*Disaster Medicine and Public Health Preparedness*

**www.cambridge.org/dmp**

# Correction in Active Cases Data of COVID-19 for the US States by Analytical Study

Ravi Solanki[1], Anubhav Varshney[2], Raveesh Gourishetty[3], Saniya Minase[4], Namitha Sivadas[5] and Ashutosh Mahajan[5]

[1]Centre for VLSI and Nanotechnology, Visvesvaraya National Institute of Technology, Nagpur, India; [2]Department of Electrical Engineering, Swami Vivekanand College of Engineering, Indore, India; [3]Department of Electrical Engineering, Indian Institute of Technology Bombay, Mumbai, India; [4]Department of Electrical and Electronics Engineering, Birla Institute of Technology and Science Pilani, Goa, India and [5]Centre for Nanotechnology Research, Vellore Institute of Technology, Vellore, India

**Brief Report**

**Cite this article:** Solanki R, Varshney A, Gourishetty R, Minase S, Sivadas N, Mahajan A (2022) Correction in active cases data of COVID-19 for the US states by analytical study. *Disaster Med Public Health Prep* **16**: 2335–2338. doi: https://doi.org/10.1017/dmp.2021.130.

First published online: 30 April 2021

**Keywords:**
COVID-19; epidemiology; mathematical model

**Corresponding author:**
Ashutosh Mahajan,
Email: ashutosh.mahajan@vit.ac.in.

## Abstract

The total coronavirus disease (COVID-19) cases caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection have reached 139 million worldwide and nearing 3 million deaths on April 16, 2021. The availability of accurate data is crucial as it makes it possible to analyze correctly the infection trends and make better forecasts. The reported recovered cases for many US states are surprisingly low. This could be due to difficulties in keeping track of recoveries, which resulted in higher numbers for the reported active cases than the actual numbers on the ground. In this work, based on the typical range of recovery rate for COVID-19, we estimate the active data from the total cases and death cases and bring out a correction for the data for all the US states reported on Worldometer.

## Introduction

The availability of accurate data of an epidemic is important as the data provide key insights on the disease spread and enable the authorities to take a decision on control measures. Worldometer is one of the very popular sources of the global coronavirus disease (COVID-19) data, and it is also trusted and used by many government bodies and agencies.[1] The available data for the COVID-19 cases can be used for the prediction and analysis of hospitalization and meeting the demands of health care facilities and setting up the critical care systems for the patients. The active cases represent the number of infected people, whether symptomatic or asymptomatic, detected through self-reporting or testing. This number is important for public health authorities to estimate the current status of the disease spread and can be calculated by subtracting death and recovered cases from the total confirmed cases.

## Method

A compartmental predictive mathematical model, SIPHERD, for COVID-19 dynamics was used where the recovery rate is a model parameter and is fixed by optimizing the model with the actual data.[2,3] The data for the total death and active cases for 364 days from March 4, 2020, were taken from Worldometer[1] and data found were close to total cases and death data from one study.[4] After running the SIPHERD model for the United States, as reported in another study,[2] the recovery rate of the active category was found to be 0.015 (corresponding to 66 days of mean recovery time), which is very low compared with other countries like Germany and India,[2,3] where the recovery rate was 0.065. The low recovery rate in the United States may be attributed to either incorrect reporting of active cases[5] or the testing of only serious cases and a longer recovery time in hospitals compared with quarantined with mild symptoms. Second, keeping the record of recoveries is difficult because some of the infected people are asked to quarantine, whereas only critical patients are hospitalized. Sometimes, the reporting of those recoveries is not accurate or incomplete. This has led to inconsistent data for active cases.

The number of mild cases is reported to be 81% in a Chinese study.[6] COVID-19 data reported from 49 states, the District of Columbia, and 3 US territories to the Centers for Disease Control and Prevention from February 12–March 16 show that 20.7 reported cases were severe and patients were hospitalized.[7] COVID-NET regions show this number to be 21.4% till April 4[8] and Institute for Health Metrics and Evaluation data from March 5–April 4 show this number to be 20.3%.[9] According to the World Health Organization, the recovery time for mild cases is 2 weeks and 3–6 weeks for severe cases. Considering 80% of mild cases, the recovery rate cannot be as low as it appears in the data for the states listed in column 2 of Table 1.

The correct estimation of the active cases can be done by subtracting the death and recovery cases, with the appropriate recovery rate, from the total cases. The Worldometer data for total

© Society for Disaster Medicine and Public Health, Inc. 2021. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Table 1.** The US states active cases data status

| Correct Data | Incorrect Data | Partially Correct Data |
|---|---|---|
| Texas | Maryland | Nebraska |
| Tennessee | Virginia | Indiana |
| Louisiana | Maine | California |
| Arizona | Hawaii | Oregon |
| North Carolina | Kentucky | Florida |
| Illinois | Minnesota | Idaho |
| Massachusetts | Rhode Island | New Mexico |
| Pennsylvania | South Carolina | Louisiana |
| Ohio | Washington | Alabama |
| Arkansas | | Michigan |
| Connecticut | | Alaska |
| Delaware | | Colorado |
| Iowa | | Missouri |
| Kansas | | New Jersey |
| Mississippi | | Georgia |
| Montana | | New York |
| North Dakota | | District of Columbia |
| New Hampshire | | |
| Oklahoma | | |
| South Dakota | | |
| Utah | | |
| Vermont | | |
| West Virginia | | |
| Wisconsin | | |
| Wyoming | | |

cases and death cases are assumed to be true as the testing for positive results and recording of diseased cases are done more stringently as compared to recovery counting. The active cases can be obtained by using the following differential equation:

$$\frac{dH(t)}{dt} = \frac{dT(t)}{dt} - \frac{dD(t)}{dt} - \sigma H(t - t_R)$$

where, $H$, $T$, and $D$ are the active, total, and death cases, respectively. The recovery from the "infected" category is defined by the 2 parameters: delay in recovery $t_R$ and the recovery rate $\sigma$. As these values are dependent upon the immune system of the community and the hospital facilities, it should not vary much within the United States. We have taken $t_R$ as 10 days and $\sigma$ as 0.048 (21 days of mean recovery time considering both mild and severe cases).

## Results and Discussion

The above delay differential equation is used for all the states in the United States, and we found 3 groups among the states according to the accuracy of the data. The active cases reported on Worldometer[1] for a few states show excellent agreement with our estimation of active cases. One example state for this group is Texas, as seen in Figure 1(a). There are few states in the second group that are largely not matching with the analytical estimation, indicating that reported active data are inaccurate. These states are listed in column 2 in Table 1 and 1 representative state is Virginia as seen in Figure 1(b), where the current active cases are reported to be 530 820 which should have been just around 31 237, according to our calculation. Interestingly, in the last group, there are some states for which the reported active cases follow the estimated active cases for some days; however, the trend of the curve changes and does not follow our estimation as represented in Indiana, shown in Figure 1(c). In Figure 1, the reported total and active cases with the estimated active cases for one of the states in all the 3 groups are shown, and the figures for the remaining states are given in the Supplementary Material.

The Network Systems Science and Advanced Computing (NSSAC) division of the Biocomplexity Institute and Initiative at the University of Virginia has created a visualization tool that presents a way of examining data curated by different data sources.[10] We compared the active cases data provided by NSSAC and found that this independent source of active data is in close agreement with the corrected active cases data. The recovery rate in individual states may vary to an extent of ±10% depending on the variation in the number of tests performed, fraction of mild and severe cases. However, we have taken a uniform value of the recovery rate as 0.048 for all the US states. The mortality rate can be calculated from the active cases data,[2] as shown:

$$\tau = \frac{DND(t)}{H(t - t_D)},$$

where, $DND$ are the daily new extinct cases, and $t_D$ is the delay associate with the extinct cases as explained in the mortality rate calculation.[2] In the initial phases of infection, many of the states show higher active cases than the true values that give a mortality rate lower than the actual rate. Since the hospital bed estimation, intensive care unit equipment requirement estimation depends on

https://doi.org/10.1017/dmp.2021.130 Published online by Cambridge University Press

https://doi.org/10.1017/dmp.2021.130 Published online by Cambridge University Press



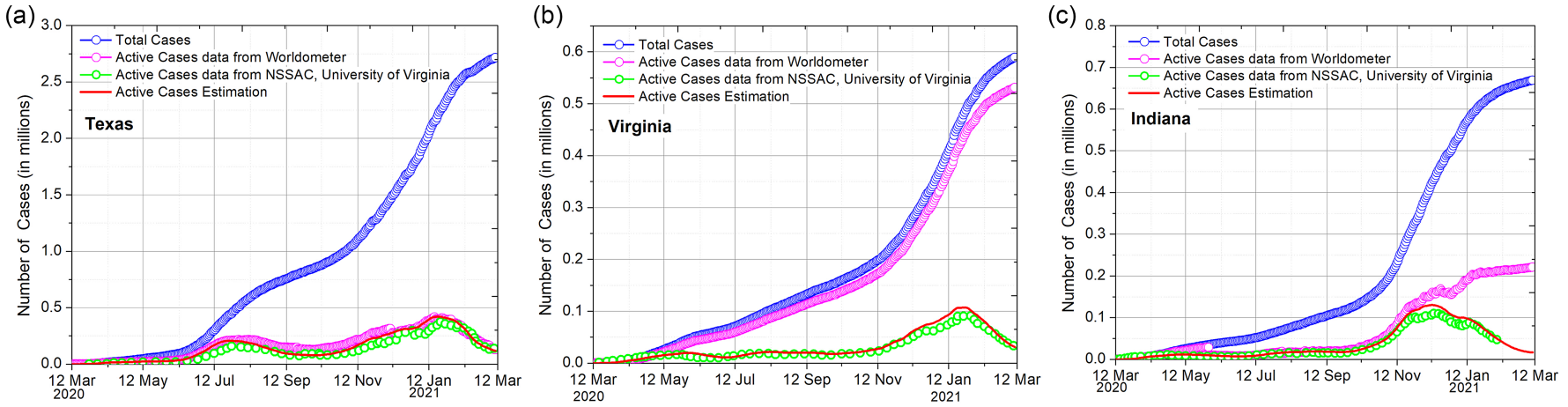



**Figure 1.** **(a)** Texas representing group states in which active data are reported correctly. **(b)** Virginia represents the second group in which data are largely incorrect. **(c)** Indiana represents the third group in which data are partially correct. The NSSAC, University of Virginia, data for the active cases are in close agreement with our analytical estimation.

the active cases currently and that are expected in the near future, and correction in the data facilitates better management of these entities. For the purpose of modeling and prediction, it is important that a mathematical model is validated against the data. Correct active cases data imply the right model parameters and a more accurate estimation of the hospital requirements.

## Conclusion

The reported active cases for a few states are consistent with the total detected cases and death cases for a recovery rate parameter value of 0.048. For a few states, the data have been corrected recently. However, for 9 states, the active cases data are still largely incorrect. We generate the corrected active case data for all states, report them in the Supplementary Material, and also keep the data available on GitHub.

**Supplementary material.** To view supplementary material for this article, please visit https://doi.org/10.1017/dmp.2021.130

**Data Availability Statement.** The data for the corrected active cases for all the US states can be downloaded using the GitHub link: https://github.com/ravisolankigithub/covid-activecases-usa.git.

**Conflict(s) of Interest.** The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this paper.

## References

1. **Worldometer**. COVID-19: Coronavirus pandemic. 2021. https://www.worldometers.info/coronavirus. Accessed March 13, 2021.
2. **Mahajan A, Solanki R, Sivadas N.** Estimation of undetected symptomatic and asymptomatic cases of COVID-19 infection and prediction of its spread in the USA. *J Med Virol.* 2021;93(5):3202-3210.
3. **Mahajan A, Sivadas NA, Solanki R.** An epidemic model SIPHERD and its application for prediction of the spread of COVID-19 infection in India. *Chaos Solitons Fractals.* 2020;140:110156.
4. The Atlantic. The COVID Tracking Project. The data. 2021. https://covidtracking.com/data. Accessed March 13, 2021.
5. **Smith-Schoenwalder C.** Why are U.S. coronavirus recovery numbers so low? *U.S. News and World Report*. April 2, 2020. https://www.usnews.com/news/health-news/articles/2020-04-02/why-are-us-coronavirus-recovery-numbers-so-low. Accessed September 5, 2020.
6. **Liu Z, Bing X, Zhi XZ.** Epidemiology Working Group for NCIP Epidemic Response, Chinese Center for Disease Control and Prevention. The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China. (In Chinese). *Chin J Epidemiol.* 2020;41(2):145-151.
7. **CDC COVID-19 Response Team**. Severe outcomes among patients with coronavirus disease 2019 (COVID-19) – United States, February 12–March 16, 2020. *MMWR Morb Mortal Wkly Rep.* 2020;69(12):343-346.
8. **Garg S.** Hospitalization rates and characteristics of patients hospitalized with laboratory-confirmed coronavirus disease 2019 – COVID-NET, 14 states, March 1–30, 2020. *MMWR Morb Mortal Wkly Rep.* 2020;69(15):458-464.
9. **Institute for Health Metrics and Evaluation (IHME)**. United States of America. Cumulative deaths. 2021. https://covid19.healthdata.org/united-states-of-america. Accessed September 5, 2020.
10. **Peddireddy AS, Xie D, Patil P, *et al.*** From 5Vs to 6Cs: operationalizing epidemic data management with COVID-19 surveillance. In: 2020 IEEE International Conference on Big Data (Big Data). IEEE. 2020:1380-1387. https://doi.org/10.1109/BigData50022.2020.9378435.