

Brief Communications

An interactive online dashboard for tracking COVID-19 in U.S. counties, cities, and states in real time

Benjamin D. Wissel ^{1,2*}, P.J. Van Camp^{1,2}, Michal Kouril^{1,2,3}, Chad Weis¹, Tracy A. Glauser^{1,3,4}, Peter S. White², Isaac S. Kohane⁵, and Judith W. Dexheimer^{1,2,3,6}

¹Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA, ²Department of Biomedical Informatics, University of Cincinnati College of Medicine, Cincinnati, Ohio, USA, ³Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, Ohio, USA, ⁴Division of Neurology, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA, ⁵Department of Biomedical Informatics, Harvard Medical School, Boston, Massachusetts, USA, and ⁶Division of Emergency Medicine, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA

*Corresponding Author: Benjamin Wissel, BS, Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, MLC 2008, 3333 Burnet Avenue, Cincinnati, OH 45229-3039, USA; Benjamin.Wissel@cchmc.org

Received 14 April 2020; Revised 16 April 2020; Editorial Decision 20 April 2020; Accepted 20 April 2020

ABSTRACT

Objective: The study sought to create an online resource that informs the public of coronavirus disease 2019 (COVID-19) outbreaks in their area.

Materials and Methods: This R Shiny application aggregates data from multiple resources that track COVID-19 and visualizes them through an interactive, online dashboard.

Results: The Web resource, called the COVID-19 Watcher, can be accessed online (<https://covid19watcher.research.cchmc.org/>). It displays COVID-19 data from every county and 188 metropolitan areas in the United States. Features include rankings of the worst-affected areas and auto-generating plots that depict temporal changes in testing capacity, cases, and deaths.

Discussion: The Centers for Disease Control and Prevention does not publish COVID-19 data for local municipalities, so it is critical that academic resources fill this void so the public can stay informed. The data used have limitations and likely underestimate the scale of the outbreak.

Conclusions: The COVID-19 Watcher can provide the public with real-time updates of outbreaks in their area.

Key words COVID-19, health informatics, data visualization

INTRODUCTION

As of April 13, 2020, the United States had 30% of novel coronavirus disease 2019 (COVID-19) cases worldwide, the most of any country.¹ At this date, New York City was the epicenter of cases in the United States, but large outbreaks were present in several other major metropolitan areas, including New Orleans, Detroit, Chicago, and Boston.

Several online tools track COVID-19 outbreaks at the county, state, and national levels.^{1–4} However, it has become apparent that tracking outbreaks at the city level is critical, as the outbreak in

China was centered within and surrounding the city of Wuhan, in Italy around Lombardy, in Spain around Madrid, and in the United Kingdom around London.

Our team developed a methodology to aggregate county-level COVID-19 data into metropolitan areas and display these data in an interactive dashboard that updates in real time. The purpose of this website was to make this information more accessible to the public, and to allow for more granular assessment of infection spread and impact.

MATERIALS AND METHODS

We assessed 3 publicly available datasets that are updated daily and include county- or state-level counts of COVID-19 confirmed cases and deaths in the United States.

The New York Times COVID-19 data

The New York Times (NYT) began tracking COVID-19 cases and deaths on the county level in January 2020, and on March 26 they released their data to the public.⁵ The NYT defined cases as individuals who tested positive for COVID-19. Cases were attributed to the county in which the person was treated and were counted on the date that the case was announced to the public. If it was not possible to attribute a case to a specific county, then it was still counted for the state in which they were treated.

Johns Hopkins University COVID-19 data

The Johns Hopkins University Center for Systems Science and Engineering was the first group aggregate COVID-19 data and release it to the public in an accessible and sizable manner.¹ This group publishes total cases, recovered cases, and deaths at the national, state, and as of March 23, county levels.⁶

COVID Tracking Project data

The COVID Tracking Project is a grassroots effort incubated by *The Atlantic* that tracks COVID-19 testing in U.S. states.⁷ This group releases daily updates for the number of positive tests, negative tests, pending tests, hospitalizations, number of patients in the intensive care unit, and deaths. Because there is a high amount of variability in state reporting, some of these data are not available for every state.

Comparing COVID-19 data sources

These 3 data resources use different strategies to aggregate COVID-19 data from multiple sources. Because a gold standard has not been established, we compared the consistency of these sources with the Centers for Disease Control and Prevention (CDC).⁸ The CDC only releases data for confirmed cases for the entire country, so that was the only metric that could be compared among all 4 sources. All 50 states, the District of Columbia, and 5 U.S. territories were included.

Metropolitan area definitions

We used the U.S. Census Bureau's lists of counties comprising major metropolitan areas⁹ to aggregate counties into the 172 combined statistical areas and 16 additional core-based statistical areas: Tuscaloosa, AL; Fayetteville-Springdale-Rogers, AR; San Diego-Chula Vista-Carlsbad, CA; Colorado Springs, CO; Tallahassee, FL; Tampa-St. Petersburg-Clearwater, FL; Champaign-Urbana, IL; Topeka, KS; Baton Rouge, LA; Lansing-East Lansing, MI; Charleston-North Charleston, SC; College Station-Bryan, TX; Austin-Round Rock-Georgetown, TX; Waco, TX; Charlottesville, VA; and Richmond, VA.

Adjusting for population

To track the proportion of each area's residents that became infected or died of COVID-19, we used the U.S. Census Bureau's 2019 population estimate for each county to normalize data to tests, cases, and deaths per 10 000 residents.¹⁰

Code

The application, referred to as the COVID-19 Watcher, checks for data updates from the NYT and COVID Tracking Project every hour. When data updates are released, they are automatically downloaded onto the server and incorporated into the web resource. New data must pass a quality control check that ensures that updated data files are the anticipated size and format.

Data visualizations are generated using the ggplot2 package¹¹ in R statistical software version 3.6.1 (R Foundation for Statistical Computing, Vienna, Austria), and the application was developed using R Shiny.¹² The Web resource is hosted in an Amazon Web Services environment behind a scalable load balancer to accommodate user load. The source code was placed in a public GitHub repository and can be accessed online (<https://github.com/wisselbd/COVID-Tracker>). The site is maintained by the Cincinnati Children's Hospital Medical Center Division of Biomedical Informatics.

RESULTS

The COVID-19 Watcher dashboard can be accessed online (<https://covid19watcher.research.cchmc.org/>). The resource includes all U.S. counties, as well as 188 metropolitan areas that are collectively inhabited by over 277 million Americans (83.3% of the population).

A screenshot of the web resource is shown in [Figure 1](#). Users can view COVID-19 cases and deaths from the NYT at the county, city, state, or national level, and the total number of tests reported by the COVID Tracking Project, including the breakdown between positive and negative tests, is shown for each state. Multiple areas can be selected at once and plots auto-generate after each selection. Options include normalizing counts by population size, linear and logarithmic axes, and a button to download a screenshot of the plots. Users can search tables that display rankings of the least and most affected areas.

A summary of the COVID-19 data sources is shown in [Table 1](#). Data are updated at the end of each day in all cases except for the NYT, where they are released the following day. The NYT, Johns Hopkins, and the COVID Tracking Project provide easy-to-access download portals, while the CDC only provides a dashboard without an option to download the data.

A comparison of confirmed cases reported in each data source is shown in [Figure 2](#). The sources were highly consistent at the national level.

DISCUSSION

In the absence of a uniform government standard for tracking COVID-19 outbreaks in the United States, academic and newsgroup-based data repositories have become the de facto standard. While these datasets are publicly available, they require informatics and data visualization to extract and display information because of their complexity and continual updates. Visualizing COVID-19 data in real time through online dashboards is a pragmatic way to meet the medical community's demand for up-to-date information.

The data displayed by the COVID-19 Watcher can be used to evaluate the effectiveness of mitigation efforts. Normalizing data by an area's population shows the relative proportion of the population that have been infected. The logarithmic scale shows the rate of spread, and flattening the exponential curve indicates the spread of the virus is slowing. Users should take caution in using these data to

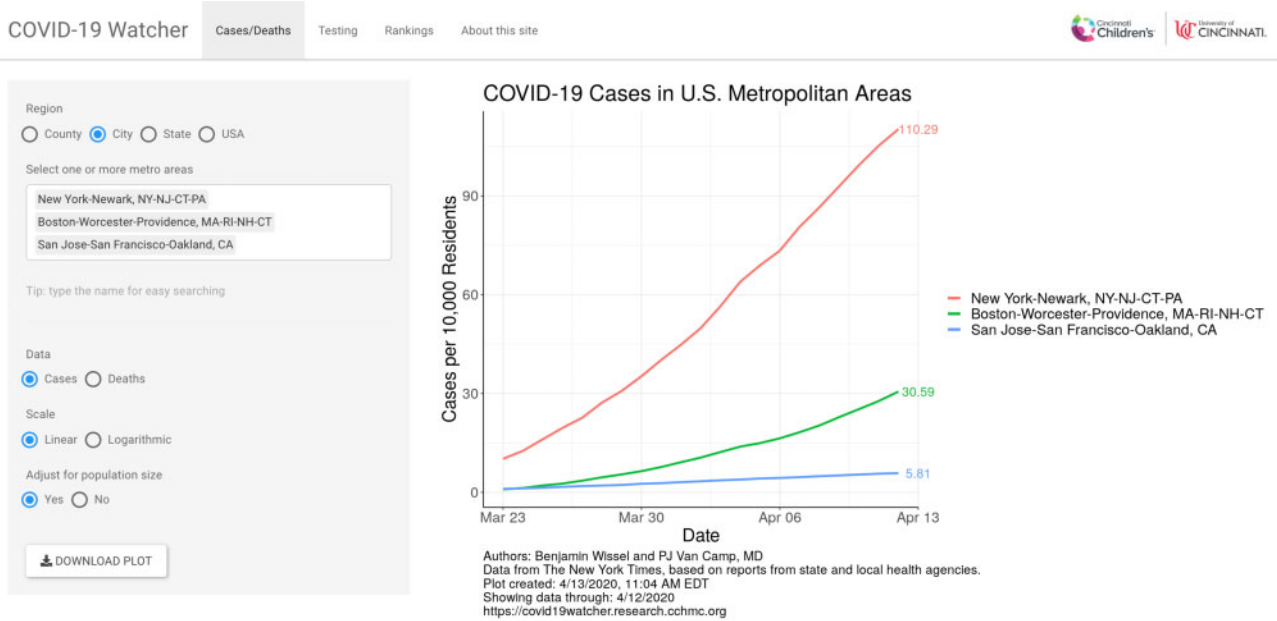


Figure 1. Screenshot of the COVID-19 Watcher web resource. Users can view data from *The New York Times* at the county, city, state, or national level. Multiple areas can be compared at once. Plots for the selected regions automatically generate and have options to view on logarithmic scale or normalize data by the population size. COVID-19: coronavirus disease 2019.

Table 1. Summary of publicly available data sources for tracking COVID-19 in the United States

Dataset	Open access	Frequency of updates	Timing of release ^a	Sources of data	Granularity of region	Data reported
CDC ⁸	No option to download data	Daily	End of same day	Case report forms submitted by state and local health departments	Nation	Cases
COVID Tracking Project ⁷	Yes	Daily	End of same day	News and public health authorities	States	Cases, deaths, hospitalizations, total tests, recovered, number in ICU ^c
<i>The New York Times</i> ⁵	Yes	Daily	Middle of next day	News and public health authorities	Counties	Cases and deaths
Johns Hopkins ⁶	Yes	Daily	End of same day	CDC and public health authorities	Counties ^b	Cases, deaths, and recoveries

As of April 15, 2020, the COVID-19 Watcher displays data from *The New York Times* and the COVID Tracking Project.

CDC: Centers for Disease Control and Prevention; COVID-19: coronavirus disease 2019; ICU: intensive care unit.

^aTiming of release is relative to Eastern Standard Time.

^bJohns Hopkins began publishing county-level data on March 23, 2020. Data from before then were reported at the state level.

^cData for the number of patients hospitalized, total number of tests, number of patients recovered, and number of patients in the ICU are sparse because many states do not report these data.

forecast future events. To make projections, these data should be used in conjunction with the University of Washington Institute for Health Metrics and Evaluation model,¹³ the University of Pennsylvania's COVID-19 Hospital Impact Model for Epidemics model,¹⁴ or other susceptible-infected-recovered models.

The authors welcome community feedback, ideas for further development, and contributions. The GitHub repository has a section for issue tracking where users can submit comments about the Web resource.¹⁵ Alternatively, contributors can make improvements to the code itself by forking the repository, modifying their copy of the code, and submitting pull requests back to the authors. These modifications will be reviewed and, if judged to be suitable, merged into

the main code. In particular, we would like to see community contributions related to geo-personalization of the website visualization, various analytics modeling, data points such as addition of countries, and timeline augmentation.

Although these datasets reviewed in Table 1 are the best that are available, they have major limitations. The procedures for reporting COVID-19 data need to be standardized. Current practices for aggregating data generally involve combining government reported data with unofficial, but reputable, media releases from public officials. Despite the differences in each source's approach, case counts were relatively similar to one another, indicating that data sources appear to reliably report available data.

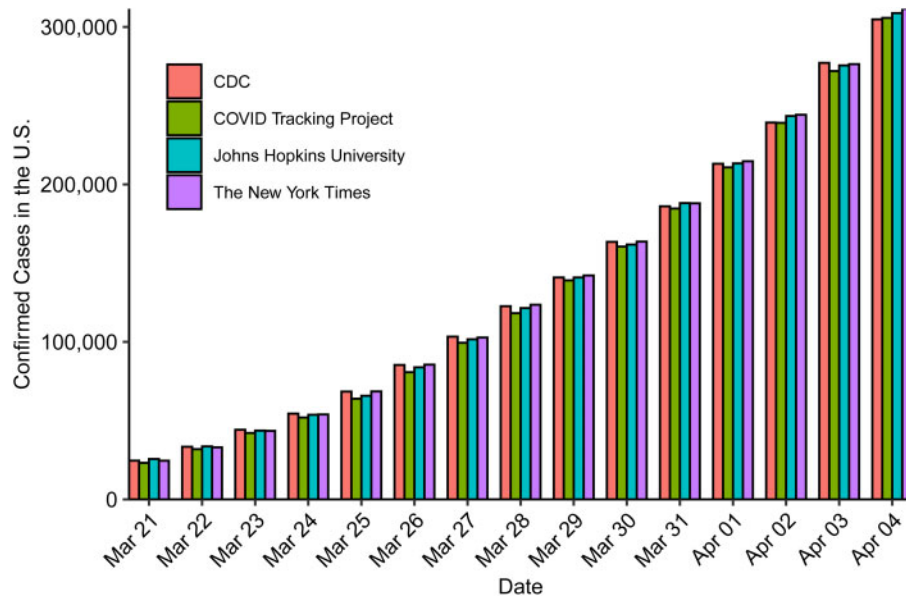


Figure 2. Comparison of coronavirus disease 2019 cases in the United States from March 21 to April 4, 2020, as reported by 4 datasets: *The New York Times*, Johns Hopkins, COVID Tracking Project, and the Centers for Disease Control and Prevention (CDC). Case numbers include U.S. territories.

However, counts for confirmed cases and deaths are likely to be underestimates because testing is limited. There is high interstate variability in the volume of testing, timeliness of results, and disclosure of the number of negative test results. States with the worst outbreaks, such as New York and Louisiana, also had the most tests per capita. There is a clear correlation between the number of tests completed and the number of confirmed cases reported. As of April 13, >40% of tests in New York came back positive, indicating that more testing is needed to understand the full scope of the outbreak.

In conclusion, we developed the COVID-19 Watcher to communicate up-to-date COVID-19 information to the medical community and general public. The Web application's pipeline was developed to be extendable, and additional data sources will be added as they become available. We hope that by making the code used by this Web resource available to the public, developers will submit ideas for improvement. Because it is possible that public data releases will be interrupted in the future, we recommend that the CDC immediately begin public releases of their entire COVID-19 data so academia can drive further innovation.

AUTHOR CONTRIBUTIONS

>All authors satisfied International Committee of Medical Journal Editors' authorship policy. BDW conceptualized the original idea to track COVID-19 data by metropolitan area and wrote the first draft of the manuscript. BDW and PJVC developed the COVID-19 Watcher application. All authors provided feedback on the application's design, submitted feedback on the manuscript for intellectual content, and approved the final version. BDW and JWD have full access to the data and source code and take responsibility for the integrity and accuracy of the report.

ACKNOWLEDGMENTS

These tools could not have been developed without many individual and selfless efforts to create resources for the public good. Special thanks to Danny T.Y. Wu, PhD, and Sander Su for their help launching the site.

CONFLICT OF INTEREST STATEMENT

None declared.

REFERENCES

- Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis* 2020; 20 (5): 533–4.
- Fagen-Ulmschneider W. An interactive visualization of the exponential spread of COVID-19. <https://91-divoc.com/pages/covid-visualization/> Accessed April 5, 2020.
- Katz J, Quealy K, Sanger-Katz M. Coronavirus in the U.S.: how fast it's growing. *The New York Times*. <https://www.nytimes.com/interactive/2020/04/03/upshot/coronavirus-metro-area-tracker.html> Accessed April 5, 2020.
- Arneson D, Bleicher P, Butte A, et al. COVID-19 County Tracker. <https://covidcounties.org/> Accessed April 5, 2020.
- Smith M, Yourish K, Almkhatar S, Collins K, Ivory D, Harmon A. An ongoing repository of data on coronavirus cases and deaths in the U.S. <https://github.com/nytimes/covid-19-data> Accessed April 5, 2020.
- 2019 Novel Coronavirus COVID-19 (2019-nCoV) Data Repository by Johns Hopkins CSSE. <https://github.com/CSSEGISandData/COVID-19> Accessed April 5, 2020.
- Miller K, Curry K. The COVID tracking project. <https://github.com/COVID19Tracking> Accessed April 5, 2020.
- Centers for Disease Control and Prevention. Coronavirus disease 2019. Cumulative total number of COVID-19 cases in the United States by report date, January 12, 2020 to April 4, 2020. <https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/cases-in-us.html#cumulative> Accessed April 5, 2020.
- U.S. Census Bureau. Core based statistical areas (CBSAs), metropolitan divisions, and combined statistical areas (CSAs). <https://www.census.gov/geographies/reference-files/time-series/demo/metro-micro/delineation-files.html> Accessed March 29, 2020.
- U.S. Census Bureau. County Population Totals: 2010-2019. https://www.census.gov/data/datasets/time-series/demo/popest/2010s-counties-total.html#par_textimage_70769902 Accessed March 30, 2020.
- Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer; 2016.

12. Chang W, Cheng J, Allaire J, Xie Y, McPherson J. Shiny: web application framework for R. *R Package Version* 2017; 1 (5).
13. IHME COVID-19 health service utilization forecasting team, Murray CJL. Forecasting COVID-19 impact on hospital bed-days, ICU-days, ventilator-days and deaths by US state in the next 4 months. *medRxiv* 2020 Mar 30 [E-pub ahead of print].
14. Weissman GE, Crane-Droesch A, Chivers C, *et al.* Locally informed simulation to predict hospital capacity needs during the COVID-19 pandemic. *Ann Intern Med* 2020 Apr 7 [E-pub ahead of print].
15. Wissel BD, Camp PV. COVID-19 Watcher. <https://github.com/wisselbd/COVID-Tracker> Accessed April 5, 2020.