

The Impact of Changes in Diagnostic Testing Practices on Estimates of COVID-19 Transmission in the United States

Virginia E. Pitzer, Melanie Chitwood, Joshua Havumaki, Nicolas A. Menzies², Stephanie Perniciaro¹, Joshua L. Warren³, Daniel M. Weinberger¹, Ted Cohen¹

Correspondence to: Virginia E. Pitzer, Yale School of Public Health, P.O. Box 208034, New Haven, CT 06520-8034 (e-mail: virginia.pitzer@yale.edu)

Author affiliations: Department of Epidemiology of Microbial Diseases and Public Health Modeling Unit, Yale School of Public Health, Yale University, New Haven, Connecticut, United States (Virginia E. Pitzer, Melanie Chitwood, Joshua Havumaki, Stephanie Perniciaro, Daniel M. Weinberger, and Ted Cohen); Department of Global Health and Population, Harvard T.H. Chan School of Public Health, Harvard University, Boston, Massachusetts, United States (Nicolas A. Menzies); and Department of Biostatistics and Public Health Modeling Unit, Yale School of Public Health, Yale University, New Haven, Connecticut, United States (Joshua L. Warren)

Funding: This work was supported by the following grants from the National Institutes of Health/National Institute of Allergy and Infectious Diseases (NIH/NIAID): R01AI137093 (VEP, JLW, DMW), R01AI112970 (VEP), R01AI123208 (DMW), and R01AI146555 (NAM, TC).

Conflict of interest: VEP has received reimbursement from Merck and Pfizer for travel expenses to Scientific Input Engagements unrelated to the topic of this manuscript. DMW has received consulting fees from Pfizer, Merck, GSK, and Affinivax for topics unrelated to this manuscript and is Principal Investigator on a research grant from Pfizer on an unrelated topic. All other authors report no relevant conflicts.

Running head: Testing Biases Estimates of COVID-19 Transmission

© The Author(s) 2021. Published by Oxford University Press on behalf of the Johns Hopkins Bloomberg School of Public Health. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com.

ABSTRACT

Estimates of the reproductive number for novel pathogens such as severe acute respiratory syndrome coronavirus 2 are essential for understanding the potential trajectory of the epidemic and the level of intervention that is needed to bring the epidemic under control. However, most methods for estimating the basic reproductive number (R_0) and time-varying effective reproductive number (R_t) assume that the fraction of cases detected and reported is constant through time. We explore the impact of secular changes in diagnostic testing and reporting on estimates of R_0 and R_t using simulated data. We then compare these patterns to data on reported cases of coronavirus disease and testing practices from different states in the United States from March 4 to August 30, 2020. We find that changes in testing practices and delays in reporting can result in biased estimates of R_0 and R_t . Examination of changes in the daily number of tests conducted and the percent of patients testing positive may be helpful for identifying the potential direction of bias. Changes in diagnostic testing and reporting processes should be monitored and taken into consideration when interpreting estimates of the reproductive number of coronavirus disease.

Keywords: basic reproduction number; coronavirus; mathematical model; reproductive number; transmission dynamics

Abbreviations: COVID-19, coronavirus disease 2019; US, United States; R_0 , basic reproductive number; R_t , time-varying effective reproductive number; SEIR, Susceptible-Exposed-Infected-Recovered; $\%_{pos}$, percent of individuals testing positive; CI, confidence interval

The initial stages of the coronavirus disease 2019 (COVID-19) epidemic in the United States (US) were characterized by difficulties in delivering and administering diagnostic tests (1). First, the real-time quantitative PCR assay developed and distributed by the US Centers for Disease Control and Prevention suffered from performance issues (2,3). As a result, all initial and confirmatory testing needed to be carried out by the Centers for Disease Control and Prevention, which led to reporting delays and capacity issues early in the epidemic (4,5). Initially, tests were only administered to individuals with a history of travel to certain countries or known contact with a positive case. By the time testing capacity increased, state and local health departments were faced with heavy demand for COVID-19 testing. Only individuals meeting specific criteria could receive a test, and these criteria have varied from state to state and over time.

Concurrently, mathematical modelers were analyzing data on reported COVID-19 cases in order to develop forecasts of future incidence and evaluate the potential impact of control measures, often at the behest of policymakers and public health officials. These models typically rely on estimates of the reproductive number of the virus. The basic reproductive number (R_0) is defined as the expected number of secondary infections produced by an infectious individual in a fully susceptible population; this can be used to derive the expected fraction of the population that will become infected in the absence of interventions and the level of control and/or immunity that is needed to eliminate the pathogen from circulation (6). The time-varying effective reproductive number (R_t) measures the average number of secondary infections per case at each time-point in the epidemic (6), and can be used for real-time

monitoring of the impact of control measures (7–9). As control measures are implemented and immunity increases in the population, R_t will decrease (6,10). The value of R_0 can be estimated from the growth rate of the number of cases early in the epidemic and estimates of the generation time (i.e. the time between infection of successive cases in a transmission chain) (11), whereas instantaneous values of R_t can be estimated based on the time series of case notifications and the distribution of the generation time (7–9). These methods have been shown to be robust to under-detection and underreporting, so long as the probability that a true case is detected and reported remains constant through time (7–9,11,12).

Here, we use simulations to explore the potential magnitude and direction of biases introduced by changes in diagnostic testing and reporting practices similar to those occurring during the early stages of the COVID-19 epidemic in the US. We then compute preliminary estimates of R_0 and R_t for different states, based on publicly available data from The COVID Tracking Project (13), and identify where these estimates are likely to be biased. We examine changes in testing practices and trends in the percent of tests that are positive to evaluate the potential for bias.

METHODS

Examining the impact of changes in testing using simulated data

We simulated data using a stochastic SEIR (Susceptible-Exposed-Infected-Recovered) model to explore the potential impact of changes in testing practices on estimates of R_0 and R_t . We modeled a population of 1 million individuals and initialized the epidemic with 10 infectious individuals to minimize the chances of early epidemic fadeout. We assumed everyone else was

susceptible at the start of the epidemic. New infections were assumed to arise according to a Poisson process (approximate tau-leaping method); the state transitions and rates are described in Table 1, and model parameters are given in Table 2. We simulated the model to day 70 using a time-step of $\Delta t=0.05$ days, and assumed a decrease in the transmission rate occurring on day 50, consistent with the impact of social distancing interventions.

We tracked the number of “true cases” on day d (Y_d) as the number of individuals entering the infectious period each day. We assumed that each true case occurring on day d had a probability $p_{\text{rep}}(d)$ of being tested, testing positive, and being reported (where $p_{\text{rep}}(d)$ is a probability conditional on being a true case). The number of tests performed was assumed to scale with the true incidence, such that for every true case that occurred on day d , there were $n_{\text{test}}(d)$ individuals with similar symptoms who were tested (i.e. $n_{\text{test}}(d)$ is the ratio of the number of tests to the number of true cases). We assumed that testing and reporting of test results occurred with some delay, which followed a gamma probability distribution, $\delta_{\text{rep},d}(t)$, with parameters a_d and b_d . Thus, we calculated the number of individuals tested (T_d) and the number of positive cases (C_d) on day d as follows:

$$T_d = \sum_{i=1}^d n_{\text{test}}(i) Y_i \delta_{\text{rep},i}(d-i)$$

$$C_d = \sum_{i=1}^d \text{Binomial}(Y_i, p_{\text{rep}}(i) \delta_{\text{rep},i}(d-i))$$

We rounded the value of T_d to the nearest integer and sampled C_d from a binomial distribution, with the additional constraint that $C_d \leq T_d$. As an indicator of changes in testing practices, we

estimated the percent of individuals testing positive as $\hat{\%}_{pos} = C_d/T_d$. The expected value of $\%_{pos}$ is given by: $E(\%_{pos}) = p_{rep}(d)/n_{test}(d)$.

As our base case, we assumed that the probability of a “true case” being reported was $p_{rep}=0.1$ and the ratio of tests performed to total number of true cases was $n_{test}=0.5$; we assumed a mean reporting delay between onset of infectiousness and testing results of 6.6 days (14). We then modelled scenarios in which the fraction of true cases detected and reported (as indicated by p_{rep}) and the testing capacity (i.e. ratio of individuals tested to true cases, n_{test}) either increased or decreased linearly between days 20 and 60 of the epidemic. To examine the impact of sudden changes to the probability of a true case being tested (e.g. associated with changes in testing criteria) and testing capacity (e.g. associated with a rapid expansion of supply), we also explored scenarios in which p_{rep} and n_{test} increased or decreased abruptly on day 40 (Table S1). Finally, we examined the effect of a two-fold increase or decrease in the average reporting delay.

We performed sensitivity analyses to explore the how the results varied for different magnitudes of increase in p_{rep} and n_{test} (Web Appendix 1). We also explored a more realistic model for the disease and observation process, in which the probability of a true case being tested and reported, as well as the associated reporting delay, depends on symptom status and disease severity, and testing capacity increases independent of the number of true cases after the initial stages of the epidemic (Web Appendix 2).

Estimation of R_0 from simulated data

We estimated R_0 from the growth rate of the epidemic, as described by Lipsitch et al (11):

$$R_0 = r^2 f(1 - f)V^2 + rV + 1,$$

where r is the growth rate, V is the generation interval (which is not typically observable and usually approximated by the serial interval (15)), and f is the proportion of the generation interval spent in the latent period. The growth rate (r) was determined by fitting Poisson regression models to the daily number of “true cases” (Y_d) and reported cases (C_d) on days $d=21$ to 40. Thus, we implicitly assumed that cases occurring over the first 20 days of the epidemic are unlikely to have been recognized. For an SEIR model, the mean generation interval is implicitly defined as the sum of the average latent period ($1/\nu$) and the average infectious period ($1/\gamma$) (16), and the proportion of the generation interval spent in the latent period is $f=1/(V\nu)$. We estimated 95% confidence intervals (CIs) for our estimate of R_0 by incorporating uncertainty in the estimated growth rate, but did not incorporate uncertainty in f or V .

Estimation of R_t from simulated data

We estimated R_t using the approach described by Cori et al (9), implemented using the EpiEstim software (17). We used a daily time step and a 7-day moving window. We assumed that the generation interval was gamma distributed (i.e. parametric) with a mean of 6.5 days and standard deviation of 4.0, consistent with data from Flaxman et al (18).

COVID-19 testing data for the United States

Daily data on the reported number of positive and negative tests for COVID-19 in the US and by state were downloaded from the COVID Tracking Project on August 31, 2020 (13). The COVID Tracking Project data come from state/district/territory public health authorities, and occasionally, from trusted news sources, official press conferences, or (rarely) social media updates from state public health authorities or governors [12]. We analyzed the data from March 4, 2020, onward, as this is the first date that negative tests for COVID-19 were consistently reported for the entire US. We also extracted information on COVID-19 testing criteria from each state health department's website during mid-March, mid-April, and mid-May. Data sources are documented in the Appendix dataset.

Estimation of R_0 from state-level testing data

To estimate R_0 from data on the daily number of reported positive COVID-19 cases in the US and different states, we fitted Poisson regression models to the first three weeks of data (March 4 through March 24, 2020) to estimate the growth rate. For states that did not report any cases early on, the Poisson regression models were fitted to data beginning the first day a positive case was reported through March 24, which is approximately one week after the national "15 Days to Slow the Spread" guidelines were announced (on March 16) (19). We assumed that the mean generation interval was 6.5 days and that the average latent period was 2.9 days (18,20,21). We calculated 95% CIs for the R_0 estimates based on uncertainty in the estimated growth rate, but did not account for uncertainty in the generation interval or latent period in order to highlight the differences associated with testing practices.

Estimation of R_t from state-level testing data

Estimates of the time-varying reproductive number in each state were generated using EpiEstim (9,17). We again used a daily time step and 7-day moving window, and assumed that the generation interval was gamma distributed with a mean of 6.5 days and standard deviation of 4.0 (18).

Analyses were implemented using MATLAB v9.3 (MathWorks, Natick, MA). Data and code are available from <https://github.com/vepitzer/COVIDtestingbias>.

RESULTS

Based on our simulations (Web Table 1, Web Figure 1), the likelihood and degree to which R_0 and R_t are biased depends on the manner in which diagnostic testing practices and reporting change over time. When the fraction of incident cases detected and reported is constant over time and testing capacity scales with the number of “true” cases, the number of confirmed positive cases provides an unbiased estimate of R_0 , despite possible delays in the reporting process (Web Figure 2). In this instance, the percent of individuals testing positive is expected to be stable over time. Estimates of R_t are also expected to be unbiased, but lag 2-4 days behind in detecting a decrease in R_t below the threshold value of 1 (i.e. when the epidemic is receding) (Figure 1A-D). If the fraction of true cases detected and reported is increasing or decreasing linearly over time, estimates of R_0 based on the growth rate of confirmed cases will be over- or under-estimated, respectively (Web Figure 2). The magnitude of the bias is positively correlated with the rate at which the fraction of true cases detected and reported is changing (Web Figure

3, Web Table 2). The time-varying reproductive number, R_t , will also be slightly over- or underestimated (Figure 1E-L). However, when p_{rep} is decreasing, the bias is partially offset by the reporting delay (Figure 1I-L). Importantly, the magnitude of the bias decreases over time, and estimates of R_t are still able to detect the impact of sustained interventions (Figure 1, Web Figure 4). A gradual increase or decrease in the percent of individuals testing positive ($\%_{pos}$) is a potential indicator of such bias. However, $\%_{pos}$ is also expected to decrease or increase over time if the testing capacity expands more or less quickly than the number of true cases, respectively (Figure 1M-T). In this instance, estimates of R_0 and R_t based on the number of confirmed cases are unbiased (Web Figure 2). Thus, contextual knowledge is needed to interpret whether a change in $\%_{pos}$ is indicative of a potential bias in estimates of the reproductive number.

Abrupt changes to testing criteria, affecting the fraction of true cases detected and reported, are also expected to bias estimates of R_0 and lead to a large but temporary bias in estimates of R_t predominantly in the days following the change (Figure 2A-H, Web Figure 5). The potential for such bias may be indicated by a sudden change in the percent of individuals testing positive, as well as an abrupt change to the daily incidence of reported cases. A similar change in $\%_{pos}$ may also occur with an abrupt change to the testing capacity (Figure 2I-P, Web Figure 5).

However, in this case, it is accompanied by a change to the daily number of individuals tested, and estimates of R_0 and R_t based on fitting to the number of positive cases are not expected to be biased (except for the reporting delay). The most difficult bias to detect may be due to a change in the reporting delay distribution (Figure 2Q-X). In this case, $\%_{pos}$ is likely to remain

roughly constant through time, but estimates of R_0 and R_t will be biased, especially when the reporting delay increases.

Our conclusions are robust to the inclusion of additional model complexity that allows for the reporting probability to be related to the severity of symptoms and the continued scale-up of testing capacity (Web Appendix 2, Web Tables 3-4, Web Figures 6-9). Temporal variation in $\%_{pos}$ may indicate the potential for bias in estimates of R_0 and R_t , but similar patterns can be due to changes in testing capacity that do not lead to bias. Furthermore, examining $\%_{pos}$ cannot detect bias due to changes in the reporting delay distribution (Figure 2Q-X), or due to concurrent changes in both the fraction of true cases detected and reported and the testing capacity (Web Figures 4-5).

Based on the number of confirmed COVID-19 cases across the US through March 24, 2020 (before any observable impact of social distancing measures, Figure 3), and assuming a fixed serial interval of 6.5 days, we estimate that R_0 for the US is 3.24 (95% confidence interval (CI): 3.21-3.26, accounting *only* for uncertainty in the growth rate). Nationally, the percent of individuals testing positive for COVID-19 increased from 10-15% in early March to around 20% in late-March (Figure 3). This early increase in the $\%_{pos}$ could be due to an increase in the fraction of true cases detected and reported or a decrease in testing capacity relative to “true” incidence; thus, our estimate of R_0 may be a slight overestimate. Estimates of R_0 for all 50 states and the District of Columbia vary from 1.53 (95% CI: 0.83, 2.44; Wyoming) to 5.03 (95% CI: 4.56, 5.50; Connecticut) (Web Table 5); trends in the $\%_{pos}$ vary by state (Figure 4).

Estimates of R_t for the US and individual states were generally high initially ($R_t > 3$) but decreased over time (Figures 3-4, Web Figure 10). This may be due to a low probability of detecting cases at the start of the epidemic in late February/early March. The $\%_{pos}$ increased gradually during March in a number of states (e.g. Arizona and New York), which could reflect a slight upward bias in estimates of R_t ; however, sharp decreases in the $\%_{pos}$ (e.g. in Arizona in mid-March) reflect increases in testing capacity, and therefore should not indicate a potential downward bias (Figure 4A,D,E,H).

Nationally and in most states, R_t hovered around 1 between early April and end of August, 2020, and increased above 1.5 in Arizona and Florida in June (Figures 3-4, Web Figure 10). Testing capacity increased steadily across all states through mid-July, leading to decreases in the $\%_{pos}$ in Michigan and New York (Figure 4K,L,O,P). While the $\%_{pos}$ increased in Arizona and Florida during June/July (Figure 4I,J,M,N), any bias in estimates of R_t should be minimal (see Web Appendix 2, Web Figure 9).

DISCUSSION

During the first three months of the COVID-19 epidemic in the US, testing practices varied dramatically over time and from state to state (22). Due to the limited availability of tests early in the epidemic, most states recommended that only those with a history of travel to affected countries or known contact with a confirmed case be tested (4,5). Once the disease became more widespread throughout the US and testing capacity increased, testing guidelines were

relaxed, but there were still considerable differences from state to state (Web Appendix 3). For example, as of April 15, 2020, Washington state had no restrictions on who could be tested for COVID-19, but prioritized hospitalized individuals and essential service providers exhibiting symptoms (23), whereas New York still recommended restricting testing to those with a known positive contact or travel history and/or symptomatic individuals who had tested negative for other infections (24). By May 15, most states had updated testing criteria to include anyone with symptoms and/or an association with a known COVID-19 case. As testing practices changed over time, we demonstrated that these changes may introduce bias into estimates of R_0 and R_t , affecting inference about how much control is needed and when control measures have reduced transmission below the critical threshold necessary to sustain the epidemic.

Monitoring the number of tests performed and the percent of tests that are positive over time can help to indicate the potential for bias in estimates of reproductive numbers. However, the reporting of test results, particularly for negative tests, has been inconsistent in many states. In California, for example, there was a more than eight-fold increase in the number of negative tests reported on March 13, and another four-fold increase on April 4, 2020 (13). These large increases in the number of negative tests were not accompanied by a corresponding increase in the number of confirmed cases. While estimates of R_0 and R_t based on the number of confirmed cases are not expected to be influenced by these abrupt changes in the number of reported tests, it becomes difficult to interpret the intervening gradual increases in the $\%_{pos}$. Moreover, we do not recommend using the $\%_{pos}$ to adjust for changes in testing practices; this will introduce bias when the change in $\%_{pos}$ results from a change in the number of tests (i.e.

the denominator). Instead, we advocate for reconstructing the under-ascertained case series based on reported deaths (for which the reporting fraction is higher and less variable over time (25)) accounting for the relevant delay distributions (26).

It is more difficult to detect whether the time between onset of infectiousness and the reporting of test results (i.e. the reporting delay) has changed over time. Our simulations suggest that such changes could bias estimates of R_0 and R_t , but would not be reflected in the percent of individuals testing positive over time. Individual-level data on the date of symptom onset, date of testing, and date of reporting are needed to resolve this potential bias. Estimates of R_t are also expected to lag behind true changes in the transmission rate due to reporting delays. Simple approaches to adjusting for the reporting delay, such as shifting the time series by the mean reporting delay or subtracting a sample of the delay distribution from each observation (i.e. convolution), fail to correct for the bias (12). Nowcasting approaches may be useful for resolving the bias by inferring the number of infections occurring on each day based on observed cases, hospitalizations and deaths, and known reporting delays (26–28).

We used a parsimonious transmission model and observation process in order to clearly demonstrate the bias that may result from changes to testing practices over time. Including additional details in the disease and observation process, such as differences in reporting of severe versus mildly symptomatic cases, does not qualitatively change our conclusions (see Appendix). However, we do not consider how additional complexities and heterogeneity in the transmission process may also relate to the probability that cases are detected and reported

over time. For example, older individuals may be both more susceptible to infection with severe acute respiratory syndrome coronavirus 2 and more likely to transmit to others, as well as being more likely to be tested for COVID-19 due to their increased risk of severe disease. Such heterogeneities are likely to affect both the true value of R_t as the disease spreads to different populations (e.g. nursing homes), as well as our ability to generate unbiased estimates of R_t . Further work is needed to explore the impact of such complexities in the transmission and reporting process.

Decisions regarding the lifting of stay-at-home orders and loosening of social distancing requirements, and when such measures may need to be reinstated, depend on having a good understanding of current levels of transmission. Reliable estimates of the reproductive number are essential for quantifying the impact of control measures on transmission and making informed decisions about future interventions, e.g. (7,18,29–31). However, changes in testing policies and practices, as well as delays in the reporting process, can lead to bias in estimates of the reproductive number, as we have demonstrated. It is important to carefully document and track such changes in testing and reporting practices in order to make correct inferences.

References

1. Cheng MP, Papenburg J, Desjardins M, et al. Diagnostic Testing for Severe Acute Respiratory Syndrome-Related Coronavirus 2: A Narrative Review. *Ann. Intern. Med.* 2020;172(11):726–734.
2. CDC Diagnostic Test for COVID-19. (<https://www.cdc.gov/coronavirus/2019-ncov/php/testing.html>). (Accessed April 20, 2020)
3. Kaplan S. C.D.C. Labs Were Contaminated, Delaying Coronavirus Testing, Officials Say. *New York Times*. 2020;(<https://www.nytimes.com/2020/04/18/health/cdc-coronavirus-lab-contamination-testing.html>) (Accessed April 18, 2020)
4. Washington Post Staff. What we know about delays in coronavirus testing. *Washington Post*. 2020;(<https://www.washingtonpost.com/investigations/2020/04/18/timeline-coronavirus-testing/?arc404=true>) (Accessed April 18, 2020)
5. Shear MD, Goodnough A, Kaplan S, et al. The Lost Month: How a Failure to Test Blinded the U.S. to COVID-19. *New York Times*. 2020;(<https://www.nytimes.com/2020/03/28/us/testing-coronavirus-pandemic.html>) (Accessed April 18, 2020)
6. Anderson RM, May RM. *Infectious Diseases of Humans*. New York: Oxford University Press; 1991.
7. Wallinga J, Teunis P. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *Am. J. Epidemiol.* 2004;160(6):509–516.
8. Cauchemez S, Boëlle PY, Donnelly CA, et al. Real-time estimates in early detection of SARS. *Emerg. Infect. Dis.* 2006;12(1):110–113.

9. Cori A, Ferguson NM, Fraser C, et al. A new framework and software to estimate time-varying reproduction numbers during epidemics. *Am. J. Epidemiol.* 2013;178(9):1505–1512.
10. Fraser C, Riley S, Anderson RM, et al. Factors that make an infectious disease outbreak controllable. *Proc. Natl. Acad. Sci. U. S. A.* 2004;101(16):6146–6151.
11. Lipsitch M, Cohen T, Cooper B, et al. Transmission dynamics and control of severe acute respiratory syndrome. *Science.* 2003;300(5627):1966–1970.
12. Gostic K, McGough L, Baskerville E, et al. Practical considerations for measuring the effective reproductive number, R_t . *PLoS Comput. Biol.* 2020;16(12):e1008409.
13. The COVID Tracking Project. (<https://covidtracking.com/data/api>). (Accessed August 31, 2020)
14. Li R, Pei S, Chen B, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). *Science.* 2020;368(6490):489–493.
15. Britton T, Tomba GS. Estimation in emerging epidemics: Biases and remedies. *J Roy Soc Interface.* 2019;16:20180670.
16. Wallinga J, Lipsitch M. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proc. R. Soc. B Biol. Sci.* 2007;274(1609):599–604.
17. Batista M. estimate_R. 2020; (https://www.mathworks.com/matlabcentral/fileexchange/78760-estimate_r), MATLAB Central File Exchange. Retrieved December 9, 2020.
18. Flaxman S, Mishra S, Gandy A, et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature.* 2020;584(7820):257–261.
19. 15 Days to Slow the Spread. *The White House.* (<https://www.whitehouse.gov/articles/15->

- days-slow-spread/). (Accessed April 15, 2020)
20. He X, Lau EHY, Wu P, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat. Med.* 2020;26(5):672–675.
 21. Bi Q, Wu Y, Mei S, et al. Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. *Lancet Infect. Dis.* 2020;20(8):911–919.
 22. Petri AE. Getting a Coronavirus Test: What is the Experience? *New York Times*. 2020;(https://www.nytimes.com/article/test-for-coronavirus.html) (Accessed July 15, 2020)
 23. Washington State Department of Health. *Interim COVID-19 Testing Guidance for Healthcare Providers*. Washington State Department of Health; 2020. (https://www.doh.wa.gov/Portals/1/Documents/1600/coronavirus/Interim-2019NovelCoronavirusQuicksheetProviders.pdf) (Accessed April 15, 2020)
 24. New York State Department of Health. *COVID-19 Testing*. Official website New York State; 2020. (https://coronavirus.health.ny.gov/covid-19-testing). (Accessed April 18, 2020)
 25. Woolf SH, Chapman DA, Sabo RT, et al. Excess Deaths from COVID-19 and Other Causes, March–July 2020. *JAMA - J. Am. Med. Assoc.* 2020;324(15):1562–1564.
 26. Russell TW, Golding N, Hellewell J, et al. Reconstructing the early global dynamics of under-ascertained COVID-19 cases and infections. *BMC Med.* 2020;18(1):332.
 27. McGough SF, Johansson MA, Lipsitch M, et al. Nowcasting by Bayesian Smoothing: A flexible, generalizable model for real-time epidemic tracking. *PLoS Comput. Biol.*

2020;16(4):e1007735.

28. Chitwood MH, Russi M, Gunasekera K, et al. Bayesian nowcasting with adjustment for delayed and incomplete reporting to estimate COVID-19 infections in the United States [preprint]. *medRxiv*. 2020. (<https://doi.org/10.1101/2020.06.17.20133983>) Accessed July 15, 2020.
29. Kucharski AJ, Russell TW, Diamond C, et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *Lancet Infect. Dis.* 2020;20(5):553-558.
30. Prem K, Liu Y, Russell TW, et al. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. *Lancet Public Heal.* 2020;5(5):e261-e270.
31. Zhang J, Litvinova M, Wang W, et al. Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei province, China: a descriptive and modelling study. *Lancet Infect. Dis.* 2020;20(7):793-802.

ORIGINAL UNEDITED MANUSCRIPT

Table 1. State transitions and rates for the stochastic simulation model.

Event	State transition	Rate ^{a,b}
New infection	Susceptible to Exposed	$\sum_{t=d+\Delta t}^{d+1} \beta I(t) \frac{S(t)}{N(t)} \Delta t$
Onset of infectiousness	Exposed to Infectious	$\sum_{t=d+\Delta t}^{d+1} \nu E(t) \Delta t$
Recovery from infectiousness	Infectious to Recovered	$\sum_{t=d+\Delta t}^{d+1} \gamma I(t) \Delta t$

^aModel parameters (β , ν , γ) are defined in Table 2. The variable d refers to the day in the simulation model, while Δt is the time-step of the simulation (=0.05 days).

^bThe total population size at time t in the model, $N(t)$, is equal to the sum of the Susceptible (S), Exposed (E), Infectious (I), and Recovered (R) states.

Table 2. Stochastic SEIR model parameters.

Parameter	Symbol	Value ^a	Reference
Transmission parameter	β	0.85 for $d < 50$ 0.5 for $d \geq 50$	Assumption (consistent with $R_0 \cong 3$ and $R_t \cong 1$)
Average duration of latent period	$1/\nu$	2.9 days	(14,20)
Average duration of infectious period	$1/\gamma$	3.7 days	(14,20)

R_0 , basic reproductive number; R_t , time-varying effective reproductive number

^aThe variable d refers to the day in the simulation model.

Figure Legends

Figure 1. Impact of gradual changes in testing practices on R_t estimation based on simulated data. The daily number of individuals tested (dashed line) and confirmed cases (solid line) per 10,000 people are plotted for days 20 to 70 of the simulated epidemic (first column), along with the percent of tests positive (second column) and the estimated time-varying reproductive number (R_t , third column) for the true cases (solid line) and confirmed cases (dashed line). The grey shaded regions represent the 95% confidence intervals (CI) around the R_t estimates. We quantified bias in the estimate of R_t as the difference between the upper or lower 95% CI of the R_t estimate for the observed cases minus the mean estimate for the true cases (black bars, fourth column). (A-D) Base case in which the fraction of true cases detected and reported and the reporting delay are constant over time and the testing capacity scales with the number of true cases. (E-H) The fraction of true cases detected increases from 5% to 25% between days 20 and 60. (I-L) The fraction of cases detected decreases from 25% to 5% between days 20 and 60. (M-P) The testing capacity increases from 0.2 individuals per case to 0.8 individuals per case between days 20 and 60. (Q-T) The testing capacity decreases from 0.8 individuals per case to 0.2 individuals per case between days 20 and 60.

Figure 2. Impact of abrupt changes in testing practices on R_t estimation based on simulated data. The daily number of individuals tested (dashed line) and confirmed cases (solid line) are plotted for days 20 to 70 of the simulated epidemic (first column), along with the percent of tests positive (second column) and the estimated time-varying reproductive number (R_t , third column) for the true cases (solid line) and confirmed cases (dashed line). The grey shaded

regions represent the 95% confidence intervals (CI) around the R_t estimates. We quantified bias in the estimate of R_t as the difference between the upper or lower 95% CI of the R_t estimate for the observed cases minus the mean estimate for the true cases (black bars, fourth column). (A-D) The fraction of true cases detected and reported increases from 5% to 25% on day 40. (E-H) The fraction of true cases detected and reported decreases from 25% to 5% on day 40. (I-L) The testing capacity increases from 0.2 individuals per case to 0.8 individuals per case on day 40. (M-P) The testing ratio decreases from 0.8 individuals per case to 0.2 individuals per case on day 40. (Q-T) The mean reporting delay increases from 6.6 days to 13.2 days on day 40. (U-X) The mean reporting delay decreases from 6.6 days to 3.3 days on day 40.

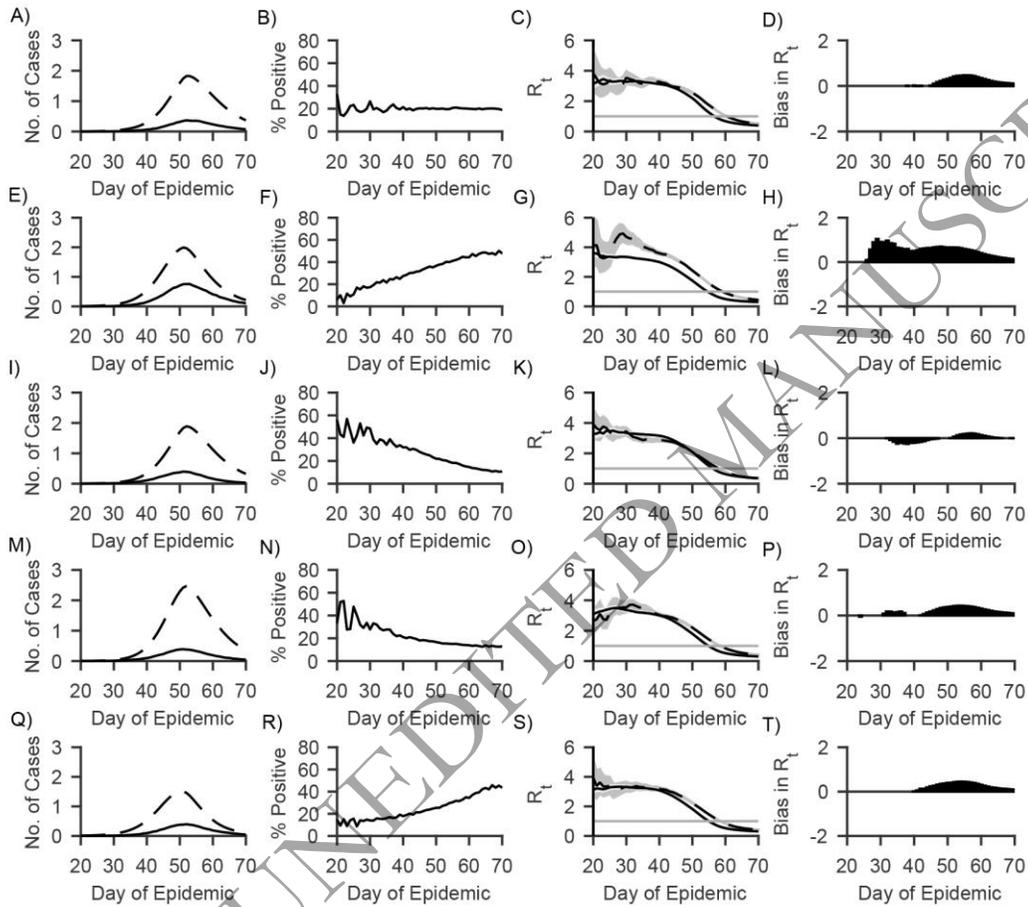
Figure 3. Reported number of COVID-19 cases and tests in the US and estimated time-varying reproductive number. (A) The daily number of individuals tested ($\times 10^5$, grey, left axis) and confirmed cases ($\times 10^4$, black, right axis) in the United States are plotted for March 4 to August 30, 2020. (B) The daily number of individuals tested (grey) and confirmed cases (black) for March 4 to April 1 are plotted on the \log_{10} scale. The dashed lines represent the fitted Poisson regression models, used to estimate R_0 from the growth rate, while the dotted vertical black line represents the March 24 cut-off date used. (C) The percent of tests positive through time is plotted for the daily data (grey) and the 15-day moving average of the daily number of cases and tests (black). (D) The estimated value of the time-varying reproductive number, R_t , is plotted for March 11 to August 30. The grey shaded region corresponds to the 95% confidence interval around estimates of R_t , while the dashed black lines represent $R_t=1$ (i.e. the threshold for epidemic growth).

Figure 4. Reported number of COVID-19 cases and tests and estimated time-varying

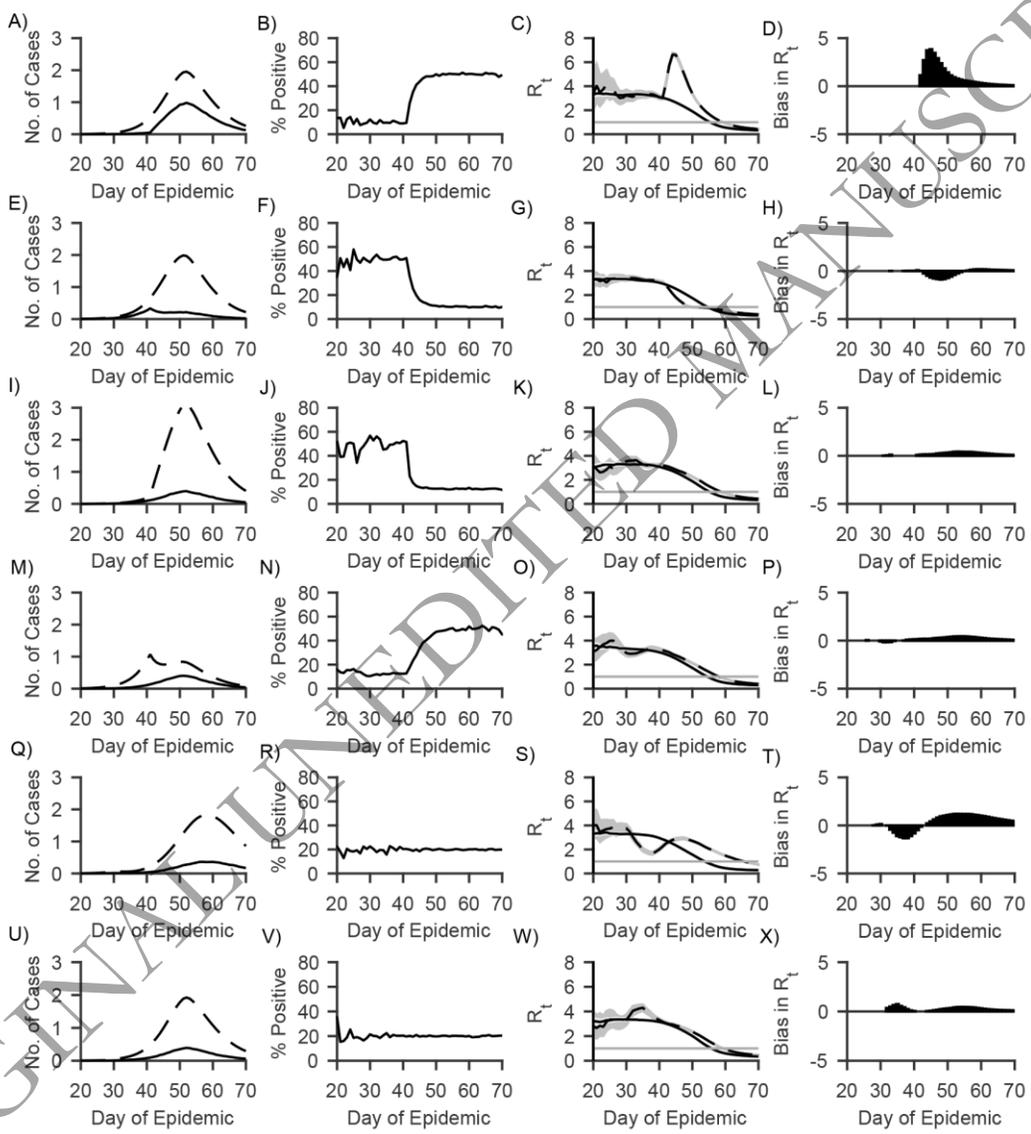
reproductive numbers for select US states. The daily number of confirmed cases ($\times 10^3$, black) and individuals tested ($\times 10^3$, grey) are plotted for (A-D) March 4 to April 1, and (I-L) April 1 to August 30, 2020, for four representative US states (Arizona: first column (A,E,I,M), Florida: second column (B,F,J,N), Michigan: third column (C,G,K,O), and New York: fourth column (D,H,L,P)). The percent of tests positive (dashed line) and the estimated value of the time-varying reproductive number (R_t , solid line) are plotted for (E-H) March 11 to April 1, and (M-P) April 1 to August 30, 2020. The grey shaded region represents the 95% confidence interval around estimates of R_t , while the grey horizontal line represents $R_t=1$ (i.e. the threshold for epidemic growth).

ORIGINAL UNEDITED MANUSCRIPT

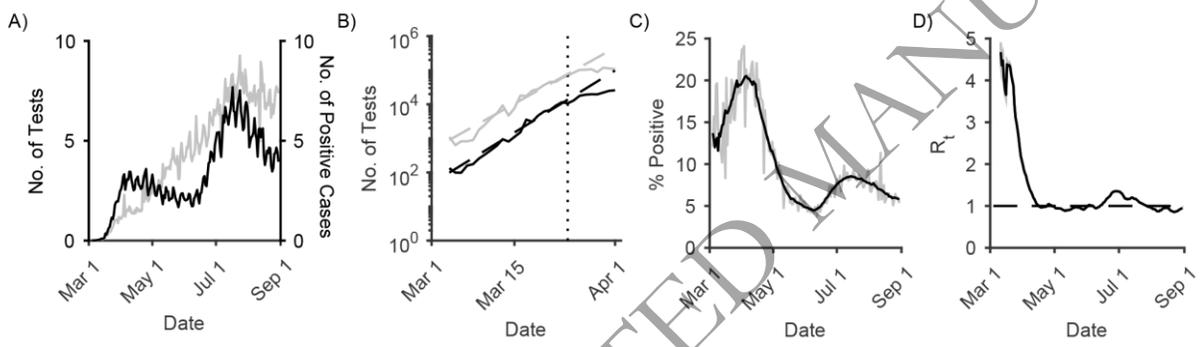
ORIGINAL UNEDITED MANUSCRIPT



ORIGINAL MANUSCRIPT



ORIGINAL UNEDITED MANUSCRIPT



ORIGINAL UNEDITED MANUSCRIPT

