Zheng, J. Yang, P. Wu, H. Jiang, Y. He, B. J. Cowling, H. Yu, Y. Shu, E. H. Y. Lau, Burden of influenza-associated outpatient influenza-like illness consultations in China, 2006-2015: A population-based study. *Influenza Other Respir. Viruses* **14**, 162–172 (2020). doi:10.1111/irv.12711 Medline

12. C. C. van den Wijngaard, L. van Asten, A. Meijer, W. van Pelt, N. J. D. Nagelkerke, G. A. Donker, M. A. van der Sande, M. P. Koopmans, Detection of excess influenza severity: Associating respiratory hospitalization and mortality data with reports of influenza-like illness by primary care physicians. *Am. J. Public Health* **100**, 2248–2254 (2010). doi:10.2105/AJPH.2009.168245 Medline

13. Q. S. Huang, L. Lopez, B. Adlam, Influenza surveillance in New Zealand in 2005. *N. Z. Med. J.* **120**, U2581 (2007). Medline

14. C. Reed, S. S. Chaves, P. Daily Kirley, R. Emerson, D. Aragon, E. B. Hancock, L. Butler, J. Baumbach, G. Hollick, N. M. Bennett, M. R. Laidler, A. Thomas, M. l. Meltzer, L. Finelli, Estimating influenza disease burden from population-based surveillance data in the United States. *PLOS ONE* **10**, e0118369 (2015). doi:10.1371/journal.pone.0118369 Medline

15. Washington State Department of Health, *Influenza Update, Week 19,* 2020.

16. New York City Department of Heath and Mental Hygiene, *Influenza Surveillance Report, Week 12,* 2020.

17. NYC Heath Data*, EpiQuery, Syndromic Surveillance,* 2020.

18. P. Rui, T. Okeyode, *National Ambulatory Medical Care Survey: 2016 National Summary Tables,* 2016.

19. The Physicians Foundation, *2018 Physician Survey,* 2018.

20. United Health Foundation, *America's Health Rankings analysis of Special data request for information on active state licensed physicians,* 2019.

21. D. A. Squires, The U.S. health system in perspective: A comparison of twelve industrialized nations. *Issue Brief (Commonw. Fund)* **16**, 1–14 (2011). Medline

22. J. T. Wu, K. Leung, M. Bushman, N. Kishore, R. Niehus, P. M. de Salazar, B. J. Cowling, M. Lipsitch, G. M. Leung, Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. *Nat. Med.* **26**, 506–510 (2020). doi:10.1038/s41591-020-0822-7 Medline

23. Centers for Disease Control and Prevention (CDC), Self-reported influenza-like illness during the 2009 H1N1 influenza pandemic—United States, September 2009 - March 2010. *Morbidity and Mortality Weekly Report* **60**, 37–41 (2011). Medline

24. M. L. Holshue, C. DeBolt, S. Lindquist, K. H. Lofy, J. Wiesman, H. Bruce, C. Spitters, K. Ericson, S. Wilkerson, A. Tural, G. Diaz, A. Cohn, L. Fox, A. Patel, S. I. Gerber, L. Kim, S. Tong, X. Lu, S. Lindstrom, M. A. Pallansch, W. C. Weldon, H. M. Biggs, T. M. Uyeki, S. K. Pillai; Washington State 2019-nCoV Case Investigation Team, First case of 2019 novel coronavirus in the United States. *N. Engl. J. Med.* **382**, 929–936 (2020). doi:10.1056/NEJMoa2001191 Medline

25. J. C. Miller, Epidemic size and probability in populations with heterogeneous infectivity and susceptibility. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **76**, 010101 (2007). doi:10.1103/PhysRevE.76.010101 Medline

26. J. O. Lloyd-Smith, S. J. Schreiber, P. E. Kopp, W. M. Getz, Superspreading and the effect of individual variation on disease emergence. *Nature* **438**, 355–359 (2005). doi:10.1038/nature04153 Medline

27. S. Sanche, Y. T. Lin, C. Xu, E. Romero-Severson, N. Hengartner, R. Ke, High Contagiousness and Rapid Spread of Severe Acute Respiratory Syndrome Coronavirus 2. *Emerg. Infect. Dis.* **26**, 1470–1477 (2020). doi:10.3201/eid2607.200282 Medline

28. M. Biggerstaff, M. A. Jhung, C. Reed, A. M. Fry, L. Balluz, L. Finelli, Influenza-like illness, the time to seek healthcare, and influenza antiviral receipt during the 2010-2011 influenza season-United States. *J. Infect. Dis.* **210**, 535–544 (2014). doi:10.1093/infdis/jiu224 Medline

29. P.-Y. Boë, C. Souty, T. Launay, C. Guerrisi, C. Turbelin, S. Behillil, V. Enouf, C. Poletto, B. Lina, S. van der Werf, D. Lévy-Bruhl, V. Colizza, T. Hanslik, T. Blanchon, Excess cases of influenza-like illnesses synchronous with coronavirus disease (COVID-19) epidemic, France, March 2020. *Euro Surveill.* **25**, 2000326 (2020). Medline

30. *The National Respiratory and enteric Virus Surveillance System,* 2020.

31. L. Pan, M. Mu, P. Yang, Y. Sun, R. Wang, J. Yan, P. Li, B. Hu, J. Wang, C. Hu, Y. Jin, X. Niu, R. Ping, Y. Du, T. Li, G. Xu, Q. Hu, L. Tu, Clinical characteristics of COVID-19 patients with digestive symptoms in Hubei, China: A descriptive, cross-sectional, multicenter study. *Am. J. Gastroenterol.* **115**, 766–773 (2020). doi:10.14309/ajg.0000000000000620 Medline

32. E. S. Rosenberg, J. M. Tesoriero, E. M. Rosenthal, R. Chung, M. A. Barranco, L. M. Styer, M. M. Parker, S.-Y. J. Leung, J. Morne, D. Greene and others, "Cumulative incidence and diagnosis of SARS-CoV-2 infection in New York," *medRxiv,* 2020.

33. *FluView Interactive,* 2020.

34. J. D. Silverman, K. Roche, Z. C. Holmes and S. David, "Bayesian Multinomial Logistic Normal Models through Marginally Latent Matrix-T Processes," *arXiv,* 2019.

35. K. Mizumoto, K. Kagaya, A. Zarebski, G. Chowell, Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. *Euro Surveill.* **25**, ••• (2020). doi:10.2807/1560-7917.ES.2020.25.10.2000180 Medline

36. Hopkins Medicine, *Coronavirus Test: What You Need to Know,* 2020.

37. J. Ma, J. Dushoff, B. M. Bolker, D. J. D. Earn, Estimating initial epidemic growth rates. *Bull. Math. Biol.* **76**, 245–260 (2014). doi:10.1007/s11538-013-9918-2 Medline

38. N. M. Linton, T. Kobayashi, Y. Yang, K. Hayashi, A. R. Akhmetzhanov, S. M. Jung, B. Yuan, R. Kinoshita, H. Nishiura, Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: A statistical analysis of publicly available case data. *J. Clin. Med.* **9**, 538 (2020). doi:10.3390/jcm9020538 Medline

39. 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. *Ann. Intern. Med.* **172**, 577–582 (2020). [doi:10.7326/M20-0504](doi:10.7326/M20-0504) Medline

40. W. E. Wei, Z. Li, C. J. Chiew, S. E. Yong, M. P. Toh, V. J. Lee, Presymptomatic Transmission of SARS-CoV-2 - Singapore, January 23-March 16, 2020. *Morbidity and Mortality Weekly Report* **69**, 411–415 (2020). doi:10.15585/mmwr.mm6914e1 Medline

41. L. Zou, F. Ruan, M. Huang, L. Liang, H. Huang, Z. Hong, J. Yu, M. Kang, Y. Song, J. Xia, Q. Guo, T. Song, J. He, H.-L. Yen, M. Peiris, J. Wu, SARS-CoV-2 viral load in upper respiratory specimens of infected patients. *N. Engl. J. Med.* **382**, 1177–1179 (2020). doi:10.1056/NEJMc2001737 Medline

42. X. He, E. H. Y. Lau, P. Wu, X. Deng, J. Wang, X. Hao, Y. C. Lau, J. Y. Wong, Y. Guan, X. Tan, X. Mo, Y. Chen, B. Liao, W. Chen, F. Hu, Q. Zhang, M. Zhong, Y. Wu, L. Zhao, F. Zhang, B. J. Cowling, F. Li, G. M. Leung, Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat. Med.* **26**, 672–675 (2020). [doi:10.1038/s41591-020-0869-5](doi:10.1038/s41591-020-0869-5) Medline

43. R. Wölfel, V. M. Corman, W. Guggemos, M. Seilmaier, S. Zange, M. A. Müller, D. Niemeyer, T. C. Jones, P. Vollmar, C. Rothe, M. Hoelscher, T. Bleicker, S. Brünink, J. Schneider, R. Ehmann, K. Zwirglmaier, C. Drosten, C. Wendtner, Virological assessment of hospitalized patients with COVID-2019. *Nature* **581**, 465–469 (2020). [doi:10.1038/s41586-020-2196-x](doi:10.1038/s41586-020-2196-x) Medline

44. P. Johnson, *adaptivetau: efficient stochastic simulations in R,* 2014.
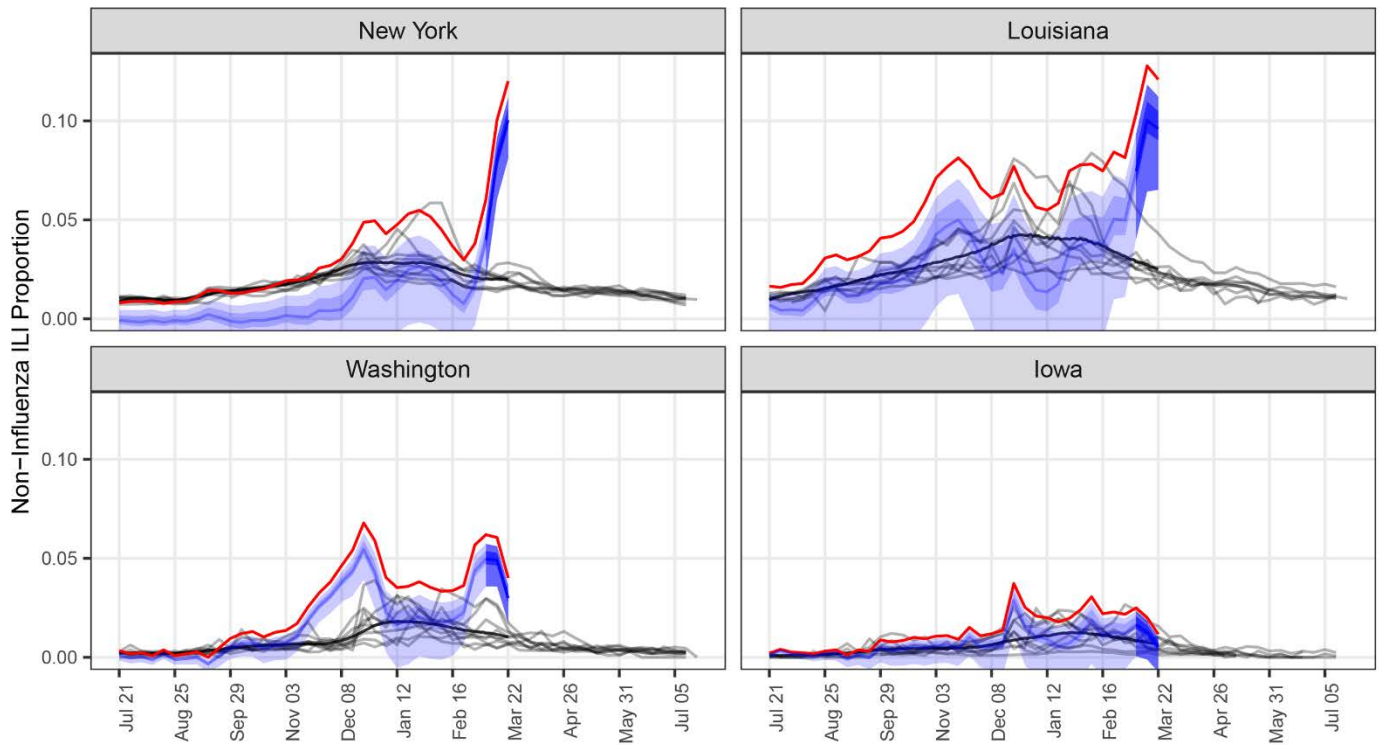
Fig. 1. An early surge of ILI visits across the US. The proportion of patients presenting with ILI that could not be explained by influenza or typical seasonal variation (that is, excess ILI) is shown for four states (blue line and ribbons represent the posterior median as well as 95% and 50% credible sets; results from all analyzed states are shown in fig. S1). ILI that could not be attributed to influenza was calculated based on influenza laboratory surveillance data (2019-2020 flu season shown in red, prior seasons are shown in black). A time-series model was used to infer seasonal variation of non-influenza ILI. Excess ILI was then calculated as the difference between non-influenza ILI from 2019-2020 and the seasonal baseline of non-influenza ILI. Excess ILI after March 7th is highlighted in darker blue as these data correlated strongly with observed COVID-19 case counts (fig. S2).
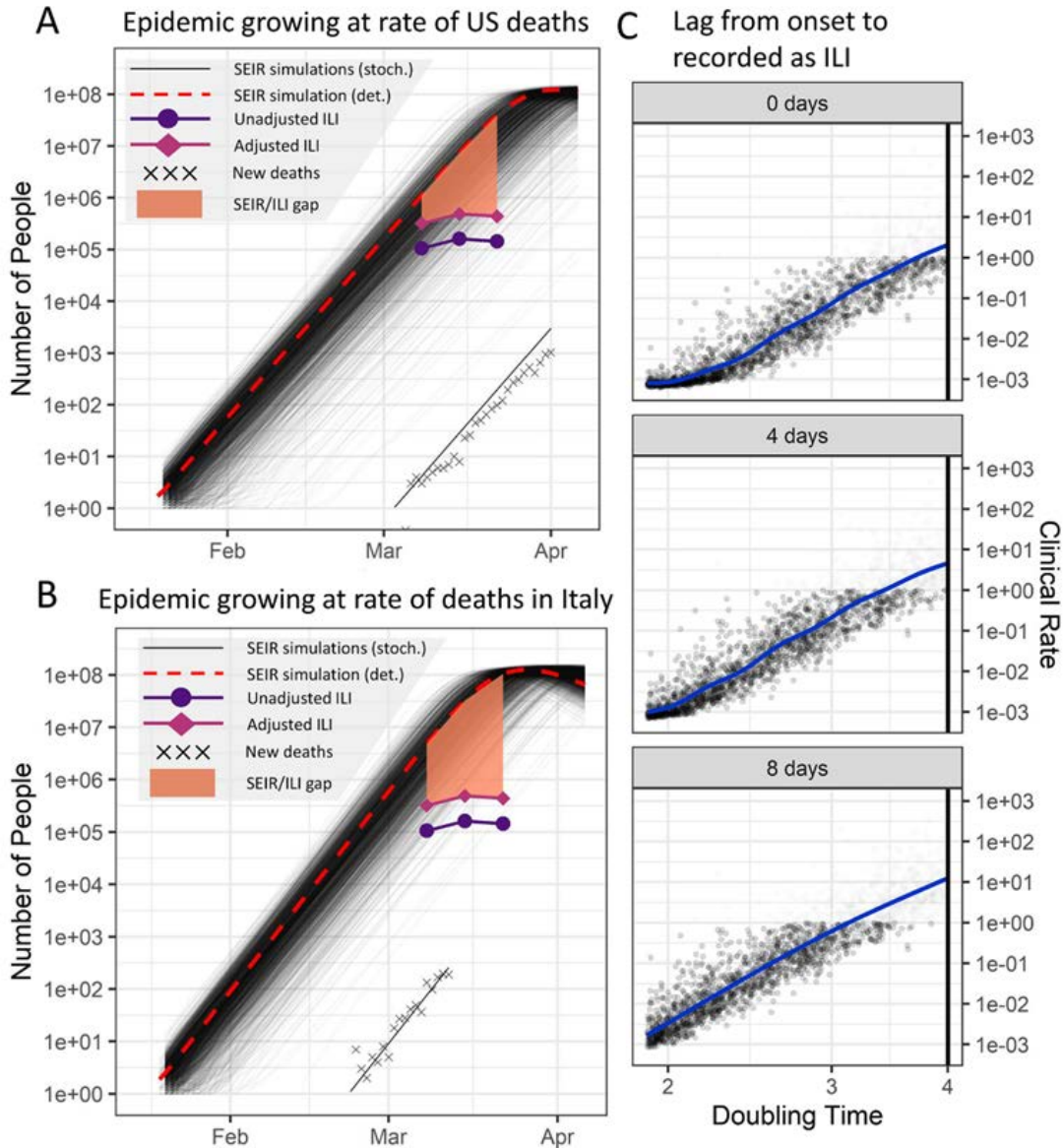
Fig. 2. The ILI surge imposes a dependence between growth rate and clinical rate in epidemiological models (A-B) SARS-CoV-2 prevalence estimates based on the ILI surge are consistent with an epidemiological model parameterized based on a January 15th epidemic start date and a doubling time equal to that observed for new deaths within the US (A) or Italy (B). Epidemiological models were either stochastic (simulated via tau-leaping) or deterministic (solved by numerical integration). In addition to our raw estimates of the ILI surge size (*unadjusted*), we provide adjusted prevalence estimates accounting for sub-clinical cases by assuming an 18% asymptomatic rate and a 40% rate of health-care seeking of symptomatic ILI patients (*adjusted*). Epidemic trajectories were simulated using an SEIR model (black lines). The increasing gap between ILI prevalence estimates and SEIR trajectories (orange) suggest the presence of additional factors such social distancing, changes in care-seeking behavior, or heterogeneity in susceptibility or transmission. (C) More generally, the size of the clinical population estimated from ILI data imposes a dependence between epidemic doubling time, the clinical rate, and the lag between onset of infectiousness and ILI reporting. Combinations of these three variables that are consistent (black) or inconsistent (gray) are shown as well as a smoothed estimate of clinical rate as a function of doubling time.