Inequitable availability of COVID-19 clinical trials for Hispanic populations in the United States

Authors: Carrie Bennette¹,², Hil Lyons¹, Bradley Wagner¹, Blythe Adamson²,³, Stewart Chang ¹ on behalf of the IDM COVID-19 Response Team

¹Institute for Disease Modeling, Seattle, Washington; covid@idmod.org
²University of Washington, Seattle, WA
³Flatiron Health, New York, NY

Results as of 2020-08-19 10:00 am

What do we already know?
Access to and patient enrollment in clinical trials of COVID-19 therapies is likely strongly impacted by where trials are launched. Our objectives were to evaluate geographic factors associated with the location of COVID-19 clinical trials launched in the US to inform future site selection and resource allocation.

What does this report add?
A third of all COVID-19 cases in the US are in counties that have not had a single recruiting trial open since the pandemic began. Despite a rapid increase in the number of trial sites open in the US from March to June 2020 (n=186 to 1156), the proportion of cases in counties without an open trial remained relatively consistent (34% in March vs 32% in June), suggesting that trials continued to overlook certain counties with patients. Counties with a higher proportion of Hispanic residents had significantly fewer COVID-19 trials (rate ratio per 10% increase in % Hispanic: 0.825; 95% CI: 0.722, 0.951), after adjusting for county-level characteristics such as socioeconomic status, healthcare system factors, case burden, and epidemic duration.

What are the implications for public health practice?
A more targeted selection of future trial sites has the potential to greatly expand the number of eligible COVID-19 cases available for enrollment, thereby accelerating efforts to discover effective treatments, and improve access for traditionally underrepresented patient populations.
Introduction

Clinical trials generate critical evidence needed to guide the treatment of patients with COVID-19 and help identify promising new therapies to combat the deadly virus. Unfortunately, there are many barriers to successful participant enrollment in clinical studies, resulting in trials that are often severely delayed or simply unable to reach their target enrollment\(^1\). Poor enrollment in clinical trials negatively affects the pace of scientific progress and increases the cost of drug development.

Prior studies have identified a lack of geographically accessible clinical trial sites as one of the strongest factors hindering patient enrollment\(^2,3\), particularly for racial and ethnic minority populations\(^4-6\). Participation of diverse populations in clinical trials is vital to ensure generalizability of the results generated, facilitate discovery of novel therapies that will be effective in traditionally underrepresented populations, and ensure equitable access to new and promising treatments\(^7\). Yet the participation rate of many minority populations remains substantially lower than the composition of the overall US population for trials\(^8\). Emerging evidence indicates that Black and Hispanic people bear a disproportionate burden of COVID-19–related outcomes\(^9\), but their access to promising clinical trials remains largely unexplored.

It is critical to evaluate how the geographic availability of COVID-19 clinical trial sites varies with population, racial and ethnic composition, healthcare infrastructure, and case burden to ensure maximal and equitable access. In this study, we therefore explore the geographic factors and temporal trends associated with the location of COVID-19 clinical trials launched in the US early in the pandemic.

Methods

Data

We downloaded a record of all COVID-19 clinical trials from Cytel’s COVID-19 clinical trial tracker on July 31, 2020\(^10\). Cytel’s publicly available tracker collates information from ClinicalTrials.gov as well as other international trial registries and uses automated algorithms and manual review to classify clinical trials studying therapies for COVID-19. We identified interventional trials for COVID-19 treatments (of any phase with both industry and government sponsors) listed as recruiting in the US. We used the Aggregate Analysis of ClinicalTrials.gov (AACT) database to obtain the zip codes for recruiting locations for more than 95% of these trials and then mapped these zip codes to US counties using the zip-code tabulation area-to-county crosswalk files\(^11\). Because case data was aggregated for several counties comprising New York City (i.e., Kings, Queens, Bronx, and Richmond counties), we similarly aggregated other measures by taking a population weighted average. Our final analysis comprised 601 unique trial locations for 231 treatment trials.
Additional data sources:

i. County population, population density, and racial and ethnic demographics from 2018 American Community Survey (ACS)^12.

ii. Daily case counts by county from the NYT\textit{imes}^13.

iii. County-level COVID vulnerability measures from the Surgo Foundation^14, which builds on the Centers for Disease Control and Prevention’s Social Vulnerability Index^15, a validated metric that captures four themes: socioeconomic status, household composition and disability, minority status and language, and housing type and transportation. The additional COVID-19 specific themes capture epidemiologic factors (high-risk populations with underlying health conditions and high population density) and healthcare system factors (health system capacity, strength, and preparedness). Specifically, we used the vulnerability indices “socioeconomic status” from US CDC (percent individuals below poverty, percent civilian unemployed, per capita income in 1999, percent persons with no high school diploma), “housing type and transportation” from US CDC (percent housing structures that are multi-unit structures, percent housing structures that are mobile homes, crowding as defined by percent households with more persons than rooms, percent households with no vehicle available, and percent of persons in group quarters), and “epidemiological factors” compiled by the Surgo Foundation (percent of adults with cardiovascular conditions, respiratory conditions, obesity, or diabetes, incidence of cancer, prevalence of HIV, mortality from influenza or pneumonia, and population density). These indices were compiled by the Surgo Foundation database at the county level as percentiles, therefore varying between 0 and 1. We did not use the aggregated minority status and language measure because we were interested in understanding relationships with specific racial and ethnic populations, derived from the Census directly. We also did not include the healthcare systems factor measure as we were primarily interested in the infrastructure to support clinical trials versus responding to an epidemic, which we modeled using a combination of hospitals, universities, and major medical centers, as described below. Lastly, because of collinearity of household composition and housing type/transportation measures, we only included the latter in our regression models.

iv. Number of hospitals and universities from the Department of Homeland Infrastructure Foundation-Level Data (HIFLD)^16. We mapped institutions to counties via zip codes as indicated above.
**Statistical analysis**

We estimated the hypothetical number of COVID-19 cases available for each “new” trial spot in each county over time accounting for both the changing case burden and number of existing or “competing” trials. To calculate the average monthly number of trial spots available within a county, we first divided each trial’s target enrollment by the planned number of months of enrollment to estimate an average monthly recruitment target for each trial. We then assigned a proportion of each trial’s monthly enrollment target to each county by allocating the total enrollment target across all counties in which the trial was open, weighted by the monthly case burden of the county. We then added the monthly recruitment target across all trials open within a county for each month from March through June 2020. Lastly, we divided the number of cases in each month by the total recruitment target for all open trials in the same month (with minimum denominator assumed to be 1) to estimate the number of cases available for a hypothetical additional trial spot. Less than 10% of trials in our sample had one or more international sites (median: 4; IQR: 2, 8 ex-US sites); our estimate of target enrollment does not account for potential recruitment from international sites given the difficulty in generalizing recruitment targets outside the US, and therefore assumes all recruitment is from US sites.

We used negative binomial models to assess whether county-level characteristics were associated with the number of open COVID-19 clinical trial sites. We followed a similar approach as used by Millet et al\(^9\), which included county population as an offset term, a variable representing the number of days since the first case of COVID-19 was reported in each county in order to control for potential confounding by length of the outbreak, and spatially structured state-level random effects. We adjusted for all county-level characteristics included in Table 1. The model was fit using integrated nested Laplace approximations\(^{18}\). We were not able to evaluate trial availability for smaller racial and ethnic minority groups, such as Native Americans or Pacific Islanders, given the overall low representation in the US and smaller variability across counties. All analyses were performed in R 4.0.2 and code is available on GitHub\(^{19}\).
Results

The overall number of COVID-19 clinical trial sites open in the US increased between March and June 2020 (n=186 in March to n=1156 in June). Despite this increase, the proportion of COVID-19 cases in the US that are in counties without a single recruiting trial site open has remained relatively constant (between 32% and 36%) since the pandemic began (Table 1). Most of these counties (92%) are communities with outbreaks of more than 100 cases and at least one hospital.

Table 1. Distribution of cases, counties and total population living in counties with and without a COVID-19 treatment clinical trial site, March-June 2020. Only counties with 100+ cases and 1+ hospital are included; no county without a case opened a trial.

<table>
<thead>
<tr>
<th></th>
<th>Proportion of COVID-affected US counties* that do not have an open COVID-19 trial</th>
<th>Proportion of all COVID-19 cases in the US that are in counties without an open COVID-19 trial</th>
</tr>
</thead>
<tbody>
<tr>
<td>March</td>
<td>0.61 (122 / 199)</td>
<td>0.34 (61,230 / 182,757)</td>
</tr>
<tr>
<td>April</td>
<td>0.74 (457 / 619)</td>
<td>0.34 (304,860 / 886,057)</td>
</tr>
<tr>
<td>May</td>
<td>0.75 (584 / 781)</td>
<td>0.36 (326,985 / 901,165)</td>
</tr>
<tr>
<td>June</td>
<td>0.78 (868 / 1109)</td>
<td>0.32 (552,854 / 1,730,570)</td>
</tr>
</tbody>
</table>

* Counties with at least 100 cases and at least one hospital

From March to June 2020, inequality in the geographic distribution of COVID-19 trial sites increased. In rankings of counties by the number of trial sites open by month, counties with the highest and lowest numbers of trial sites open were generally consistent between March and June (Kendall’s rank correlation coefficient: 0.58; p < 0.001). For example, 84% of counties that launched at least one COVID-19 clinical trial site in March launched at least one additional site in the subsequent three months, compared to less than 6% of counties without a trial site in March.

Across all counties, a median average of 26 cases (IQR: 6, 102) were available for each additional trial site in June (Figure 1), compared to 2 cases (IQR: 0, 10) in March (data not shown). However, many counties with high numbers of cases had launched few or no trials as of June 2020 (left side of Figure 1), and major metropolitan areas with large case counts had already launched multiple trials (red points on right side of Figure 1), reducing the effective number of cases available for a hypothetical future trial.
The availability of open trials differed significantly by ethnic makeup at the county level. Counties with a higher proportion of Hispanic residents had fewer open COVID-19 trials (Figure 2a). For example, communities with ~1% Hispanic residents had 6.7 (95% CI: 4.7, 8.7) open trials per 100,000 residents while communities with 10% Hispanic residents had only 1.5 (95% CI: 0.5, 2.5) on average. This trend was not substantially changed after adjusting for percent urban, local case burden, COVID vulnerability measures, and medical research infrastructure (rate ratio per 10% increase in Hispanic population: 0.825; 95% CI: 0.722, 0.951; Table 2). In contrast, we observed that the number of open COVID-19 trials in a county varied non-linearly with the proportion of Black residents, reaching a minimum at counties comprised of ~10% Black residents (Figure 2b). After adjusting for county-level characteristics, there was a significant association between a higher proportion of Black residents and higher rates of open COVID-19 trials (rate ratio per 10% increase in Black population: 1.18; 95% CI: 1.04, 1.33; Table 2). Urban populations, housing types associated with vulnerable populations, and the number of hospitals and major medical centers were also significantly associated with the number of trials open at the county level, indicating that trials tended to be in metropolitan areas with stronger medical research infrastructure (Table 2). Importantly, we also found that communities with few Black residents had higher numbers of hospitals relative to their population (see Appendix), which partially explained the higher number of trials in these communities seen in Figure 2b.
Figure 2. Number of open COVID-19 trials (per 1000 cases) in a county by the proportion of (a) Hispanic residents or (b) Black residents. Trials open at any time March – June 2020 are counted. Circle size area is proportional to number of cases.

(A)

(B)
Table 2. Association of county-level characteristics with the rate of open COVID-19 treatment trials in US counties. All results are from a multivariable negative binomial regression model fit with spatially structured state random effects. Rate ratios less than one mean that higher levels of a given characteristic are associated with lower rates of open COVID-19 trial sites. (Significant factors highlighted in bold.)

<table>
<thead>
<tr>
<th>Rate Ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Hispanic (per 10%)</td>
</tr>
<tr>
<td>% Black (per 10%)</td>
</tr>
<tr>
<td>% total population that is urban (per 10%)</td>
</tr>
<tr>
<td>COVID-19 cases (per 1% increase)</td>
</tr>
<tr>
<td>COVID-19 deaths (per 1% increase)</td>
</tr>
<tr>
<td>Vulnerability Indices (per decile; higher values = more vulnerable)</td>
</tr>
<tr>
<td>Socioeconomic Status</td>
</tr>
<tr>
<td>Housing Type &amp; Transportation</td>
</tr>
<tr>
<td>Epidemiologic Factors</td>
</tr>
<tr>
<td>Time since first case detected (per week)</td>
</tr>
<tr>
<td>Per university</td>
</tr>
<tr>
<td>Per hospital</td>
</tr>
<tr>
<td>Major medical center in county</td>
</tr>
</tbody>
</table>

Discussion

We used publicly available data to explore the geographic factors and temporal trends associated with the location of COVID-19 clinical trials launched in the US between March and June 2020. Overall, we found COVID-19 clinical trials launched in the United States were concentrated in a small number of counties, with less trial availability in counties with more Hispanic residents. These trends have left substantial numbers of COVID-19 cases without access to clinical trials of new treatments, potentially slowing down trial enrollment and ultimately scientific progress. Going forward, a more targeted selection of trial sites has the potential to expand the number of diverse, representative people available for enrollment, thereby accelerating efforts to discover effective treatments, and improve access for traditionally underrepresented patient populations.

Geographic inequity in COVID-19 clinical trials

We found evidence of important, but nuanced, disparities in the availability of COVID-19 trials by county-level racial and ethnic composition. First, there was a consistent trend towards fewer trials being available in increasingly Hispanic communities, which remained statistically...
significant after accounting for other county-level characteristics such as socioeconomic status and healthcare infrastructure. Second, after adjusting for other county-level characteristics, we found that an increased proportion of Black residents was associated with significantly more open trials, which was counter to our original hypothesis. Importantly, our main regression model also accounts for a county’s medical research infrastructure, such as the number of hospitals relative to the population, which we found was substantially higher in communities with few Black residents (but not those with few Hispanic residents; see Appendix). Thus the increased availability of hospitals among predominantly non-Black communities appears to confound the relationship with the availability of clinical trials and a county’s proportion of Black residents. On the one hand, there is some good news: after accounting for disparities in medical research infrastructure, we find evidence that communities with a larger Black population have significantly more open COVID-19 trials. On the other hand, our results are a sobering reminder of the unequal allocation of hospital infrastructure that supports such trials.

Improving future clinical trial site selection

Leveraging epidemiologic data will likely be useful for trial sponsors identifying future sites that may not already be part of a sponsor or research organization’s list. Prior to initiating site selection, researchers for the National Drug Abuse Treatment Clinical Trials Network (CTN) examined epidemiologic data to identify areas with highest potential for recruitment and then reached out to potential sites to gauge interest in participating\(^{21}\). They selected several sites that were outside their network and would not have otherwise been considered. Subsequently one of the best performing sites in terms of recruitment was a site with no prior research experience, but with a large pool of available patients as suggested by epidemiological data.

Importantly, when selecting a site for a future COVID-19 trial it is important to consider not only the future incidence of cases, but also the potential competition from any existing trials. Prior research has shown that competition for patients is one of the most important factors leading to poor enrollment rates\(^{22-24}\). We found that counties with the largest case burden generally also had the greatest number of open trials and highest target trial enrollment rates, leading to many of these areas having fewer patients available for a future trial. Conversely, many areas with fewer cases also had considerably fewer open trials competing for these patients. For example, Dallas County, TX had more than 14,000 cases in June, but also had 16 trials open and a target enrollment that resulted in fewer than 25 patients per target trial spot, whereas Clark County, NV had only 10,000 cases in June, but a target enrollment that left more than 250 patients per trial spot. There are more than 100 counties without a single open trial through June, but more than 1000 cases in the month of June alone; many of these counties are likely excellent candidates for a future trial site. To aid decision makers, we have created a list of some of the most promising counties based on the recent case burden, healthcare
infrastructure, number of competing trials, and ability to address inequities in access for Hispanic populations (see Appendix).

Identifying new or previously overlooked trial sites with strong recruitment potential may require additional investment from trial sponsors to train or hire additional research staff, but these investments are likely to be offset by the ability to accelerate enrollment. Study sponsors have stated that they consider access to the relevant patient population and recruitment-related factors to be the most important factors, and the majority would prefer reaching their target enrollment faster, even if it increased costs substantially\textsuperscript{25, 26}. These preferences from the pre-pandemic era are likely to be heightened in the current pandemic when delays in therapeutic and vaccine development will be associated with additional morbidity, mortality, and economic loss at a global scale.

Limitations

We highlight several limitations to these analyses. First, our main analyses are at the county-level, which may not align perfectly with how individuals seek healthcare, particularly in communities that span multiple counties. An alternative is to rely on hospital referral regions (HRRs)\textsuperscript{30}. However, the alignment between counties and HRRs is also imperfect and requires assumptions regarding the distribution of COVID-19 cases and deaths. The 306 HRRs are intended to reflect referral patterns for tertiary care, which is likely too broad a catchment area for treatment trials as most COVID-19 patients that are hospitalized will need urgent, but not tertiary-level care. Lastly, the much broader population served by each HRR may obscure potentially important relationships between sociodemographic characteristics and availability of clinical trial sites seen at a more granular county-level analysis. Secondly, counties still represent relatively aggregated measures of a population and our analysis may miss important disparities or inadequate access to a clinical trial site within a county. Having an open trial site nearby is a necessary but not sufficient condition for broad access and ultimately diverse enrollment. Ensuring more equitable enrollment in COVID-19 treatment trials will require additional efforts, such as community involvement, cultural adaptation of recruitment materials and increased diversity of principal investigators\textsuperscript{20}. That said, ensuring unbiased geographic availability of trial sites is an important step that may often be overlooked in selecting trial sites in favor of other metrics such as area case burden or an existing relationship with a site’s investigator.

Conclusion

The rising number of COVID-19 cases in the US is unquestionably a tragedy, but it presents an opportunity to discover new therapies to combat this deadly disease at a rapid rate. Moreover, efforts to expand access to clinical trials and diversify sites are likely to extend far beyond COVID-19. The National Academy of Sciences\textsuperscript{31} has argued that the current clinical trials
enterprise is inadequate to provide evidence needed for clinical practice and called for an expansive and harmonized improvement to the United States’ clinical trials infrastructure, yet little has materialized from such calls. The growing tragedy of COVID-19 could be a catalyst to usher in a new paradigm for clinical research in the US that is broadly distributed and able to efficiently and robustly generate evidence that improves population health.
References


Appendix

Number of hospitals per 100k residents by proportion of Hispanic residents

Number of hospitals per 100k residents by proportion of Black residents
Table of top 10 recommended counties for launching future COVID-19 trials to alleviate inequities in access for Hispanic populations. Recommendations based on ranking a county’s overall case burden and proportion Hispanic and healthcare infrastructure. Specifically, a county’s rank score = number of cases in prior month * (1 / number of trials open in county in most recent month) / 500 + proportion Hispanic and is limited to counties with at least 500 cases in prior month and one hospital.

<table>
<thead>
<tr>
<th>County</th>
<th>Cases (June)</th>
<th>Number trials (June)</th>
<th>Proportion Hispanic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Starr, TX</td>
<td>552</td>
<td>0</td>
<td>0.99</td>
</tr>
<tr>
<td>Webb, TX</td>
<td>992</td>
<td>0</td>
<td>0.98</td>
</tr>
<tr>
<td>Hidalgo, TX</td>
<td>3380</td>
<td>1</td>
<td>0.96</td>
</tr>
<tr>
<td>Cameron, TX</td>
<td>1532</td>
<td>1</td>
<td>0.91</td>
</tr>
<tr>
<td>El Paso, TX</td>
<td>3161</td>
<td>0</td>
<td>0.90</td>
</tr>
<tr>
<td>Imperial, CA</td>
<td>4151</td>
<td>1</td>
<td>0.88</td>
</tr>
<tr>
<td>Santa Cruz, AZ</td>
<td>1451</td>
<td>0</td>
<td>0.84</td>
</tr>
<tr>
<td>Yuma, AZ</td>
<td>5120</td>
<td>0</td>
<td>0.74</td>
</tr>
<tr>
<td>Miami-Dade, FL</td>
<td>18,820</td>
<td>18</td>
<td>0.67</td>
</tr>
<tr>
<td>Tulare, CA</td>
<td>2277</td>
<td>0</td>
<td>0.64</td>
</tr>
</tbody>
</table>