Association of COVID-19 With Race and Socioeconomic Factors in Family Medicine

Niharika Khanna, MD, MBBS, DGO, Elena N. Klyushnenkova, PhD, and Alexander Kaysin, MD

**Introduction:** Recent data demonstrated that socioeconomic, environmental, demographic, and health factors can contribute to vulnerability for coronavirus 2019 (COVID-19). The goal of this study was to assess association between severe acute respiratory syndrome coronavirus (SARS CoV-2) infection and demographic and socioeconomic factors in patients from a large academic family medicine practice to support practice operations.

**Methods:** Patients referred for SARS CoV-2 testing by practice providers were identified using shared patient lists in the electronic health records (Epic). The Health Information Exchange (CRISP) was used to identify additional practice-attributed patients receiving testing elsewhere.

**Results:** Compared with white non-Hispanic patients, the odds of COVID-19 detection were higher in black non-Hispanic (odds ratio [OR] = 1.75; 95% CI, 1.18–2.59, \( P = .0052 \)) and Hispanic patients (OR = 5.40; 95% CI, 3.11–9.38, \( P < .0001 \)). The latent class analysis revealed additional patterns in health disparities. Patients living in the areas with Area Deprivation Index 8–10 who were predominantly black had higher risk for SARS CoV-2 infection compared with patients living in less socioeconomically deprived areas who were predominantly white (OR = 1.68; 95% CI, 1.25–2.28; \( P = .0007 \)).

**Conclusion:** Our data provide insight into the association of COVID-19 with race/ethnic minority patients residing in highly socioeconomically deprived areas. These data could impact outreach and management of ambulatory COVID-19 in the future. (J Am Board Fam Med 2021;34:S40–S47.)

**Keywords:** COVID-19 Diagnostic Testing, Family Medicine, Health Information Exchange, Latent Class Analysis, Logistic Models, Minority Groups

**Introduction**
The novel coronavirus disease 2019 (COVID-19) was first encountered and isolated in December 2019 in China,\(^1\) rapidly evolving into an unprecedented global pandemic that has endangered many lives.\(^2\)–\(^7\) During this outbreak, epidemiologic data demonstrate poorer outcomes and higher risk of severe COVID-19 among people aged \( \geq 65 \) years, those with underlying health conditions such as hypertension, cardiovascular disease, chronic lung conditions, diabetes mellitus, obesity, immunodeficiency, cancer, and tobacco use compared with those who are younger and/or without these conditions.\(^2\)–\(^4\),\(^6\),\(^8\),\(^9\) Furthermore, a host of interdependent socioeconomic, environmental, demographic, and health factors likely contribute to gradations of vulnerability for COVID-19.\(^5\),\(^10\) Social determinants of health are known to be powerful influencers of medical illness, behavioral health characteristics, and outcomes.\(^11\)–\(^14\) Higher severe acute respiratory syndrome coronavirus (SARS CoV-2) infection rates among racial and ethnic minority communities are also suggestive of deep-rooted health disparity issues.\(^10\),\(^15\) Data are emerging on the association of race and poverty with COVID-19 in high-density and socioeconomically deprived neighborhoods with high numbers of Medicaid patients.\(^15\)

During SARS CoV-2 outbreak, primary care practices had to reconfigure their workflow to
become responsive to the pandemic. Facing a high volume of patients needing SARS CoV-2 screening, the University of Maryland family medicine and immediate care (UFM) practices utilized the Learning Health System guidance from the National Academies of Science, Engineering, and Medicine to create an adaptive response to COVID-19.\textsuperscript{16,17} As a part of this response, demographic data for all patients undergoing SARS CoV-2 testing were collected. These data have led us to understand that our practices provide screening for COVID-19 to individuals residing in highly socioeconomically deprived neighborhoods and that African Americans or black patients formed the largest number of SARS CoV-2-positive individuals.\textsuperscript{17}

The Area Deprivation Index (ADI) had been recently adopted in several studies as a comprehensive, composite census-based socioeconomic index comprised of 17 elements that measure neighborhood’s socioeconomic disadvantage such as poverty, education, unemployment rates, crime, household composition, median home value, median rent, home ownership, education, and access to a telephone or motor vehicle.\textsuperscript{13,14,18–21} Building on our prior work, the goal of this study was to assess the association between demographic characteristics, Maryland ADI and COVID-19 in patients from the UFM practices during the first 3 months of the pandemic (March 12, 2020, to June 4, 2020).

Methods

The UFM offices are located in urban Baltimore City and suburban Howard County where diverse populations are served. Both sites use the Epic (Verona, WI) electronic health record (EHR), and each site has a colocated primary care and immediate care practice. Collectively, the department serves approximately 10,000 patients with 35,000 visits annually and offers testing for COVID-19 at each site. Samples were collected via nasopharyngeal or nasal swabs and analyzed at licensed commercial and hospital-based labs using SARS-CoV-2 reverse transcriptase–polymerase chain reaction.

Data collection utilized shared patient lists in Epic EHR to identify patients referred for SARS CoV-2 testing by practice providers. The state-designated Health Information Exchange (CRISP) was used to identify patients from our practice panel who received testing elsewhere. Demographic data (age, sex, race, Hispanic ethnicity, street address, ZIP codes) and test results were extracted from Epic EHRs. Race and ethnicity were self-reported and grouped for downstream analyses as (1) African American or black non-Hispanic/unknown ethnicity, (2) white non-Hispanic/unknown ethnicity, (3) Hispanic (regardless of race), and (4) other/unknown race non-Hispanic/unknown ethnicity. Race distribution for Hispanic patients is shown in Supplemental Table 1; race distribution for other group is shown in Supplemental Table 2.

COVID-19 data for the state of Maryland were obtained through the Maryland Department of Health COVID-19 dashboard, and data were downloaded from the Maryland Geographic Information System (GIS) data catalog.\textsuperscript{22,23} Average SARS CoV-2 positivity rate for the State of Maryland was estimated based on daily testing volumes and number of positive tests between March 23, 2020, and June 4, 2020. Demographic characteristics of the UFM patients were obtained from the claims data for the year preceding COVID-19 pandemic (February 1, 2019, to January 31, 2020).

To characterize socioeconomic status (SES), we used the ADI derived from the neighborhood atlas database.\textsuperscript{13,19} The ADIs are constructed based on the 2010 census data using Census Block Groups with a unique 12-digit Federal Information Processing Standards (FIPS) code. In the database, state-specific ADIs are expressed as deciles and are constructed by ranking the ADI from lowest (1, least disadvantaged) to highest (10, most disadvantaged) for each state. The neighborhood atlas database also includes 9-digit ZIP codes (5 + 4) matched to the 12-digit FIPS code, which allowed for assigning ADI ranks to individual patients based on their street address. For the ZIP codes (5 + 4) that were not included into the database, patient-level ADI ranks were imputed using median ADI from nearest neighborhoods that have the same first 7 digits (5 + 2) ZIP codes (n = 88) or median for corresponding 12-digit FIPS code (n = 8). For patients with invalid addresses, ADIs were assigned based on available (5 + 4) digit ZIP codes for the nearest building in the vicinity (n = 4).

Descriptive statistics data were obtained using SAS 9.4 software (SAS Institute Inc, Cary, NC). Associations between groups of comparison were assessed using a $\chi^2$ test. Multivariable binary logistic regression model was used to assess association between COVID-19 and age, sex, race/ethnicity groups and State of Maryland ADI rank. Statistical significance was established at $\alpha = 0.05$. Latent class
analysis (LCA) for correlated categorical variables (race/ethnicity and ADI) was performed using JMP Pro 13 software (SAS Institute Inc). After preliminary data exploration, 2 classes were prespecified. The most likely cluster for each participant was determined using mixture probabilities of the cluster, determining the highest probability of membership.24

Results

COVID-19 Prevalence in Maryland

Between March 23, 2020, and June 4, 2020, the results for 400,437 SARS CoV-2 tests were reported to the Maryland Department of Health electronically, and 66,168 positive cases were identified (unadjusted
positivity rate = 16.52%, 95% CI, 16.41–16.64). Among positive cases identified during this period, 28.7% were African American or black, 19.5% were white non-Hispanic, and 25.7% were Hispanic.

University Family Medicine Patient Characteristics

In a year preceding pandemic (February 1, 2019, to January 31, 2020), 86,843 invoices for 24,441 patients were recorded at the UFM practices. On average, the UFM patients were 37.1 ± 19.4 years old, 64.2% were females, 53.5% were black, and 30.5% were white.

A descriptive summary of demographic characteristics for patients tested for SARS CoV-2 in the UFM clinics between March 12, 2020, and June 4, 2020, is shown in Table 1. Among 1781 tested patients, the average age was 43.2 ± 16.4 years, 69.7% were females, 59.5% were black, 26.5% were white non-Hispanic, 4.7% were Hispanic, and 49.9% were residing in areas with state of Maryland ADI 8–10 (Table 1).

Overall, 272 of patients (15.3%) were positive for SARS CoV-2 (95% CI, 13.5–17.0%). Among positive cases, 80.1% were between 25 and 64 years old, 72.8% were females, 64.7% were black, 16.2% were white non-Hispanic, 11.4% were Hispanic, and 58.1% lived in areas with ADI 8–10 (Figure 1 and data not shown).

SARS CoV-2 Testing and the Impact of Age and Gender

The highest SARS CoV-2 infection rate was observed in 34- to 45-year-old patients and was slightly higher in females compared with males; however, the associations between COVID-19 and either age or sex were not statistically significant (Table 1). Association between age and sex was also not statistically significant (Supplemental Table 3).

SARS CoV-2 Testing and the Impact of Race/Ethnicity and Socioeconomic Deprivation

The SARS CoV-2 positivity rate was highest in Hispanic patients (36.9%) followed by black non-Hispanic (16.6%, Table 1). The SARS CoV-2 positivity rate was highest for patients living in areas with ADI 8–10 (Table 1). There was a strong correlation between race/ethnicity and ADI, with 69.6% of black patients living predominantly in the highly deprived areas with ADI 8–10 and white patients living in areas with ADI 5–7 or 1–4 (Table 2). The majority of other non-Hispanic patients (84.3%) also lived in areas with ADI 1–4 or 5–7 (Table 2). The distribution of Hispanic patients was spread approximately evenly across all ADI ranks; however, 41.7% of these patients lived in areas with ADI 8–10 (Table 2).

In the multivariable logistic regression model, only race/ethnicity but not ADI rank, age, or sex was significantly associated with COVID-19. The interaction terms between age and sex as well as race/ethnicity and ADI were also not statistically significant and were excluded from the final model (data not shown). Regardless of age, sex, and ADI rank, the odds of SARS CoV-2 infection were 1.8 times higher in black compared with white patients (odds ratio [OR] = 1.75; 95% CI, 1.18–2.59, \( P = .0052 \)). The odds of SARS CoV-2 infection among Hispanic patients were 5.4 times higher compared with white non-Hispanic (OR = 5.40; 95% CI,
3.11–9.38, \( P < .0001 \)). The odds of SARS CoV-2 infection among other non-Hispanic patients were higher compared with white; however, the difference was not statistically significant (OR = 1.40; 95% CI, 0.80–2.44, \( P = .2287 \)).

**Latent Class Analysis**

We also conducted the LCA using ADI ranks as an ordinal variable with 3 levels (1–4, 5–7, and 8–10) and race/ethnicity as a nominal variable with 4 levels as defined in Table 1. We tested models with 2 to 4 classes and found that a model with 2 classes had the best fit based on minimizing values for the Bayes Information Criteria and Akaike’s Information Criteria (data not shown). Cluster 1 was identified as predominantly black living in highly deprived areas. Cluster 2 was identified as a predominantly white living in areas with low or intermediate ADI ranks. Cluster membership proportion were 0.66 and 0.34 for Cluster 1 and Cluster 2, respectively.

Parameter estimates for 2 classes are shown in Figure 2. Given Cluster 1 membership, probabilities for a patient to live in highly deprived areas (ADI 8–10) were 0.72, to live in the areas with ADI 5–7 were 0.20, and to live in areas with ADI 1–4 were 0.08; probabilities of being black were 0.87, being white were 0.07, being Hispanic were 0.04, and have other or unknown race were 0.02 (Figure 1 and data not shown). Given Cluster 2 membership, probabilities for a patient to live in the areas with ADI 8–10 were 0.07, to live in the areas with ADI 5–7 were 0.32, and to live in areas with ADI 1–4 were 0.61; probabilities of being black were 0.08, being white were 0.62, being Hispanic were 0.06, and have other or unknown race were 0.23 (Figure 2 and data not shown).

After latent class membership was identified, we treated it as a categorical independent variable in a logistic regression model with SARS CoV-2 positivity rate as a dependent variable. Initial analysis indicated that the effects of age and sex were not statistically significant (data not shown). The odds of COVID-19 were 1.7 times higher for Cluster 1 members who were predominantly black patients living in highly deprived areas compared with Cluster 2 members who were predominantly white patients living in areas with low or moderate level of socioeconomic deprivation (OR = 1.68; 95% CI, 1.25–2.28; \( P = .0007 \)).

**Discussion**

An emerging role for primary care during the pandemic is to provide ambulatory management of COVID-19, with outreach and remote patient monitoring for home-dwelling patients and seniors. During the first 3 months of this pandemic, the family medicine ambulatory practices were on the front line.

An unadjusted estimate for the proportion of SARS CoV-2 cases (15.3%) in our study population was similar to the estimates obtained by Martinez et al25 for the same population in the Baltimore-Washington area (16.3%) as well as our estimate for the state of Maryland, 16.5% daily average positivity rate during the same time period, with overlapping 95% confidence intervals.

Consistent with UFM practice locations in Baltimore and the overall UFM patients’ demographics, black patients were the largest group tested for the SARS CoV-2 (59.5%) followed by white patients (29.7%). The proportion of Hispanic patients tested during the study period was relatively small (4.7%), reflective of their low overall proportion of total patients in our catchment areas. State-specific data for SARS CoV-2 testing volumes by race and ethnicity, especially during the first months of the outbreak, are sparse. Among SARS CoV-2–positive cases, black patients were overrepresented, and Hispanic patients were underrepresented in our study population compared with the statewide race/ethnicity distribution22,23.

Our data support that racial/ethnic minorities have a higher risk of COVID-19 compared with non-Hispanic white and other nonblack racial groups. The unadjusted estimates for the proportion of SARS CoV-2–positive patients by race/ethnicity were comparable with estimates obtained by Martinez et al25 for the same population, although patients from that study were sicker based on a high hospitalization rate for SARS CoV-2–positive patients (35.9%). Patients with COVID-19 in our data were predominantly ambulatory with a hospitalization rate of approximately 8.3%, and 1 confirmed COVID-attributed death (unpublished observation).

Based on the Centers for Disease Control and Prevention data for the first 6 months of this pandemic, COVID-19 prevalence was estimated to be 2.6 times higher in black non-Hispanic and 2.8 times higher in Hispanic persons compared with white non-Hispanics.3 The estimates for black patients in our study population were comparable with the nationwide data, whereas estimates for Hispanic patients were higher. This could be due to the different time frame for the available data and small number of Hispanic patients in our study population as well as due to low SES, nonconformity to preventive practices, more limited access, and distrust of health care institutions.15 Further studies are needed to unveil the reasons for health disparities in Hispanic patients in our practices.

Historically, Baltimore is one of the geographic areas in the United States that has marked geographic segregation of ethnic/racial groups owing to structural racism such as discriminatory housing policies.26 In the multivariate logistic regression model, we did not detect a significant association between COVID-19 and ADI after adjustment for race/ethnicity. An alternative approach using the LCA identified health disparity patterns based on the race and SES that were not revealed by logistic regression analysis. In our study, a cluster of patients who were predominantly black and lived in highly SES-deprived areas had higher risk of COVID-19 compared with a cluster of patients who were predominantly white and lived in areas with higher SES. These data are in concordance with our previous observation about spatial race and ADI distribution of SARS CoV-2–positive cases among Baltimore suburban and urban populations.17 High ADI communities with a disproportionate number of racial/ethnic minorities are known to experience marked disparities in health and SES.24 Higher risk for COVID-19 in the underserved communities reflect the underlying inequities in health, income, access to government resources and participation, incarceration, and education.11

In our study population, age and gender did not have a significant association with SARS CoV-2 infection rate, although other studies have demonstrated greater propensity for infection and adverse outcomes among older patients.27–29 This finding may be due in part to the fact that our data were collected from community-dwelling individuals, whereas published data may reflect institutional living in long-term care facilities and existing chronic conditions. In addition, based on our empiric observation, the entire UFM practice population may be skewed younger than the community.

Our study has several limitations, including the use of observational data from clinical sources. Data were collected from a single practice with 4 sites and represent a convenience sample of patients who were seeking SARS CoV-2 testing or developed symptoms that required medical intervention. The UFM practice followed the Centers for Disease Control and Prevention data for the COVID-19 response in Maryland, screening and testing approximately 2000 patients for SARS CoV-2 using reverse transcriptase–polymerase chain reaction by mid-June 2020.
Control and Prevention guidelines to prioritize SARS CoV-2 testing to high-risk populations in the early months of the pandemic. Race and ethnicity were self-reported, which could have led to underestimation of the proportion of ethnic minorities, especially Hispanic, in the study population. The ADI ranks were computed based on 2010 Census data and were based on the neighborhood characteristics rather than SES for individual patients. In addition, we captured only a limited number of demographic and SES characteristics. Other risk factors, including underlying chronic medical conditions, should be taken into consideration in the future studies.

Given the expected higher prevalence of underlying risk factors in underserved communities, there may be a higher probability of poor outcomes from COVID-19. Therefore, the role of primary care would include not only SARS CoV-2 testing and management but also management of chronic conditions and preventive care delivery supported by telehealth. In addition, there is a need to reinvent the practice-patient relationship, with a high degree of patient outreach, remote patient monitoring, phone supports, and telehealth. Patients living in high ADI communities, especially the high risk, elderly population, are more likely to lack internet connectivity and devices to connect with their primary care practices and to self-manage chronic conditions, which must be considered and addressed in the design of interventions. There is also a need to develop payment models to support the management of defined populations with the use of remote patient monitoring and centralized monitoring by practices using novel workflows. New paradigms of care delivery need to be taught to medical students and residents to ensure that the pipeline for medical professionals is prepared to manage the challenges of the COVID-19 pandemic. Lastly, health care policy makers and regulators need to work together with primary care groups to optimize the implementation of new technologies, care delivery, and payment models to entrench these new systems of care beyond temporary emergency authorizations.

Conclusion

Neighborhood deprivation provides insight into the needs of SARS CoV-2–positive patients and supports a better understanding of the characteristics of COVID-19 spread in white, black and Hispanic populations. In addition, these data provided our practices with information necessary to tap into health system and state-funded resources for community outreach and COVID-19 prevention. Policy-relevant observations from this population of COVID-19 patients will be provided to health system leaders, local policy makers and regulators to enable public health programming to better address community needs.

To see this article online, please go to: http://jabfm.org/content/34/Supplement/S40.full.

References

23. Maryland’s GIS Data Catalog. Available at: https://data.imap.maryland.gov/.