Mechanisms of socioeconomic differences in COVID-19 screening and hospitalizations

Jarrod E. Dalton¹, Douglas D. Gunzler², Vardhmaan Jain¹, Adam T. Perzynski², Neal V. Dawson², Douglas Einstadter², Yasir Tarabichi², Peter B. Imrey¹, Michael Lewis², Michael W. Kattan¹, James Yao¹, Glen Taksler¹, Kristen A. Berg², Nikolas I. Krieger¹, David Kaelber², Lara Jehi¹, Ankur Kalra¹,*

¹ Cleveland Clinic Lerner College of Medicine at Case Western Reserve University, Cleveland, Ohio, 2 Case Western Reserve University at MetroHealth Medical Center, Cleveland, Ohio

☯ These authors contributed equally to this work.
* kalraa@ccf.org

Abstract

Background

Social and ecological differences in early SARS-CoV-2 pandemic screening and outcomes have been documented, but the means by which these differences have arisen are not well understood.

Objective

To characterize socioeconomic and chronic disease-related mechanisms underlying these differences.

Design

Observational cohort study.

Setting

Outpatient and emergency care.

Patients

12900 Cleveland Clinic Health System patients referred for SARS-CoV-2 testing between March 17 and April 15, 2020.

Interventions

Nasopharyngeal PCR test for SARS-CoV-2 infection.

Measurements

Test location (emergency department, ED, vs. outpatient care), COVID-19 symptoms, test positivity and hospitalization among positive cases.
Results
We identified six classes of symptoms, ranging in test positivity from 3.4% to 23%. Non-Hispanic Black race/ethnicity was disproportionately represented in the group with highest positivity rates. Non-Hispanic Black patients ranged from 1.81 [95% confidence interval: 0.91–3.59] times (at age 20) to 2.37 [1.54–3.65] times (at age 80) more likely to test positive for the SARS-CoV-2 virus than non-Hispanic White patients, while test positivity was not significantly different across the neighborhood income spectrum. Testing in the emergency department (OR: 5.4 [3.9, 7.5]) and cardiovascular disease (OR: 2.5 [1.7, 3.8]) were related to increased risk of hospitalization among the 1247 patients who tested positive.

Limitations
Constraints on availability of test kits forced providers to selectively test for SARS-Cov-2.

Conclusion
Non-Hispanic Black patients and patients from low-income neighborhoods tended toward more severe and prolonged symptom profiles and increased comorbidity burden. These factors were associated with higher rates of testing in the ED. Non-Hispanic Black patients also had higher test positivity rates.

Introduction
Differences in SARS-CoV-2 infection rates, COVID-19 hospitalizations, and COVID-19 deaths relating to race, ethnicity and geography have been documented worldwide [1–11]. These alarming disparities were noted early in the pandemic in the UK and USA [12,13]. However, the underlying relationships that may operate to produce these differences are less understood.

In the USA specifically, Black or African American persons are substantially more likely to contract, be hospitalized, and die of COVID-19 than their White counterparts [3]. Geographically, COVID-19 has disproportionately impacted regions with more racial and ethnic minority representation [3,5,6,10,11,14]. Racial differences in risk of death from COVID-19 are likely complex and multifactorial [15–20]. These differences do not appear to be fully accounted for by variations in area-based socioeconomic position [6], or medical comorbidities [2,4].

The goal of this study was to further understand the social and biological processes that underlie these ecological differences. Specifically, we identified and analyzed potential mechanisms of social and geographic differences in early (March–April 2020) SARS-CoV-2 testing access, test positivity, and COVID-19 hospitalization rates, among a regional cohort of Northeast Ohio residents who underwent SARS-CoV-2 screening.

Methods
Development of hypothesized causal mechanisms
We conducted a series of 13 weekly videoconferences to construct hypothesized mechanisms linking race/ethnicity and neighborhood income to i) presentation to the emergency department (ED) as opposed to an outpatient testing location for SARS-CoV-2 screening, ii)
SARS-CoV-2 test positivity, and iii) COVID-19 hospitalization among those who screened positive for the SARS-CoV-2 virus. We identified several intersecting hypotheses supported by prior literature, and assembled these into a directed acyclic graph (see Fig 1). We detail below the rationale for each of these hypotheses.

We hypothesized that racial, ethnic, and income-related differences in obesity rates [21], asthma [22], chronic obstructive pulmonary disease (COPD) [23–25], type 2 diabetes mellitus (T2DM) [26–29], and other comorbid conditions [30] may result in differences in the extent and severity of COVID-19-related symptom patterns. The Cleveland Clinic Health System (CCHS) collects information on whether or not patients referred for testing have the following symptoms: cough, fever, flu-like symptoms, shortness of breath, diarrhea, loss of appetite, and vomiting. Rather than model individual symptoms, we sought to characterize patterns of co-occurring symptoms. We conducted a sub-analysis which sought to i) identify groups (classes) of patients with similar symptom profiles; ii) characterize the distribution of these symptom classes as a function of sociodemographic and comorbid characteristics (see “Symptom Class” in Fig 1), and iii) relate the derived symptom classes to the likelihood of testing occurring in the ED, likelihood of test positivity, and likelihood of hospitalization. Details on the derivation of these symptom classes are provided below, under Statistical Methods.

Smoking is disproportionately prevalent in disadvantaged communities [31,32], and the prevalence of smoking decreases in older populations [33]. Smoking by itself is associated with COVID-19 disease progression [34], and is a primary risk factor for COPD [35]. It thus represents a potential mechanism to which any observed variation in symptom patterns across the socioeconomic spectrum might be attributable.

Lack of vehicle ownership, whether due to urbanized living environments, neighborhood disadvantage or racial/ethnic disparities in employment, presents a significant barrier to timely outpatient testing utilization. Health care utilization, in general, is reduced among disadvantaged populations due to lack of financial resources, distrust, lack of time due to other responsibilities, lack of transportation, and other factors [36–39]. Therefore, neighborhood-level median income and vehicle ownership rates (defined as the percentage of households within a census tract owning at least one vehicle) may be related to the likelihood that a patient presents to the ED (vs. all other settings) for screening; as well as the nature, duration and severity of symptoms at the time of testing.
Data sources and inclusion criteria

The Cleveland Clinic COVID-19 registry, approved by the Cleveland Clinic Institutional Review Board, was designed to serve as a large case series study of the clinical characteristics and outcomes of all CCHS patients who are referred for COVID-19 testing. This study population is identified daily through built-in electronic health record (EHR) queries and linked to patients’ demographic, clinical and residential characteristics through the EHR. We included every adult patient who resided in one of 17 Northeast Ohio Counties and who was referred for testing from the registry’s inception on March 17, 2020 through April 15, 2020. We excluded patients who did not receive a test or whose test results were deemed erroneous. We also excluded patients whose symptoms were not documented. Addresses of residence for patients in the registry were mapped to U.S. Census’ Geographic Identifiers (GEOIDs). We extracted from the 2018 American Community Survey neighborhood-level characteristics (such as median income and percent of households owning one or more vehicles) corresponding to census tracts embedded within these GEOIDS using the R packages sociome and tidy-census [40,41].

Statistical methods

We developed regression models for each variable depicted in Fig 1. For each given variable in the network, we included all pre-specified parent nodes as predictors. Continuous and dichotomous variables were modeled using multivariable linear regression and multivariable logistic regression, respectively. Duration of symptoms was modeled using proportional odds logistic regression, with the following ordered levels: asymptomatic, or symptoms of 0–2 days, 3–7 days, 1–2 weeks, and >2 weeks duration. Neighborhood vehicle access, defined for each patient as the proportion of households in their census tract with 1 or more vehicle, was modeled using quasibinomial regression [42], a technique for modeling proportions that are not necessarily expressed as counts of events out of a total number of trials. Multinomial logistic regression was used to model smoking status (non-smoker, current smoker, former smoker, unknown smoking status).

Latent class analysis [43,44]—a method for empirically deriving homogenous groups of patients with respect to a set of binary indicators—was used to derive symptom classes. We estimated, using Mplus software version 8.4 [45], and compared solutions of up to 10 classes. We chose the solution for the number of classes that optimized model goodness of fit criteria (e.g., Bayesian Information Criteria [BIC]; entropy; and sequential likelihood ratio tests [LRTs]). Resulting classes were ordered with respect to increasing incidence of test positivity. Class membership was summarized across age, neighborhood income, race/ethnicity, sex, comorbidity status (set of clinical diagnoses and treatments depicted in Fig 1) and duration of symptoms.

We used quadratic polynomial terms for age and neighborhood income. For models including age, race/ethnicity and neighborhood income as predictor variables, we considered two- and three-way interactions among these variables. Interactions were removed from those models in an iterative manner, beginning with the 3-way interaction, when they did not meaningfully improve model goodness-of-fit criteria (e.g., residual deviance); we chose this approach instead of one strictly driven by statistical significance criteria because our sample size enabled detection of relatively minor interaction effects.

We used the RStudio Integrated Development Environment [46], running R version 3.5.0 (2018-04-23) [47] for all analyses, unless otherwise noted above. A reproducible analysis workflow, including code, analytic datasets and automated statistical reporting were developed using the projects and RMarkdown R packages [48,49].
Results

A total of 20392 patients were accrued into the Cleveland Clinic COVID-19 registry over the study period, of whom 14887 were documented as residing in Northeast Ohio. Of these, we removed 36 patients for whom a test was ordered but not performed, 66 additional patients for whom nasopharyngeal test results were deemed erroneous and 1885 additional patients whose symptom data were not recorded (total of 13.3% removed).

Within the patients who met study inclusion criteria, age distributions were comparable between non-Hispanic Black and non-Hispanic White populations (median [first quartile, third quartile] of 52 [37, 65] and 51 [36, 65] years, respectively). Hispanic patients and patients of other racial or ethnic backgrounds tended to be younger (43 [31, 57] and 42 [31, 57] years, respectively)). Sociodemographic and comorbid disease characteristics are summarized across levels of neighborhood median income in Table 1: Patients from lower-income communities were disproportionately non-Hispanic Black and/or Hispanic (see S1 File for distributions of neighborhood median income by race) and had disproportionately higher prevalence of T2DM, obesity, smoking, asthma, cardiovascular disease (CVD) and COPD. Vehicle ownership rates in patients’ neighborhoods of residence were not related to the likelihood of testing in the ED after accounting for neighborhood income and other factors. Curves of the prevalence of comorbid conditions, as well as proportions of households with neighborhood vehicle access, as a function of age are presented by race/ethnicity in Fig 2 and by levels of neighborhood median income in Fig 3. Increasing age was generally related to increased prevalence of CVD, COPD, T2DM and immune disease (see S1 File). Non-Hispanic Black and Hispanic patients exhibited a higher prevalence of CVD and T2DM, as well as increased obesity (based on body mass index recorded in the electronic health record) prevalence in mid-life. Generally, income gradients in these characteristics were more prominent than gradients corresponding

Table 1. Sociodemographic and clinical characteristics of patients who met study inclusion criteria, by levels of 2018 neighborhood median income corresponding to census tracts of residence in Northeastern Ohio.

<table>
<thead>
<tr>
<th>Neighborhood Median Income (USD, thousands)</th>
<th>p-value2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\leq$25k N = 3891</td>
<td>$25-40k$ N = 31041</td>
</tr>
<tr>
<td>Age (years)</td>
<td>54 (37, 65)</td>
</tr>
<tr>
<td>Female Sex</td>
<td>230 (59%)</td>
</tr>
<tr>
<td>Race/Ethnicity</td>
<td></td>
</tr>
<tr>
<td>Non-Hispanic White</td>
<td>80 (21%)</td>
</tr>
<tr>
<td>Non-Hispanic Black</td>
<td>267 (69%)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>17 (4.4%)</td>
</tr>
<tr>
<td>Other</td>
<td>25 (6.4%)</td>
</tr>
<tr>
<td>Autoimmune Disease</td>
<td>136 (37%)</td>
</tr>
<tr>
<td>Diabetes Mellitus</td>
<td>115 (32%)</td>
</tr>
<tr>
<td>Obesity</td>
<td>158 (41%)</td>
</tr>
<tr>
<td>Current or Former Smoker</td>
<td>227 (58%)</td>
</tr>
<tr>
<td>Asthma</td>
<td>129 (35%)</td>
</tr>
<tr>
<td>Cardiovascular Disease</td>
<td>242 (65%)</td>
</tr>
<tr>
<td>Chronic Obstructive Pulmonary Disease</td>
<td>70 (19%)</td>
</tr>
<tr>
<td>Neighborhood Vehicle Access (% households)</td>
<td>0.63 (0.49, 0.75)</td>
</tr>
</tbody>
</table>

1 Statistics presented: median (IQR); n (%).
2 Statistical tests performed: Kruskal-Wallis test; chi-square test of independence.

https://doi.org/10.1371/journal.pone.0255343.t001
Fig 2. Prevalence of selected comorbid conditions and proportion of households in patients’ respective neighborhood of residence owning at least one vehicle, as a function of age, for groups of patients defined according to race and ethnicity. Estimates adjusted to the observed median neighborhood income of $69,324. Estimates for cardiovascular disease adjusted to modal values of other predictors in the model, i.e., non-diabetic, non-obese, no autoimmune disease, and no documented smoking history. Similarly, estimates for COPD adjusted to no documented smoking history.

https://doi.org/10.1371/journal.pone.0255343.g002

Fig 3. Prevalence of selected comorbid conditions and proportion of households in patients’ respective neighborhood of residence owning at least one vehicle, as a function of age, for groups of patients defined according to selected levels of median neighborhood income. Estimates adjusted to the modal race/ethnicity (non-Hispanic White). Estimates for cardiovascular disease adjusted to modal values of other predictors in the model, i.e., non-diabetic, non-obese, no autoimmune disease, and no documented smoking history. Similarly, estimates for COPD adjusted to no documented smoking history.

https://doi.org/10.1371/journal.pone.0255343.g003
to differences in race/ethnicity. Odds ratio estimates from multivariable models that also adjusted for other risk factors as depicted in Fig 1 are provided in S1 File.

Regarding the latent class analysis to identify relatively homogeneous groups of patients with respect to their symptoms at testing, the 6-class solution optimized the BIC, and entropy and (0.62) was near optimum relative to solutions involving other numbers of classes. Sequential LRT's all were strongly significant (p<0.001) for all solutions up to the 6-class solution, while the LRT for the 7-class solution relative to the 6-class solution was not significant (p = 0.11). These profiles, along with incidence estimates of SARS-CoV-2 test positivity and a summary of sociodemographic factors are provided in Table 2. Class 1 (n = 2510), which had the lowest incidence of SARS-CoV-2 test positivity at 3.4%, was mostly characterized by cough (incidence: 67%) and shortness of breath (99%) and had the highest incidence of testing in the ED (vs. outpatient locations, 60%). Class 1 patients were older (median [IQR] age: 57 [41, 71] years) and had generally more extensive comorbid disease burden than the other classes. Class 2 (n = 3463; 5.4% test positivity) was the largest and generally lower in symptom burden, with moderate incidence of cough (48%) and relatively short duration of symptoms (57% asymptomatic or having symptoms 2 or fewer days). Class 2 had a higher prevalence of non-Hispanic White race/ethnicity (68%), relatively higher neighborhood median income ($71000 [$49000, $95000]) and relatively low rates of presentation to the ED. Class 3 (n = 684; 8.9% test positivity) had a generally higher symptom burden, with particularly higher prevalence of gastrointestinal symptoms (diarrhea, loss of appetite and vomiting) class. Class 3 was largely characteristic of the broader cohort with respect to demographic and clinical factors, and relatively more likely to have been tested in the ED. Class 4 (n = 2955; 11% test positivity) had symptoms most closely matching influenza (cough, fever and flulike symptoms); symptoms were relatively prolonged in this relatively younger class. Class 5 (n = 2593; 16% test positivity) was similar to Class 2, except that the incidence of cough was higher (90%) and accompanied by fatigue (100%) and fever (55%), and Class 2 patients had over twice the rate of test positivity (5.7% and 16% test positivity for Classes 2 and 5, respectively). Class 6 (n = 695; 23% test positivity) had the highest overall symptom burden—all documented symptoms had incidences of at least 59%—and also had the longest duration of symptoms (31% with symptoms for >1 week). These patients were disproportionately non-Hispanic Black (27%, vs. 21% in the overall sample), were more likely to present to the ED (52%), the most likely to have tested positive (23%) and the most likely to have required hospitalization (11% vs. 5.7% for Class 5, which had the next highest test positivity and hospitalization rates).

Curves of the proportion tested in the ED and test positivity as a function of race/ethnicity and neighborhood income are given in Fig 4, and multivariable odds ratio estimates for these outcomes comparing groups defined by race/ethnicity and levels of median neighborhood income (respectively) are given in S1 File. S1 File provides confidence interval estimates for the predicted probabilities given in Fig 4 for selected age values. The analyzed sample of early-pandemic SARS-CoV-2-screened individuals contained too few positive cases to produce reliably similar curves for hospitalization rates among positive cases. Non-Hispanic Black and Hispanic patients, as well as those from lower-income neighborhoods, were more frequently tested in the ED, with the difference in testing location compared to non-Hispanic White and higher-income neighborhoods (respectively) attenuating slightly among individuals of older age. Those whose symptoms were present for more than a week were less likely to present to the ED. Neighborhood vehicle access was not independently related to presentation to the ED.

Non-Hispanic Black patients were between 1.81 [95% confidence interval: 0.91–3.59] times (at age 20) to 2.37 [1.54–3.65] times (at age 80) more likely to test positive for the SARS-CoV-2 virus, while test positivity was not significantly different across the neighborhood income spectrum. Differences in test positivity remained across the six symptom classes after multivariable
adjustment, with the highest risk class (Class 6) being 10.0 [7.4–13.6] times as likely to test positive as the lowest risk class (Class 1).

Among the 1247 patients who tested positive, 404 (32.4%) were hospitalized. Results from the multivariable model for hospitalization were not precise due to the relatively low number
of hospitalizations in the sample, although testing in the emergency department (OR: 5.4 [3.9, 7.5]) and cardiovascular disease diagnosis (OR: 2.5 [1.7, 3.8]) were related to increased risk.

**Discussion**

This study was aimed at identifying socioeconomic and chronic disease-related processes that may influence access to SARS-CoV-2 testing, test positivity, and COVID-19-related hospitalization among a cohort of Northeast Ohio residents who were tested for SARS-CoV-2. The key findings of this study were that tested patients residing in lower-income neighborhoods, who were more likely to be Non-Hispanic Black and/or Hispanic, i) had a higher prevalence of comorbidities such as T2DM and CVD; ii) were more likely to present to the ED for testing than any other location; and iii) were more likely to be hospitalized for COVID-19. Furthermore, non-Hispanic Blacks were more likely to test positive for COVID-19 compared with all others. The increased likelihood of presenting to the ED for testing among Hispanic and non-Hispanic Black patients (compared to non-Hispanic White patients) was more apparent among younger individuals, and appeared to attenuate in the older populations.

Since the beginning of the pandemic, much has been published on the associations of prevalent comorbidities, such as hypertension, T2DM, obesity, coronary artery disease with elevated COVID-19 incidence and case-fatality [50,51]. Such outcomes-based research has allowed policy-makers and public-health specialists to identify individuals who may benefit the most from social distancing and other preventive measures against COVID-19 and highlighted racial/ethnic and socio-economic disparities in healthcare access, and a possible disproportionate impact of the pandemic on persons from racial and ethnic minority backgrounds. The earliest reports came from New York City, where ecological researchers found a disproportionately higher rate of COVID-19 infection and its associated mortality among...
residents of the Bronx, a predominantly lower-income, Hispanic- and non-Hispanic Black-inhabited borough compared with Manhattan, a predominantly middle-to-high income, non-Hispanic White-inhabited borough [5]. Similar trends were reported in other urban hotspots and predominantly non-Hispanic Black-inhabited counties across the United States [52]. Identifying the factors that may be at play behind such glaring healthcare disparities is of paramount importance, with the ultimate aim of addressing them to ensure that all individuals, regardless of their socio-economic background are able to benefit from the advancements of medical care.

The adverse effects of social determinants of health are more pronounced in an airborne pandemic such as COVID-19 [53]. Better living facilities that allow for adequate social distancing, adequate access to healthcare, job and income security providing opportunities for working from home, and other changes to work status and work environments all may impact an individual’s risk for contracting and dying from the disease. Historically, non-Hispanic Black and Hispanic persons have not enjoyed a full share of these opportunities in the social hierarchy of American society. Our study, of a large healthcare system serving a socio-economically diverse population, allowed us to explore individual level differences in race, income, vehicle access and other social determinants and their impact on COVID-19 outcomes. This is in contrast to other ecological studies [5], which have measured associations at population aggregate levels such as counties and boroughs. The finding that socioeconomically disadvantaged, Non-Hispanic Black and Hispanic groups were more likely to present to the ED for testing than any other location may be due to downstream effects of policy that, despite measures such as the Fair Housing Act, have resulted in generatively entrenched, multi-generational disparities that have led to concentrated poverty and hardship. Coincident with lack of access and means for accessing outpatient testing facilities are a lack of healthcare literacy, inadequate of primary care services in these neighborhoods, concerns, perceptions and fears of social biases associated with the healthcare system writ large, and possible implicit bias within the system connecting those who are in need of testing to those who provide testing. Despite the implementation of the Affordable Care Act in Ohio, health insurance inevitably remains tied with employment. It is likely that given the high proportion of non-Hispanic Blacks and Hispanics employed in lower-paying, contingent jobs, the economic burden of the pandemic disproportionately affects lower-income communities and racial and ethnic minorities in particular, rendering workers uninsured and with few options other than emergency care. Recognition of these disparities offers a unique opportunity to target interventions that help vulnerable populations in high need areas, including providing transportation to clinics, food and meal delivery, connections to primary care services, and peer support. Recently published research from our region, that examined patients who utilized a physician-staffed telephone hotline with wrap-around social services, suggests that such approaches have promise for improving health access, reaching patients earlier and reducing social health disparities.

The high test-positivity rate in the ED is likely due (at least in part) to severe regional constraints on the availability of testing early in the pandemic, even to persons presenting to primary care with some symptoms. Nonetheless, the findings that racial/ethnic and economic disparities in testing access, test positivity, as well as hospitalization from COVID-19 were more pronounced in younger individuals compared with older individuals likely reflect a combination of phenomena. First, in a younger population with fewer comorbidities, effects associated with social and economic deprivation may be more pronounced, whereas adverse selection processes (such as premature mortality) may lead to attenuation of observed differences in older individuals. Second, younger persons may have had a higher likelihood of exposure due to working front line jobs and not being designated as an ‘at risk’ group. Third, research suggests that persons in lower-income communities exhibit signs of accelerated or
premature aging, inflammation and immune dysregulation; these vulnerabilities are likely to translate into differences in the severity of COVID-19 disease.

The finding that individuals with a higher prevalence of disease symptoms (class 6) were more likely to be Non-Hispanic Blacks and/or Hispanic, and, in turn, were more likely to test positive and be hospitalized, is especially concerning. What we cannot describe from these data is the degree to which better access to earlier screening and care might have reduced this disparity. Thus, further understanding of processes both endogenous and exogenous to the healthcare system that coalesce to produce racial and economic inequalities in COVID-19 outcomes—and policy changes to ameliorate these inequalities—is critically needed.

Supporting information

S1 File. S1 File contains details on additional supplementary tables and figures and can be found online.

(DOCX)

Author Contributions

**Conceptualization:** Jarrod E. Dalton, Vardhmaan Jain, Adam T. Perzynski, Neal V. Dawson, Yasir Tarabichi, Peter B. Imrey, Michael Lewis, Michael W. Kattan.

**Data curation:** Jarrod E. Dalton, Douglas D. Gunzler.

**Formal analysis:** Jarrod E. Dalton, Douglas D. Gunzler.

**Investigation:** Jarrod E. Dalton.

**Methodology:** Jarrod E. Dalton, Vardhmaan Jain.

**Supervision:** Douglas D. Gunzler.

**Writing – original draft:** Jarrod E. Dalton, Vardhmaan Jain, Adam T. Perzynski, Yasir Tarabichi.

**Writing – review & editing:** Douglas D. Gunzler, Vardhmaan Jain, Adam T. Perzynski, Neal V. Dawson, Douglas Einstadter, Yasir Tarabichi, Peter B. Imrey, Michael Lewis, Michael W. Kattan, James Yao, Glen Taksler, Kristen A. Berg, Nikolas I. Krieger, David Kaelber, Lara Jehi, Ankur Kalra.

References


