
Brief Communications

Misalignment between COVID-19 hotspots and clinical trial sites

Lauren Franks^{1,*}, Hao Liu^{2,*}, Mitchell S.V. Elkind^{3,4,5}, Muredach P. Reilly^{3,5}, Chunhua Weng^{2,5,#}, and Shing M. Lee^{1,5,#}

¹Department of Biostatistics, Mailman School of Public Health, Columbia University, New York, New York, USA, ²Department of Biomedical Informatics, Columbia University, New York, New York, USA, ³Department of Neurology, Vagelos College of Physicians and Surgeons, Columbia University, New York, New York, USA, ⁴Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, New York, USA, and ⁵Irving Institute for Clinical and Translational Research, Columbia University, New York, New York, USA

*Lauren Franks and Hao Liu contributed equally.

#Chunhua Weng and Shing M. Lee contributed equally.

Corresponding Author: Shing Lee, PhD, Department of Biostatistics, Mailman School of Public Health, Columbia University, 722 W. 168th Street, 6th Floor, New York, NY 10032, USA; sml2114@cumc.columbia.edu

Received 31 March 2021; Revised 13 July 2021; Editorial Decision 26 July 2021; Accepted 2 August 2021

ABSTRACT

Hundreds of interventional clinical trials have been launched in the United States to identify effective treatment strategies for combating the coronavirus disease 2019 (COVID-19) pandemic. However, to date, only a small fraction of these trials have completed enrollment, delaying the scientific investigation of COVID-19 and its treatment options. This study presents novel metrics to examine the geographic alignment between COVID-19 hotspots and interventional clinical trial sites and evaluate trial access over time during the evolving pandemic. Using temporal COVID-19 case data from USAFacts.org and trial data from ClinicalTrials.gov, U.S. counties were categorized based on their numbers of cases and trials. Our analysis suggests that alignment and access have worsened as the pandemic shifted over time. We recommend strategies and metrics to evaluate the alignment between cases and trials. Future studies are warranted to investigate the impact of the misalignment of cases and clinical trial sites on clinical trial recruitment.

Key words: clinical research informatics, clinical trial recruitment, COVID-19, clinical trial site location

INTRODUCTION

Coronavirus disease 2019 (COVID-19) is one of the most elusive and yet widespread infectious diseases in modern medicine. Our knowledge regarding its etiology, transmission, and treatment remains limited. Since January 2020, when the first case of COVID-19 was confirmed in the United States, over 800 COVID-19 interventional clinical trials have been launched in the United States to evaluate potential therapies and strategies for combating the unprecedented COVID-19 pandemic.¹ However, as of October 31, 2020,

only 7.5% of interventional trials have completed enrollment.² While there have been over 30 million COVID-19 cases in the United States, clinical trials have been largely challenged by slow recruitment, delaying answers to clinical questions that are vital to advancing the scientific understanding of COVID-19 and its treatment.³ Many factors could be contributing to this dilemma, including restrictive or unrepresentative clinical trial eligibility criteria, patient reluctance or inability to enroll in clinical trials, and poor patient awareness of or access to clinical trials.⁴ With the dynamic na-

ture of the pandemic and its rapid geographical shifts over time, many clinical centers have seen a paradoxical decline in the number of patients just as the number of trials had been rising. We developed a novel approach for examining the geographic alignment of cases and trials sites and applied it to investigate if the poor geographic alignment between COVID-19 cases and interventional trials sites could have contributed to low enrollment and hindered the completion of clinical trials evaluating treatments for patients with COVID-19. Furthermore, we discuss the important considerations for data-driven clinical trial site selection and across-sites collaboration in response to future rapidly evolving pandemics.

MATERIALS AND METHODS

This study is not a human subject study because all the data are obtained from the public domain—ClinicalTrials.gov,² USAFACTS.org,⁵ and Census.gov.⁶ COVID-19 clinical trials data were extracted from ClinicalTrials.gov through October 31, 2020. The search term “COVID-19” was used, with the advanced search options for type being “interventional study” and country being “United States.” Studies that closed before enrolling patients, had not started (with start dates later than October 31, 2020), or had status “Unknown” or “Not yet recruiting” were excluded. Two assessors (L.F. and H.L.) independently reviewed the title and ClinicalTrials.gov details for each returned clinical trial to identify trials providing active intervention, treatment, or supportive care to COVID-19 patients. Therefore, prevention, vaccine, diagnostic, and behavioral trials not providing interventions to COVID-19 patients were excluded. For each trial, clinical site information (including city, state, zip code) was extracted, and the address for each enrolling site was mapped to its corresponding county using the Google Map Geocoding API service.⁷ The number of COVID-19 cases for each county in the United States were obtained from USAFacts.org. For each county, we calculated the average number of new daily cases for each month. We further categorized each county based on the presence of interventional clinical trials in the county. The 2019 population estimates were obtained from Census.gov⁶ to calculate the number of COVID-19 cases per 100 000 in each state as of October 31, 2020.

To evaluate the temporal geographic increase in the number of trials in each state, regression lines were fit to each state’s monthly trial count, and the estimated slope was categorized into the following categories—<1, 1 to 2, 2 to 3, 3 to 4, and 4+—and visualized in a tile map. To examine the temporal alignment between COVID-19 cases and interventional trials, we examined the maximum number of trials actively enrolling in each county relative to the average number of new COVID-19 cases by month between March and October 2020. Trial start date, primary end date, and site information were used to obtain the maximum number of actively enrolling trials by county for each month. Counties without trials were categorized into 1 of the following 3 groups: (1) <10 average daily cases, (2) 10 to 50 average daily cases, and (3) ≥ 50 average daily cases. For counties with trial presence, we used the ratio of cases to trials and classified them as active counties if they had ≥ 10 new daily cases per trial and inactive counties if they had <10 new daily cases per trial. A threshold of ≥ 10 average daily cases per trial was used because it yields at least 300 new cases monthly per trial. Assuming a threshold of 2% of patients are eligible and enrolled based on the prior literature, there would be at least 6 new patients per trial, worthwhile to open a new site.⁸ Given that inactivity could be due to either a low number of new daily cases (<10 daily cases) or an abundance of tri-

als that leads to a ratio of <10 average new daily cases per trial, we further classified the inactive counties as (1) <10 cases and <10 cases per trial or (2) ≥ 10 cases and <10 cases per trial.

These county-level categorizations were developed to quantify case-trial alignment and displayed graphically by month to visualize the alignment over time, as well as in a map format to illustrate the geographic alignment of trials and cases for April, July, and October peaks of the pandemic. We also summed the number of counties with ≥ 10 daily cases, from which we reported the proportion of counties without trials by month. We also contrasted the increase in the total number of counties with 10 to 50 new daily cases but no trial access vs the increase in the total number of counties with trials by month. The proportion of counties with trials categorized as inactive due to an abundance of trials was also reported in early March versus early fall.

RESULTS

We identified 366 actively recruiting COVID-19 interventional trials that met our trial search criteria between March and October 2020 and were conducted across 3141 counties in the United States. These trials had sites distributed across 2689 counties in 49 states. The agreement between the assessors was 100% for the identification of the COVID-19 interventional trials. The trial-to-county mapping and each trial’s recruiting status are provided in the [Supplementary Appendix](#).

Figure 1A displays the estimated regression slope of the number of trials by month in each state. Three states shaded in gray (ie, Alaska, Delaware, and Wyoming) did not have any trial. States colored in the lightest blue had regression slopes of <1, indicating a very slow growth of trial sites over time, whereas states colored in dark orange had the highest rate of trial growth through the course of the pandemic. The smallest slope of 0.04 was seen in Hawaii and the largest was in Texas with 8.76. There were no negative slopes estimated, indicating that there were no states that experienced an overall decline in trial site openings. While the majority of the states had only slight increases in the number of trials, states with large urban areas had significant increases. For comparison, it is worth noting that many of the states with the highest number of cases per 100 000 as of October 31, 2020 are states with large rural areas (Figure 1B).

Figure 2 shows the breakdown of counties by the specified categories over the first 8 months of the pandemic. The orange and red bars indicate the number of counties with 10 to 50 and >50 new daily cases without trial presence, respectively. The blue bars indicate the number of counties with trial presence. The inactive counties are colored in lighter blues with inactivity owing to low number of cases being colored in the lightest blue. The active counties with >10 new daily cases per trial are shaded in dark blue. Counties with <10 daily cases and no trial presence are not depicted in the figure. The number of counties with <10 daily cases and no trial presence decreased from March to October 2020 as the COVID-19 became more widespread across the country over time. The large increase in the orange and red bars indicates that the number of counties with ≥ 10 new daily cases and no trial presence increased significantly between March ($N=22$) and October ($N=848$). The sum of the blue bars indicates that the number of counties with trials more than doubled from 113 in March to 295 in October, suggesting that trial site presence did not keep up with this spread of the pandemic across the United States. March recorded the fewest number of counties with trials and cases, documenting only 79 counties with ≥ 10 new daily

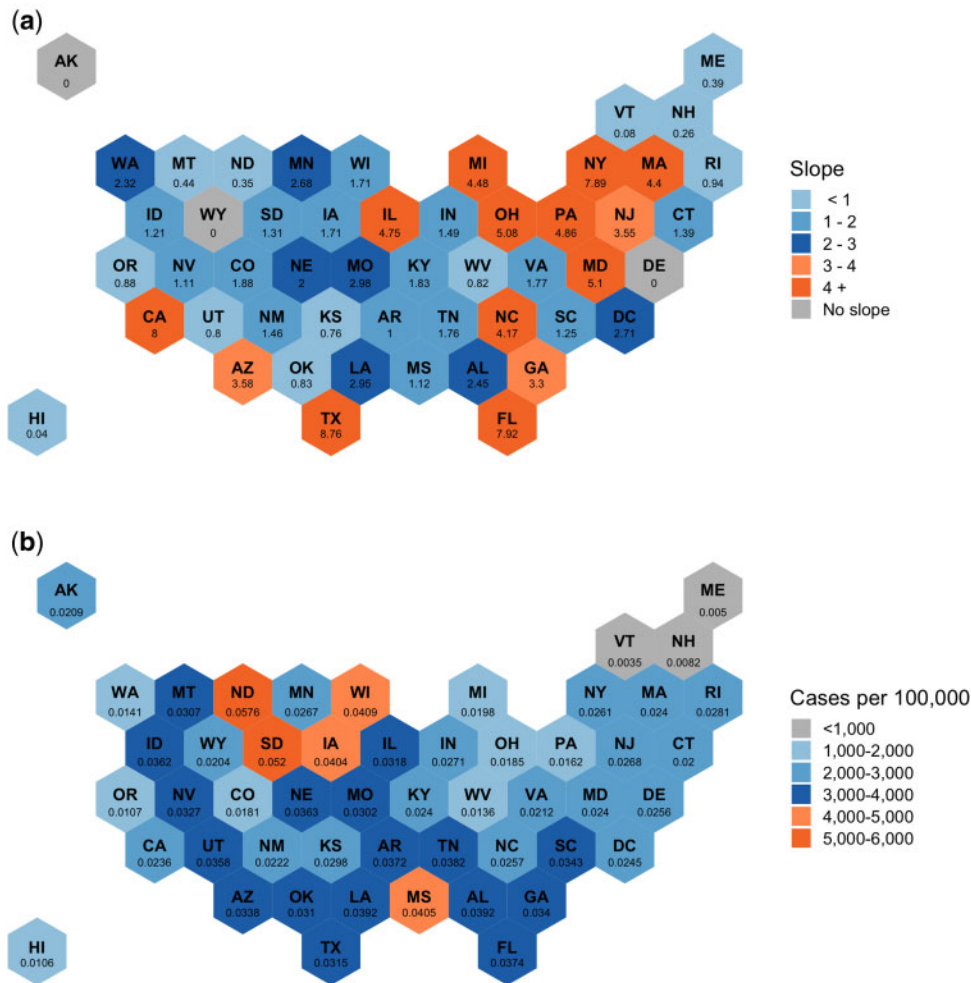


Figure 1. Tile maps for the estimated regression slope of (A) the number of trials per month by state from March to October 2020 and (B) the number of COVID-19 (coronavirus disease 2019) cases per 100 000 by state as of October 31, 2020.

cases, of which 22 (28%) did not have a clinical trial presence. By October, these numbers increased to 1118 counties with ≥ 10 new daily cases, and 848 (76%) of those counties had no clinical trial presence. With the shifting pandemic, a large number of these counties remained inactive. The number of trials facing inactivity due to an abundance of trials increased from March to September, with 98 counties reporting ≥ 10 daily cases and < 10 daily cases per trial in September compared with 31 counties in March. This number decreased slightly in October, with only 70 counties categorized as inactive. October also experienced the greatest number of counties with 50+ daily cases and no clinical trials, with 106 counties in October vs 5 in March.

Figure 3 displays the widespread nature of the pandemic, highlighting the 3 peak months (ie, April, July, and October 2020). The same shading as that of Figure 2 is used, with the addition of gray which represents counties with < 10 daily cases and without trial presence. Orange and red shades represent counties with cases but no trial presence, while shades of blue represent counties with a trial presence, varying shades indicating the different levels of active and inactive categorization. With the large number of counties shaded in gray, orange, and red, it is evident that there is a lack of trial available for many counties affected by COVID-19. Moreover, over the 3 peak months, the number of counties with significant number of cases but without access to trials was worsening and

more widespread geographically for each consecutive peak month, highlighting the increasing lack of access to interventional COVID-19 interventional trials in less densely populated and rural areas. This is illustrated by the increasing number of counties in orange and red in the middle of the map over time.

DISCUSSION

Our results indicate that as the pandemic shifted geographically in the United States, trial access and geographic alignment have worsened. An increasing number of cases are located in counties without trials, while a large number of counties have low case to trial ratios, likely owing to an overabundance of competing trials. The overall population size and the existence of local clinical trial infrastructure can be contributing factors to the disparity in trial distribution because it is easier for large academic medical centers to launch more trials, resulting in competition for patients within large metropolitan areas or even within academic medical centers, which can slow down the recruitment of individual trials.

With the large number of COVID-19 trials, investigators should work in a regionally and nationally coordinated manner and consider closing trials as evidence becomes available, limiting opening trial sites in already saturated areas and prioritizing new sites in

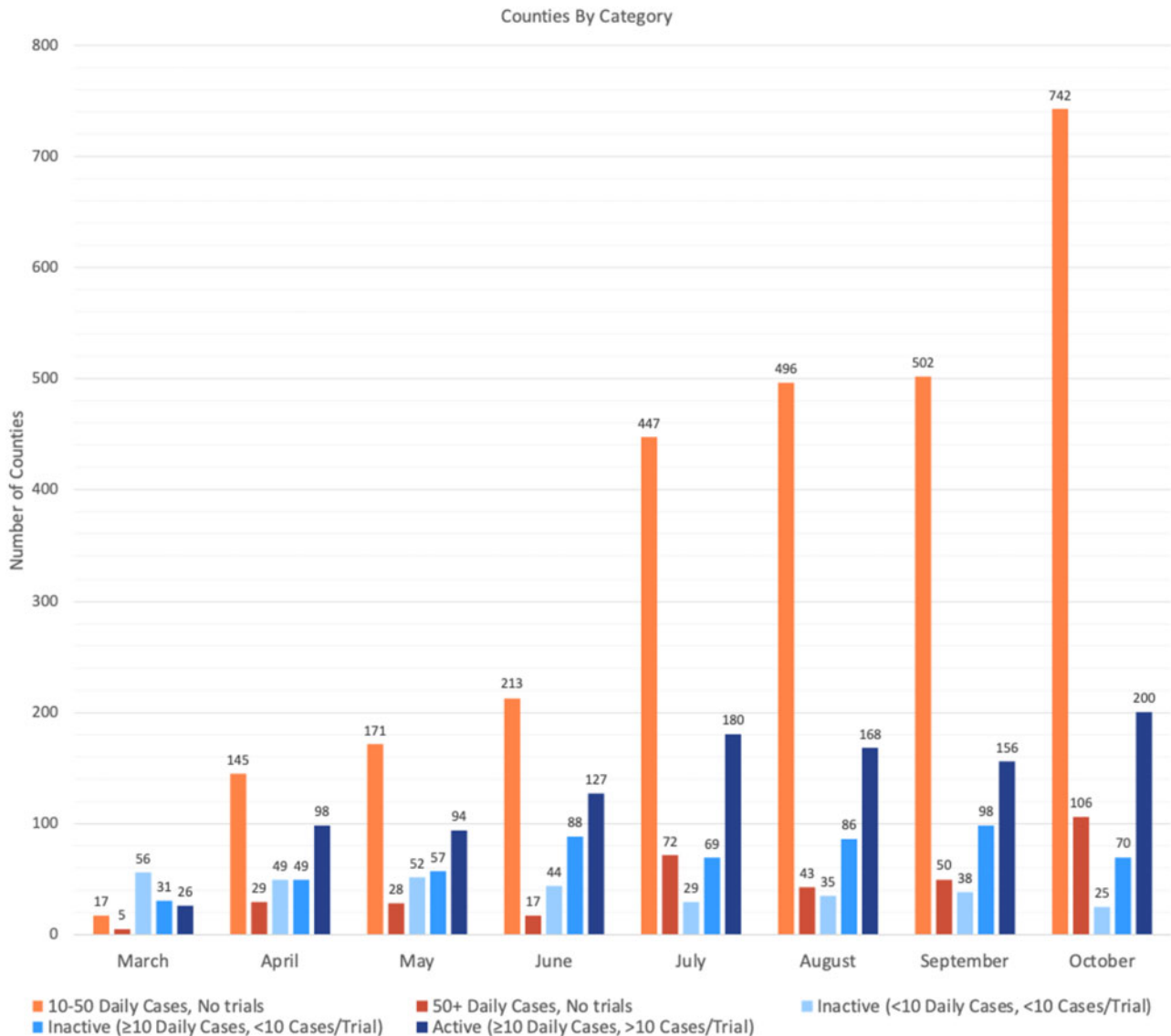


Figure 2. Number of counties by case levels and trial activity over time. Counties are colored based on trial presence and average new daily COVID-19 (coronavirus disease 2019) cases. Orange bars indicate counties without trial presence. Blue bars indicate counties with trial presence.

counties with high case levels and no trials.⁹ The increasing number of counties with cases, but without access to trial further highlights disparity in access to COVID-19 trials and emphasizes the need to expand trial sites to rural areas to accelerate recruitment and ensure generalizability of results.¹⁰ Leaders of clinical trial networks and trial investigators should leverage already established networks in rural areas and regions of poor access such as the Appalachian Translational Research Network to improve trial access.¹¹ Increasing collaboration between major academic medical institutions and local medical centers will maximize enrollment and effectively utilize resources, while improving a much-needed national infrastructure for medical collaboration.

The COVID-19 pandemic has highlighted many challenges surrounding trial design and execution. Social media and the news have made it increasingly difficult to ensure that patients are receiving accurate information regarding the virus, leading to patient uncertainty in treatment decisions.⁴ This could have impacted the number of patients willing to enroll in interventional

COVID-19 trials. However, investigators can expand trial access through adding sites in areas with high number of cases but no trial. This article proposes a data-driven approach with novel metrics and visualization for identifying these locations. The thresholds and mapping of geographic areas used for visualization can be tailored for specific disease areas and easily applied to other settings. Competitions for similar patients among trials can impede recruitment for all related trials conducted within neighboring areas. Thus, efficient identification of similar or competing trials in specific geographic regions during the planning of trials can guide clinical trial sponsors and investigators with optimal trial site selection or facilitate meaningful collaboration among trial designers and sponsors to minimize redundant and competing trials. A limitation of this article is that lack of recruitment data to evaluate the impact of the misalignment on recruitment given that as of May 17, 2021 only 23% of the trials had completed, with the majority of them being multicenter trials with trials sites across various counties. Future studies are needed to investigate

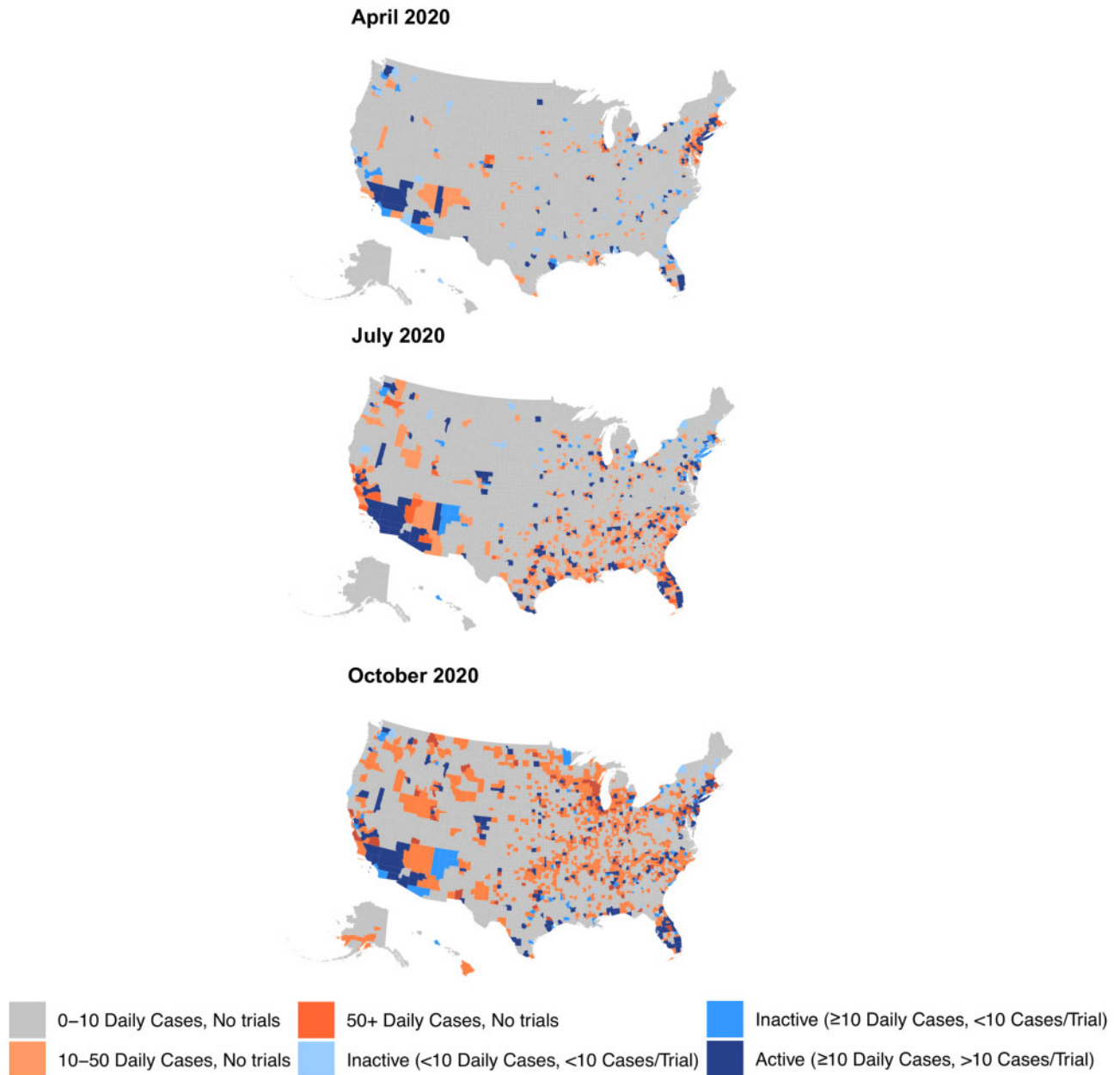


Figure 3. Geographic case levels and trial activity by county in April, July, and October 2020. The counties are color coded based on trial access and average daily cases.

the impact of the misalignment of cases and clinical trial sites on clinical trial recruitment and completion.

CONCLUSION

Most of COVID-19 interventional clinical trials have suffered from slow recruitment, potentially affected by the misaligned case hot-spots and trial sites as well as by local competition. Clinical trial site selection should be more data driven and account for patient population sizes, especially for interventional clinical trials launched to combat rapidly evolving pandemics.

FUNDING

All authors are funded through National Center for Advancing Translational Sciences, National Institutes of Health Grant Number

UL1TR001873. CW and HL were funded by National Library of Medicine grant R01LM009886.

AUTHOR CONTRIBUTIONS

SML, CW, MSVE, and MPR developed the idea for the manuscript; LF and HL provided data analysis; LF, CW, and SML wrote the first draft of the article; LF developed the figures; and all authors provided written input and edits for the final draft.

SUPPLEMENTARY MATERIAL

[Supplementary material](#) is available at *Journal of the American Medical Informatics Association* online

DATA AVAILABILITY STATEMENT

The datasets were derived from sources in the public domain: ClinicalTrials.gov (<https://clinicaltrials.gov>), USAFacts.org (<https://usafacts.org/visualizations/coronavirus-covid-19-spread-map>), and Census.gov (<https://www.census.gov/data/tables/time-series/demo/popest/2010s-national-total.html>). Part of the derived data generated in this research are provided in the [supplementary materials](#), the rest will be shared on reasonable request to the corresponding author.

CONFLICT OF INTEREST STATEMENT

None.

REFERENCES

1. Holshue ML, DeBolt C, Lindquist S, *et al.*; Washington State 2019-nCoV Case Investigation Team. First case of 2019 novel coronavirus in the United States. *N Engl J Med* 2020; 382 (10): 929–36.
2. ClinicalTrials.gov. <https://clinicaltrials.gov>. Accessed November 2, 2020.
3. Centers for Disease Control and Prevention. COVID Data Tracker. https://covid.cdc.gov/covid-data-tracker/#cases_casesinlast7days. Accessed March 31, 2021.
4. Burki TK. Completion of clinical trials in light of COVID-19. *Lancet Respir Med* 2020; 8 (12): 1178–80.
5. USAFacts.org. U.S. coronavirus cases and deaths. <https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/>. Accessed November 2, 2020.
6. Census.gov. National population totals and components of change: 2010-2019. <https://www.census.gov/data/tables/time-series/demo/popest/2010s-national-total.html>. Accessed May 24, 2021.
7. Google Geocoding API. <https://developers.google.com/maps/documentation/geocoding/overview>. Accessed May 13, 2021.
8. Weng C, Batres C, Borda T, *et al.* A real-time screening alert improves patient recruitment efficiency. *AMIA Annu Symp Proc* 2011; 2011: 1489–98.
9. North CM, Dougan ML, Sacks CA. Improve clinical trial enrollment – in the COVID-19 era and beyond. *N Engl J Med* 2020; 383 (15): 1406–8.
10. Dandachi D, Reece R, Wang EW, *et al.* Treatment COVID-19 in rural America. *J Rural Health* 2020; 35: 205–6.
11. Appalachian Translational Research Network. <http://appalachianresearch.org>. Accessed November 4, 2020.