



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Contents lists available at ScienceDirect

Annals of Epidemiology

journal homepage: www.annalsofepidemiology.org

Rapid report

Rapid report on using data to make standardized decisions about enrollment during the COVID-19 pandemic: perspectives from the MoTrPAC study



Stephanie M. George, PhD^{a,*}, Haiying Chen, MD, PhD^b, Michael E. Miller, PhD^b, W. Jack Rejeski, PhD^c, Cynthia L. Stowe, MPM^b, Christopher Webb, MPP^b, William E. Kraus, MD^d, Nicolas Musi, MD^e, John M. Jakicic, PhD^f

^a Division of Extramural Research, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, MD

^b Department of Biostatistics and Data Science, Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC

^c Department of Health and Exercise Science, Wake Forest University, Winston-Salem, NC

^d Duke Molecular Physiology Institute and Division of Cardiology, Duke University School of Medicine, Durham, NC

^e Barshop Institute for Longevity and Aging Studies, University of Texas Health San Antonio, San Antonio, TX

^f University of Pittsburgh, Healthy Lifestyle Institute, Pittsburgh, PA

ARTICLE INFO

Article history:

Received 14 January 2021

Revised 22 March 2021

Accepted 18 April 2021

Available online 3 May 2021

Keywords:

Epidemiology

COVID-19

Incidence

Data

Introduction

The emergence and persistence of Coronavirus Disease 2019 (COVID-19) in the United States has presented challenges for ongoing, large clinical and observational research studies that involve in-person components and measurements. To facilitate continuous operations, a need has arisen to develop systems to use COVID-19 incidence data to make standardized decisions about starting and restarting enrollment across multiple clinical sites. However, inci-

dence rates based on raw county-level data can be dramatically higher than actual values. We report here on an approach to 1) centrally monitor COVID-19 incidence rates using available, rolling epidemiological data, 2) identify solutions for addressing data limitations during monitoring when data are available from multiple imperfect data sources each with important differences in methodology, scope, and quality, and 3) use the results for informing temporary suspension or restarting of individual clinic activities.

Methods

The Molecular Transducers of Physical Activity Consortium (MoTrPAC), the largest investment in physical activity research to date by the National Institutes of Health, was established to elucidate how exercise improves health and ameliorates diseases by building a map of the molecular responses to acute and chronic exercise [1]. Suspension or restarting of MoTrPAC recruitment and in-person clinic activities is based on thresholds of 14-day county-level incidence rates. When the 14-day average incidence rate exceeds 10 cases per 100,000 population, COVID-19 testing is re-

Abbreviations: MoTrPAC, Molecular Transducers of Physical Activity Consortium; COVID-19, Coronavirus Disease 2019; ARIMA, autoregressive integrated moving-average; AO, additive outliers.

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

* Corresponding author. Division of Extramural Research, National Institute of Arthritis, Musculoskeletal and Skin Diseases, 6701 Democracy Boulevard, Suite 800, Bethesda, MD 20892 Tel.: 301-594-4974.

E-mail address: Stephanie.george@nih.gov (S.M. George).

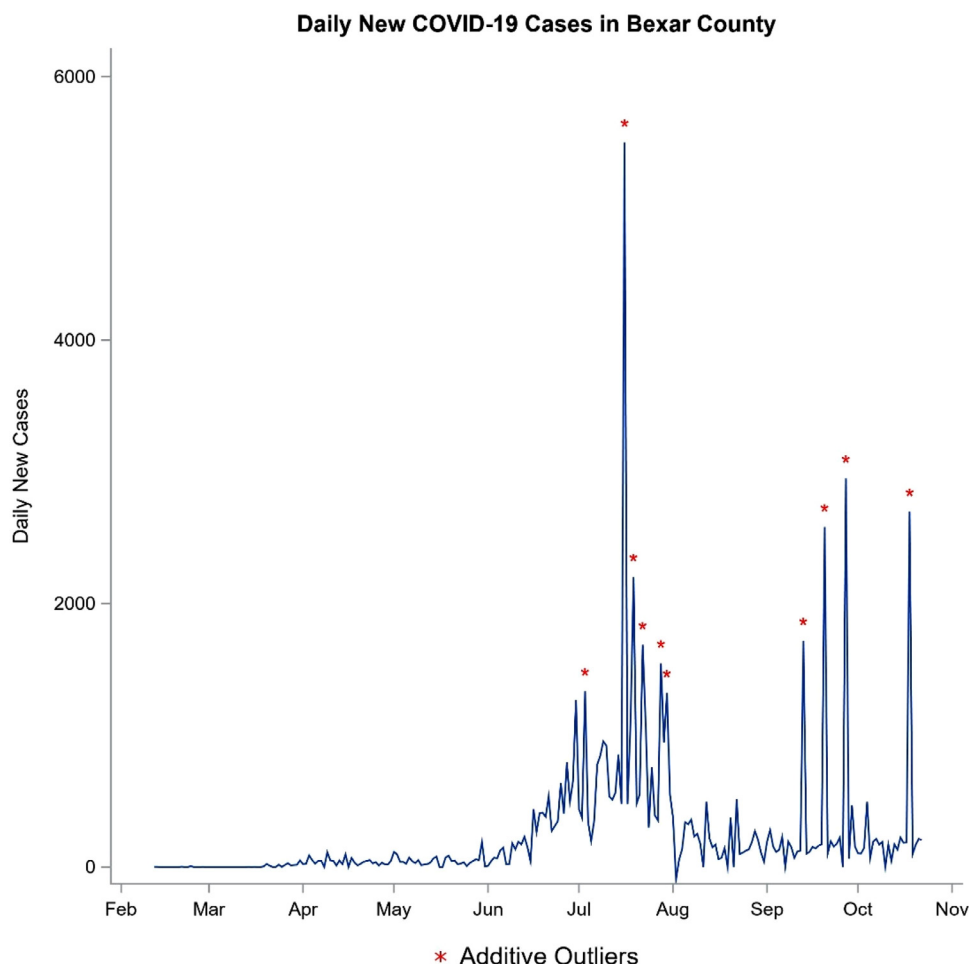


Figure 1. Daily new COVID cases in Bexar County.

quired before all in-person study visits; when incidence rates meet or exceed 30 cases per 100,000 population, new recruitment is halted. We developed dynamic reports utilizing a publicly available dataset updated daily and released by the New York Times through GitHub ([link](#)) to rapidly monitor the daily incidence rates of new COVID-19 cases for counties with potential participants that may be recruited by each clinical site. Depending on the clinical site and state, the targeted recruitment catchment areas ranged from 1 to 8 counties. During the process, it became evident that there were periodic data dumps due to electronic lab reporting on backlogs of old cases. Thus, we adopted an autoregressive integrated moving-average (ARIMA) modeling approach for active monitoring of the new COVID-19 cases. ARIMA modeling has been widely used for analyzing univariate time series [2]. It can identify additive outliers (AO) in the response series that are not accounted for by the estimated model.

Briefly, each Friday after the New York Times updates their data, the process involves sending to each site county-specific reports that contain both the daily new cases with AOs flagged and corresponding 14-day averages for incidence rates. Depending on where the 14-day averages fall ($<10/100k$, $\geq 10/100k$ and $<30/100k$, or $\geq 30/100k$), the sites may choose not to investigate the potential outliers and just follow usual rules for clinical operations, or the sites may choose to investigate the reason for the spikes with the local health department. If spikes are found to be related to backlog of cases, evidence of data dumps with written justification is sent back to the Data Coordinating Center. If some of the cases in

the dump occurred within the recent 14-day span (i.e., sometimes sites can only ascertain that the dump contained cases from a date range), all of the cases are included in the 14-day average. Only if all of the cases in the data dump are at least 2 weeks old, they are removed from the 14-day average and the report for the clinical site is revised. Based on the new 14-day average incidence rate (over same time period), the sites then follow the usual rules for clinical operations and determine if their operations for the upcoming week need to change from the previous week. The study protocol was changed to incorporate the above procedure.

In this paper, we use the report for Friday 10/23 for the clinic site in Bexar County, Texas as an example to illustrate this process.

Results

Figure 1 shows the time series of reported daily new cases for Bexar County/San Antonio through Thursday, October 22, 2020. For illustrative purposes, 10 most statistically significant AOs across the series identified by an ARIMA (0,1,1) model were indicated as * on the graph. We focused on the outliers from October 9 to October 22, the 14-day data that were used to generate the current rate estimate to guide clinical operations for the upcoming week. The site reviewed the outliers and verified with the Bexar County/San Antonio Metropolitan Health District that a backlog of 2583 cases was added to Bexar's cumulative case count on October 18 in addition to the cases that were identified on that date. Further investigation revealed that all the backlog cases occurred more than 14

Impact of Excluding Backlogged Cases on COVID-19 Rates in Bexar County

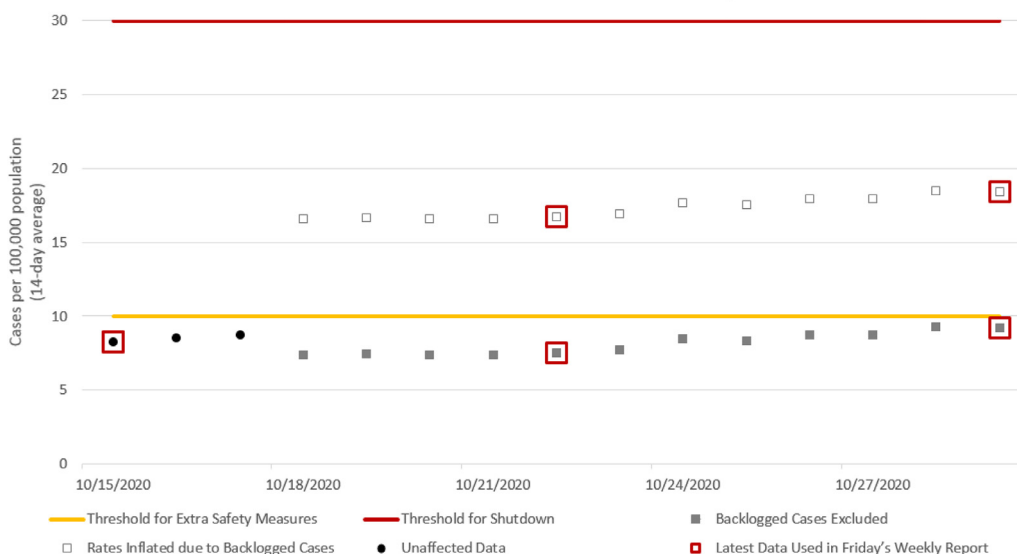


Figure 2. Impact of backlogged cases on COVID-19 rates in Bexar County.

days prior, thus they were excluded from the 14-day average and the rate was revised to 7.5 cases per 100,000 population. Figure 2 shows that if this correction were not made, the 14-day average would be skewed to 16.7 cases per 100,000 population and the clinical operation for the upcoming week would be impacted due to the crossing of the threshold of 10 cases per 100,000 population. This would also affect the subsequent Friday report on 10/30.

Discussion

The COVID-19 pandemic has brought unforeseen challenges to conducting in-person research, and there are still many unknowns ahead. The method we have applied can aid consortia in systematically identifying outliers due to delays in uploading of COVID-19 tests, not actual evidence of increase in community spread. A larger backlog than the one observed in Bexar County, or a similar backlog in a county with a higher rate, could raise the rate enough to bring recruitment from that county to a halt in our study. Our weekly centralized reporting on the data allow for clinical sites to be more aware of the rise and fall in rates and allow them to follow the prespecified procedures related to the ongoing recruitment and enrollment of individuals to the study. We view this process as an exploratory aid to help the clinical sites manage operations. During a pandemic, in-person research necessitates comprehensive public health mitigation measures, including use of personal protective equipment, to keep participants and staff safe. Our work highlights the additional value of building novel analytic approaches for monitoring and understanding community spread during current and future infectious disease outbreaks.

Author statement

All authors contributed equally. SMG: conceptualization, methodology, investigation, writing-original draft, writing-review & editing HC, MEM, CLS, and CW: conceptualization, data curation, formal analysis, methodology, investigation, visualization, writing-review & editing, verified underlying data. WJR, BK, NM, JJ: conceptualization, methodology, investigation, writing-review & editing. All authors approved the final version.

Acknowledgments

This work from the MoTrPAC Study is supported by NIH grants, U01AR071133, U01AR071130, U01AR071124-01, U01AR071128, U01AR071150, U01AR071160, U01AR071158 (Clinical Centers), U24AR071113 (Consortium Coordinating Center).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.annepidem.2021.04.016](https://doi.org/10.1016/j.annepidem.2021.04.016).

References

- [1] Sanford JA, Nogiec CD, Lindholm ME, et al. Molecular Transducers of Physical Activity Consortium (MoTrPAC): mapping the dynamic responses to exercise. *Cell* 2020;181(7):1464–74.
- [2] Box GEP, Jenkins GM. *Time series analysis: forecasting and control*, San Francisco: Holden Day; 1976. Rev. ed. ed.