Clinical Infectious Diseases
MAJOR ARTICLE

County-level Predictors of Coronavirus Disease 2019 (COVID-19) Cases and Deaths in the United States: What Happened, and Where Do We Go from Here?

John M. McLaughlin, Farid Khan, Sarah Pugh, Frederick J. Angulo, Heinz-Josef Schmitt, Raul E. Isturiz, Luis Jodar, and David L. Swerdlow

Pfizer Vaccines, Collegeville, Pennsylvania, USA

Background. The United States has been heavily impacted by the coronavirus disease 2019 (COVID-19) pandemic. Understanding microlevel patterns in US rates of COVID-19 can inform specific prevention strategies.

Methods. Using a negative binomial mixed-effects regression model, we evaluated the associations between a broad set of US county-level sociodemographic, economic, and health status–related characteristics and cumulative rates of laboratory-confirmed COVID-19 cases and deaths between 22 January 2020 and 31 August 2020.

Results. Rates of COVID-19 cases and deaths were higher in US counties that were more urban or densely populated or that had more crowded housing, air pollution, women, persons aged 20–49 years, racial/ethnic minorities, residential housing segregation, income inequality, uninsured persons, diabetics, or mobility outside the home during the pandemic.

Conclusions. To our knowledge, this study provides results from the most comprehensive multivariable analysis of county-level predictors of rates of COVID-19 cases and deaths conducted to date. Our findings make clear that ensuring that COVID-19 preventive measures, including vaccines when available, reach vulnerable and minority communities and are distributed in a manner that meaningfully disrupts transmission (in addition to protecting those at highest risk of severe disease) will likely be critical to stem the pandemic.

Keywords. risk-factors; disparities; vulnerable populations; transmission; vaccine distribution.

The coronavirus disease 2019 (COVID-19) pandemic, caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), is ongoing, and the United States has been heavily impacted. The US population, however, is geographically and sociodemographically diverse, and understanding microlevel patterns in rates of COVID-19 cases and deaths can inform specific prevention strategies and the titration of public health responses at the federal, state, and local levels. This need is heightened as the US economy and schools begin reopening and daily life gets back on track against the backdrop of uncertainty about whether a resurgence of COVID-19 will emerge with the upcoming flu season.

Although previous studies have evaluated the impact of various sociodemographic or environmental factors on the risk of developing or dying from COVID-19 (eg, race/ethnicity [1–11], poverty [2], air pollution [12], mobility [13], population density [14], chronic medical conditions [15–18]), these factors have largely been examined in isolation. Moreover, most analyses were conducted early in the pandemic. The Centers for Disease Control and Prevention (CDC) recently presented preliminary data that describe the association between an aggregated "social vulnerability index" and the likelihood of becoming a CDC-designated COVID-19 "hot spot" [19]. However, additional comprehensive evaluations of COVID-19 disease trends are needed to inform future public health strategies against the complexities of COVID-19. To help pinpoint prevention strategies, including vaccination once available, we evaluated the associations between a broad set of county-level environmental, sociodemographic, economic, and health status–related characteristics on rates of COVID-19 cases and deaths in the United States.

METHODS

Outcome Data

We obtained county-level records of the cumulative number of COVID-19 laboratory-confirmed cases and deaths from the Johns Hopkins University Coronavirus Resource Center available between 22 January 2020 and 31 August 2020. This source tracks and makes publicly available county-level COVID-19 data reported by the CDC and state health departments. Cumulative county-level rates of COVID-19 cases and deaths through 31 August 2020 were expressed per 100 000 county residents.

Received 21 October 2020; editorial decision 9 November 2020; published online 19 November 2020.

Correspondence: John M. McLaughlin, Pipeline Vaccines, Pfizer, Inc, 500 Arcola Rd., Collegeville, PA 19426 (e. john.mclaughlin@pfizer.com).

Clinical Infectious Diseases® 2021;73(7):e1814–21

DOI: 10.1093/cid/ciaa1729

Exposure Data

County-level environmental, sociodemographic, economic, and health status characteristics hypothesized to be associated with transmission or mortality of COVID-19 were obtained from several publicly available databases maintained by the US government or private institutions. These data were collated and then combined with Johns Hopkins county-level COVID-19 data to form the analysis database. Environmental factors included population density, urbanicity, residential crowding (housing with >1 person per room [20]), and air pollution (particles per million [PPM]). Sociodemographic and economic variables included gender, age, race/ethnicity, a residential housing segregation index (0–100 scale, with 100 being most segregated counties between Whites and non-Whites [21]), high school education status, unemployment status, state-adjusted median household income, and income inequality (ratio of household incomes at the 80th vs the 20th percentile [22]). Health status–related variables included prevalence of diabetes, obesity, smoking, and, as a potential indicator of risky close-contact behavior, rates of sexually transmitted infections (STIs) [23]. Finally, as a proxy for adherence to stay-at-home orders and recommendations to minimize travel [24], we obtained Google community mobility reports that describe percent change in county-level travel to nonresidential locations during the pandemic compared with a prepandemic baseline period [25]. The baseline period was defined as the median value from the 5-week period between 3 January 2020 and 6 February 2020 [25]. A list of all exposure variables, including definitions and data sources, is provided in Supplementary Table 1.

Statistical Analyses

County-level characteristics were summarized with descriptive statistics. Missing county-level characteristics (in <1% of the US population) were imputed using state-level values (Supplementary Table 1). Google mobility data, when missing from the least-populous counties due to privacy concerns, were not imputed (Supplementary Table 1). Using the *menbreg* command in Stata version 14.0 (StataCorp LLC, College Station, TX), we fit negative binomial mixed-effects regression models (which allow for overdispersion) [26] to estimate county-level predictors of cumulative rates of COVID-19 cases and deaths. To estimate rates, we modeled cumulative cases and deaths by county, controlling for county population size as an independent variable. For all models, we included state (n = 51; 50 states and the District of Columbia) as a group-level random intercept to account for potential correlation in counties within the same state (eg, state-level testing practices, lockdown measures, and other health-related, social, and cultural differences). Because exposure variables were likely to independently predict COVID-19 rates and confound the relationship between one another, we constructed univariate and multivariable models. If a large change in point estimates occurred between univariate and multivariable models, we constructed stepwise parsimonious models to understand which covariates were key confounders. We assessed multicollinearity using variance inflation factors (VIFs) to ensure multivariable models were not overfitted.

RESULTS

County Characteristics

Between 22 January 2020 and 31 August 2020, the numbers of laboratory-confirmed COVID-19 cases and deaths in the United States were 5 916 357 and 180 886, respectively. Cases across 3142 US counties ranged from 0 to 241 768, with Los Angeles County, California, having the most (4% of all US cases). Only 41 of 3142 (1%) counties reported no cases. No deaths were reported in 686 of 3142 counties (22%); however, these counties made up only 3% of the US population. The most deaths, 7290, occurred in Kings County, New York. Table 1 summarizes county characteristics. Google mobility data were not available for 309 of 3142 (10%) counties, which accounted for <1% of the US population.

Rates of COVID-19 Cases

County-level rates of COVID-19 cases ranged from 0 to 14 338 per 100 000 persons, with mean = 1422 (95% confidence interval [CI] = 1377–1466) and median = 1059 with interquartile range (IQR) = 568–1897 (Table 1). The highest COVID-19 rate occurred in Trousdale County, Tennessee, driven by an outbreak of >1300 cases at a prison [27]. Overall, 33 of 51 (65%) and 44 of 51 (86%) states had ≥1 county in the top decile and quartile of rates, respectively. Supplementary Table 2 compares county characteristics by quartiles of COVID-19 rates.

In univariate results, counties with higher proportions/rates of population density, urbanicity, crowded housing, air pollution, females, persons aged 30–49 years, racial/ethnic minorities, residential housing segregation, adults without a high school degree, obesity, STIs, and travel outside the home during the pandemic had higher rates of COVID-19 cases (all $P < .05$; Table 2). Counties with higher proportions of adults aged 50–64 and ≥80 years, diabetes, and who had higher household income had lower rates at the univariate level. Multivariable models (n = 2833 when restricted to counties with Google mobility data; 51 states) that adjusted for all exposure variables simultaneously revealed generally similar trends to univariate results; however, the magnitude of some variables (ie, population density, crowded housing) was reduced in multivariable models (Table 2; Supplementary Table 3). Additionally, while significant in univariate results, in the multivariable model, the following were no longer related to COVID-19 rates and seemed to be explained by other factors in the model: Asian race, age groups 50–64 and ≥80 years, high school education, obesity, and STI rates (Table 2). Supplementary Table 3 shows stepwise

Table 1. Summary of County-Level Characteristics Across 3142 US Counties

County-level Characteristic	Mean (Standard Deviation)	Median (Interquartile Range)	Min.	Max.
Outcome variables (22 January 2020–31 August 2020)				
Laboratory-confirmed COVID-19 cases	1883.0 (8111.5)	295.0 (83.0 to 968.0)	0	241 768
Rate of laboratory-confirmed COVID-19 cases	1421.7 (1277.1)	1058.9 (567.6 to 1896.9)	0	14 339
Laboratory-confirmed COVID-19 deaths	57.6 (310.0)	5.0 (1.0 to 22.0)	0	7290
Rate of laboratory-confirmed COVID-19 deaths	33.5 (46.7)	17.2 (3.5 to 44.0)	0	461
Environmental exposure variables				
Population size	104 468 (333 457)	25 726 (10 901 to 68 098)	86	10 039 107
Population density (persons per square mile of land)	272.7 (1785.8)	44.8 (16.5 to 118.6)	0	71 341
Percent urban[a]	41.3 (31.5)	40.5 (11.5 to 66.6)	0	100
Percent living in crowded housing (>1 person per room [20])	2.4 (2.4)	1.9 (1.2 to 2.9)	0	52
Air pollution (parts per million)[b]	8.9 (2.1)	9.3 (7.6 to 10.4)	0	20
Sociodemographic and economic exposure variables				
Percent female	49.9 (2.3)	50.3 (49.4 to 51.0)	27	57
Percent aged 0–19 years	24.4 (3.6)	24.4 (22.3 to 26.3)	0	45
Percent aged 20–29 years	12.2 (3.1)	11.7 (10.4 to 13.0)	0	37
Percent aged 30–49 years	23.3 (2.7)	23.2 (21.7 to 24.7)	12	38
Percent aged 50–64 years	20.3 (2.4)	20.5 (19.1 to 21.8)	7	31
Percent aged 65–79 years	14.9 (3.6)	14.6 (12.8 to 16.7)	3	46
Percent aged ≥80 years	4.8 (1.6)	4.6 (3.8 to 5.6)	0	24
Percent White	76.0 (20.2)	83.4 (64.3 to 92.3)	3	98
Percent Black	9.0 (14.3)	2.2 (0.7 to 10.2)	0	85
Percent Asian	1.6 (3.0)	0.7 (0.5 to 1.4)	0	43
Percent other race	2.5 (7.8)	0.7 (0.4 to 1.5)	0	93
Percent Hispanic	9.7 (13.8)	4.4 (2.4 to 10.0)	1	96
Residential housing segregation scale (0–100, with 100 being most segregated between Whites and non-Whites [21])[c]	32.4 (13.4)	32.0 (23.3 to 41.6)	0	90
Percent without high school degree[d]	11.4 (7.1)	10.3 (6.4 to 15.0)	0	74
Percent unemployed	4.0 (1.5)	3.7 (3.0 to 4.6)	1	19
Median household income (in 2019 dollars)[e]	52 794 (13 880)	50 568 (43 681 to 58 848)	25 385	140 382
Percentage of median state household income[e]	89.4 (20.1)	86.9 (76.2 to 99.2)	44	264
Income inequality ratio (comparing 80th percentile of household income vs 20th percentile [22]) [f]	4.5 (0.8)	4.4 (4.0 to 4.9)	3	12
Percent uninsured[g]	13.6 (6.2)	12.5 (8.6 to 17.4)	3	42
Health status to related variables				
Percent with diabetes	12.1 (4.1)	11.6 (9.2 to 14.5)	2	34
Percent obese	32.9 (5.5)	33.1 (29.2 to 36.5)	12	58
Percent current smokers	17.5 (3.6)	17.0 (14.9 to 19.7)	6	41
Rate of sexually transmitted infections per 1000 persons[h]	4.1 (2.8)	3.4 (2.3 to 5.0)	0	61
Travel outside the home during pandemic[i]				
Percent change in travel outside the home during the pandemic compared with prepandemic baseline	–12.3 (10.4)	–11.9 (–18.6 to –6.1)	–67	43

Abbreviation: COVID-19, coronavirus disease 2019.

[a]Urbanicity was missing for 7 counties (<1% of US population), which were imputed using state-level values.

[b]Air pollution was missing for 34 counties (<1% of US population), which were imputed using state-level values.

[c]Residential housing segregation scale was missing for 351 counties (<1% of US population), which were imputed using state-level values.

[d]High school education status was missing for 96 counties (<1% of US population), which were imputed using state-level values.

[e]Annual household income was missing for 1 county (<1% of US population), which was imputed using state-level values.

[f]Income inequality was missing for 2 counties (<1% of US population), which were imputed using state-level values.

[g]Health insurance status was missing for 1 county (<1% of US population), which was imputed using state-level values.

[h]Sexually transmitted infection rate was missing for 152 counties (<1% of US population), which were imputed using state-level values.

[i]n = 2833. Google mobility data were not available for 309 of 3142 (10%) counties (due to privacy concerns in less-populous counties), which accounted for <1% of the US population. These missing values were not imputed.

modeling for independent variables with large changes in the point estimate between univariate and fully adjusted models (ie, population density, crowded housing, and Asian race) to elucidate which other covariates were key confounding factors in these instances.

The strongest predictors of COVID-19 rates in the multivariable model were higher proportions of persons aged 30–49 years (incidence rate ratio [IRR] = 3.17; 95% CI = 2.48–4.05 for each 10% increase) and persons aged 20–29 years (IRR = 2.18; 95% CI = 1.76–2.70 for 10% increase) vs persons aged 0–19 years, uninsured (IRR = 1.70; 95% CI = 1.49–1.94 for 10% increase), women (IRR = 1.59; 95% CI = 1.31–1.93 for 10% increase), crowded housing (IRR = 1.57; 95% CI = 1.24–2.00 for 10% increase), population density (IRR = 1.51; 95% CI = 1.38–1.64 for highest quartile vs lowest 3 quartiles), and travel outside the home during the pandemic (IRR = 1.38; 95% CI = 1.34–1.42 for 10% increase; Table 2). Additionally, for each 1 PPM increase in air pollution or 10% increase in county-level urbanicity, income, proportion racial/ethnic minorities, residential housing segregation, income inequality, or diabetes, COVID-19 rates were 1.09–1.24 times higher in the multivariable model (all $P < .05$; Table 2).

Rates of COVID-19 Deaths

County-level rates of COVID-19 deaths ranged from 0 to 461 per 100 000 (Table 1), with the highest rate in Hancock

Table 2. County-level Characteristics Associated With Rates of Laboratory-Confirmed Coronavirus Disease 2019 Cases and Deaths Through 31 August 2020 in Univariate and Multivariable-Adjusted Mixed-Effects Negative Binomial Regression Models

	Univariate Model (n = 3142)				Multivariable, Final Model (n = 2833)[a]			
	Cases		Deaths		Cases		Deaths	
County-level Characteristic	IRR	95% CI	IRR	95% CI	IRR	95% CI	IRR	95% CI
Environmental								
Highest quartile of population density (vs lowest 75%)	3.61	3.23–4.04	3.42	3.00–3.90	1.51	1.38–1.64	1.46	1.29–1.65
10% increase in proportion living in urban area[b]	1.38	1.36–1.40	1.35	1.33–1.37	1.11	1.09–1.13	1.12	1.09–1.15
10% increase in proportion living in crowded housing (>1 person per room [20])	4.26	3.24–5.59	4.16	3.07–5.63	1.57	1.24–2.00	1.73	1.23–2.45
1 part per million increase in air pollution[c]	1.52	1.46–1.58	1.60	1.53–1.68	1.24	1.21–1.28	1.31	1.26–1.37
Sociodemographic and economic								
10% increase in proportion female	1.43	1.19–1.72	2.72	2.20–3.35	1.59	1.31–1.93	2.73	2.06–3.62
10% increase in proportion aged 20–29 years	0.95	.76–1.19	0.76	.57–1.02	2.18	1.76–2.70	2.09	1.53–2.86
10% increase in proportion aged 30–49 years	1.63	1.25–2.12	1.09	.78–1.52	3.17	2.48–4.05	3.47	2.42–4.97
10% increase in proportion aged 50–64 years	0.18	.14–.25	0.22	.15–.33	1.05	.82–1.36	1.65	1.13–2.40
10% increase in proportion aged 65–79 years	1.17	.93–1.46	0.90	.67–1.22	1.70	1.39–2.08	1.69	1.24–2.29
10% increase in proportion aged ≥80 years	0.04	.02–.06	0.08	.04–.15	0.74	.47–1.16	2.12	1.07–4.23
10% increase in proportion Black	1.15	1.10–1.20	1.20	1.15–1.26	1.09	1.05–1.13	1.16	1.10–1.22
10% increase in proportion Asian	3.92	2.88–5.33	2.74	1.97–3.81	1.09	.94–1.26	1.21	.98–1.50
10% increase in proportion Native American or Hawaiian/Other Pacific Islander	1.08	1.01–1.14	1.22	1.13–1.32	1.07	1.00–1.14	1.28	1.15–1.41
10% increase in proportion Hispanic	1.33	1.27–1.39	1.27	1.21–1.34	1.17	1.12–1.23	1.24	1.16–1.31
10-unit increase in residential housing segregation scale (0–100, with 100 being most segregated between Whites and non-Whites [21])[d]	1.07	1.03–1.11	1.15	1.10–1.20	1.10	1.07–1.13	1.11	1.08–1.15
10% increase in proportion without a high school degree[e]	1.45	1.34–1.57	1.50	1.37–1.66	1.03	.98–1.09	1.03	.95–1.11
10% increase in proportion unemployed	0.76	.56–1.03	1.84	1.27–2.69	1.11	.84–1.47	1.48	.98–2.22
10% increase in state-adjusted household income[f]	1.14	1.11–1.16	1.08	1.05–1.10	1.11	1.08–1.14	1.10	1.05–1.14
1-unit increase in income inequality ratio (comparing 80th percentile of household income vs 20th percentile [22])[g]	1.04	.98–1.11	1.20	1.12–1.30	1.10	1.05–1.16	1.15	1.06–1.24
10% increase in proportion without health insurance[h]	1.11	.99–1.25	1.18	1.03–1.36	1.70	1.49–1.94	1.48	1.22–1.78
Health-status related								
10% increase in prevalence of diabetes	0.71	.63–.80	0.89	.77–1.03	1.12	1.03–1.22	1.13	.99–1.29
10% increase in prevalence of obesity	1.18	1.08–1.29	1.36	1.23–1.51	1.04	.97–1.12	1.11	1.00–1.22
10% increase in prevalence of current smoking	1.19	1.00–1.42	1.47	1.20–1.81	0.79	.64–.97	0.74	.55–1.01
1 per 1000 increase in rates of sexually transmitted infections[i]	1.17	1.14–1.19	1.18	1.15–1.21	1.02	1.00–1.04	1.01	.98–1.03
Travel outside the home during pandemic[a]								
10% increase in travel outside the home during the pandemic based on Google mobility data	1.57	1.51–1.63	1.51	1.44–1.59	1.38	1.34–1.42	1.38	1.32–1.45

Abbreviations: CI, confidence interval; IRR, incidence rate ratio.

[a]n = 2833. Google mobility data were not available for 309 of 3142 (10%) counties (due to privacy concerns in less-populous counties), which accounted for <1% of the US population. These missing values were not imputed and counties with missing Google mobility data were not included in the multivariable, final model.

[b]Urbanicity was missing for 7 counties (<1% of US population), which were imputed using state-level values.

[c]Air pollution was missing for 34 counties (<1% of US population), which were imputed using state-level values.

[d]The residential housing segregation scale was missing for 351 counties (<1% of US population), which were imputed using state-level values.

[e]High school education status was missing for 96 counties (<1% of US population), which were imputed using state-level values.

[f]Annual household income was missing for 1 county (<1% of US population), which was imputed using state-level values.

[g]Income inequality was missing for 2 counties (<1% of US population), which were imputed using state-level values.

[h]Health insurance status was missing for 1 county (<1% of US population), which was imputed using state-level values.

[i]Sexually transmitted infection rate was missing for 152 counties (<1% of US population), which were imputed using state-level values.

County, Georgia, driven by nursing home outbreaks in a rural, predominately minority, and underserved community [28]. Supplementary Table 4 compares county characteristics by quartiles of mortality rate.

In univariate results, all county-level variables except diabetes prevalence were related to mortality rates (Table 2). The final multivariable model of mortality was similar to the model that predicted rates of confirmed cases with a notable exception. Namely, in addition to higher proportions of adults aged 20–29 years (IRR = 2.09; 95% CI = 1.53–2.86 for 10% increase) and 30–49 years (IRR = 3.47; 95% CI = 2.42–4.97 for 10% increase) being related to higher mortality rates (as with rates of confirmed cases), a 10% increase in the proportion of adults aged 50–64, 65–79, and ≥80 years (vs 0–19 years), while not related to rates of COVID-19 cases, was also associated with 1.7–2.1 times higher mortality rates. In addition to age, other county-level predictors strongly related to mortality were increasing proportions of females (IRR = 2.73; 95% CI = 2.06–3.62 for 10% increase), crowded housing (IRR = 1.73; 95% CI = 1.23–2.45 for 10% increase), uninsured adults (IRR = 1.48; 95% CI = 1.22–1.78 for 10% increase), higher population density (IRR = 1.46; 95% CI = 1.29–1.65 for highest quartile vs lowest 3 quartiles), and more travel outside the home during the pandemic (IRR = 1.38; 95% CI = 1.32–1.45 for 10% increase; Table 2; Supplementary Table 3). VIFs for variables included in multivariable models (for cases and deaths) were all <3 with mean <2, suggesting no evidence of multicollinearity.

DISCUSSION

To our knowledge, this study provides results from the most comprehensive multivariable analysis of county-level predictors of rates of COVID-19 cases and deaths conducted to date. Our findings, current through the end of August 2020, have significant implications for COVID-19 prevention strategies, including vaccination. While many county-level factors were related to COVID-19 rates, there are 2 key takeaways from our research.

First, our findings confirm and expand on earlier reports [1–12] and preliminary data from the CDC [19] that the pandemic has taken a disproportionate toll on minority and other vulnerable [29] US populations. Specifically, rates of COVID-19 cases and deaths were higher in counties with more racial/ethnic minorities, residential housing segregation, income inequality, uninsured persons, air pollution, and adults with diabetes. Our findings on this topic, however, are novel in that they confirm these disparities exist even after adjustment for other potentially confounding factors. For example, even after adjustment for mobility during the pandemic, population density, urbanicity, crowded housing, age, education, employment, and health insurance status and for the prevalence of diabetes, obesity, and smoking, for every 10% increase in the proportion of a US county that was Black or Hispanic, there was a corresponding 9% and 17% increase in the rate of COVID-19 cases and a 16% and 24% increase in mortality, respectively. Compounding this, more residential housing segregation and income inequality were both independently related to higher county-level rates of cases and deaths. These findings confirm that there may be larger structural forces behind racial/ethnic differences in COVID-19 rates beyond the factors we measured, and this warrants continued research.

Recent reports have highlighted that many of the vulnerable populations we identified as being at increased risk for COVID-19 (eg, minorities, uninsured, and those without a high school degree) disproportionately serve in "essential" pandemic front-line jobs (eg, grocery clerks, food and agriculture jobs, facilities and janitorial workers, and social services) [7, 30–32]. These jobs often cannot be done at home, which increases workplace exposure to SARS-CoV-2 [5, 6]. Indeed, we confirmed that counties with more travel outside the home during the pandemic had higher rates of COVID-19 cases and deaths. Future studies should evaluate the link between vulnerable and minority communities and workplace exposure with individual-level data, and more studies of occupation-specific risks for COVID-19 are needed. In the near term, redirecting public health resources (eg, testing, contact tracing, ensuring safe working conditions, health promotion and education efforts, and eventually vaccination) to vulnerable and minority communities and to communities with a disproportionate share of "essential" workers is likely warranted. A leading example includes the Rapid Acceleration of Diagnostics in Underserved Populations initiative, launched by the National Institutes of Health, that provides support to expand availability, accessibility, and acceptance of SARS-CoV-2 testing for underserved and vulnerable populations [33]. This strategy, in addition to mitigating exposure to individuals at highest risk of severe disease (eg, frail elderly, nursing homes) [34], may be an additional way to help stem the pandemic.

Although our finding that counties with higher state-adjusted household incomes had higher rates of COVID-19 cases and deaths initially seemed counterintuitive to our other findings that highlight vulnerable communities, several potential explanations for this exist. For example, COVID-19 hit coastal counties, where incomes are highest, especially hard early on. Further, nursing home death rates were also especially high among high-income states on the East Coast [35]. Additionally, there may be better access to testing (and thus more confirmed cases) in areas with higher income [36]. Another possibility is that county-level income inequality, rather than income level alone, may better predict vulnerable communities, as we found that higher county-level income inequality predicted higher COVID-19 rates. This is consistent with previous reports that showed that even within counties with high median household incomes, vulnerable pockets of communities with more economic and social stress and less access to medical care can exist and often experience disparate health outcomes [37]. This finding ultimately suggests that identification of populations at increased risk for COVID-19 is multifaceted and that a multivariable approach like ours or a multidimensional risk-score approach (as is being explored by the CDC [19]) will be needed to accurately pinpoint areas at high risk of becoming COVID-19 hotspots.

Our second major finding was that our study confirms anecdotal reports [38] that efforts to interrupt COVID-19 transmission, including with vaccination when available, may be as equally impactful on mortality as is protecting individuals at highest risk for severe disease (eg, the elderly and those with comorbidities [39, 40]). Specifically, we identified several county-level factors (eg, population density, urbanicity, crowded housing, and mobility outside of the home during the pandemic) that independently predicted county-level COVID-19 mortality rates, despite not being related to COVID-19 case fatality or the development of severe disease [39, 40]. One interpretation is that COVID-19 has hit hardest in communities where adhering to social distancing guidelines may be more difficult due to high population density, an urban setting (with potentially more reliance on public transit and multiunit housing), or crowded living arrangements (eg, multigenerational families [41]). These readily available metrics could be used to prioritize early vaccination efforts when the number of doses may be limited. Moreover, while it was perhaps not surprising that counties with more persons aged 20–49 years seemed to have higher rates of COVID-19 illness (given presumably more exposure or a perceived lower risk for severe disease and thus taking social distancing guidelines less seriously), it was unexpected that higher proportions of persons aged 20–49 years also predicted higher county-level mortality rates. Because individual-level case fatality rates are markedly lower in this age group [42], this finding suggests that adults aged <50 years are likely driving transmission (and thus indirectly impacting county-level mortality rates). Similarly, although individual-level reports have previously identified men as being at increased risk of developing severe COVID-19 [43], we unexpectedly found that counties with more women had higher rates of COVID-19 cases and deaths. Future studies should also explore the role of women in driving transmission (eg, disproportionately working in healthcare or other "essential" jobs [30] or caring for children or other family members during the pandemic). Finally, while the proportion of children aged <20 years was not related to higher rates of COVID-19, this age group will be returning to daycare and school and engaging in more extracurricular activities over the coming months. Thus, their role in determining COVID-19 rates should be continuously monitored to further elucidate the role children play in driving community-level disease rates and the impact that interrupting transmission in this age group might have [44].

It remains unclear whether communities with higher COVID-19 rates to date would again be at highest risk during a potential resurgence this fall or winter or if herd immunity in these communities is approaching levels needed to meaningfully slow transmission [45, 46]. For example, a recent report suggested that in some hard-hit, vulnerable communities in New York City, antibody levels could already be >50% [47]. Thus, despite our findings to date, it is also possible that communities with lower rates of COVID-19 until now may be more susceptible (because of lower levels of immunity) to future waves of COVID-19. However, while it was hypothesized that communities first hit hard in the spring during the H1N1 influenza pandemic would be less likely to experience a subsequent "second wave" during the following influenza season (due to higher levels of herd immunity), this was not the case, suggesting that elevated spring illness did not protect against an autumn resurgence [48]. Thus, continuous monitoring of whether the same trends in COVID-19 rates we report here are observed throughout the rest of 2020 may be an indication of the level of immunity in communities that have been most susceptible to date.

Our study was ecological, and our findings should be confirmed with individual-level data. We did not have county-level data about specific social distancing measures such as mask-wearing; bar, restaurant, and retail closures; and other local-level restrictions on large gatherings. However, we included county-level data that describe mobility during the pandemic, which is a proxy for social distancing measures [13]. Another limitation is that our data, apart from our outcome variables and Google mobility data, were historical. Thus, data about unemployment and health insurance status, household income, and other sociodemographic and environmental factors did not necessarily reflect the situation during the pandemic. Additionally, not all exposure data came from the same year. However, we obtained the most-recent estimates from all data sources, and most of the data that describe county-level characteristics were based on estimates from the last 2 years. When modeling COVID-19 mortality rates, 22% of counties reported no deaths. These counties, however, accounted for only 3% of the US population. Moreover, negative binomial regression models, which we used in our analysis, allow for overdispersion (which can result from excess zeros) and straightforward interpretation and have been shown to model count data with zeros as well as other zero-inflated Poisson models [49]. Finally, we did not have data that described county-level SARS-CoV-2 testing practices. Vulnerable communities, which had higher COVID-19 rates in our study, have historically had reduced access to healthcare [8] and to SARS-CoV-2 testing [36]. Thus, disparities in COVID-19 rates among vulnerable and minority communities could be more pronounced after adjusting for local testing practices. Lower testing rates in minority neighborhoods [36] may also explain why we saw more pronounced racial/ethnic differences in mortality compared with rates of confirmed cases. More research about community-specific testing and its impact on disparities in COVID-19 rates is needed.

Our study gives a comprehensive, granular, and contemporary overview of which areas were most affected by COVID-19 in the United States through the summer of 2020. While the outbreak has now spread across the entire country at a macro level without a great deal of discrimination, microlevel county-by-county disparities in how the pandemic spread were more pronounced. A vaccine is likely the only alternative to balancing restrictive measures, such as forced lockdowns and closures to protect vulnerable and minority populations who have been disproportionately impacted by the COVID-19 pandemic to date, and the dire economic consequences these measures often bring to the same working class communities. Our findings make clear that ensuring that COVID-19 preventive measures, including vaccines when available, reach vulnerable and minority communities and are distributed in a manner that meaningfully disrupts transmission (in addition to protecting those at highest risk of severe disease) will likely be critical to stem the pandemic. Historically speaking [38, 50, 51], this too will be a formidable public health challenge.

Supplementary Data

Supplementary materials are available at *Clinical Infectious Diseases* online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

Notes

Financial support. This study was sponsored by Pfizer Inc.

Potential conflicts of interest. All authors are employees and shareholders of Pfizer, Inc. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. Millett GA, Jones AT, Benkeser D, et al. Assessing differential impacts of COVID-19 on black communities. Ann Epidemiol **2020**; 47:37–44.
2. Adhikari S, Pantaleo NP, Feldman JM, Ogedegbe O, Thorpe L, Troxel AB. Assessment of community-level disparities in coronavirus disease 2019 (COVID-19) infections and deaths in large US metropolitan areas. JAMA Netw Open **2020**; 3:e2016938.
3. Price-Haywood EG, Burton J, Fort D, Seoane L. Hospitalization and mortality among black patients and white patients with Covid-19. N Engl J Med **2020**; 382:2534–43.
4. Moore JT, Ricaldi JN, Rose CE, et al; COVID-19 State, Tribal, Local, and Territorial Response Team. Disparities in incidence of COVID-19 among underrepresented racial/ethnic groups in counties identified as hotspots during June 5-18, 2020–22 states, February-June 2020. MMWR Morb Mortal Wkly Rep **2020**; 69:1122–6.
5. Tenforde MW, Billig Rose E, Lindsell CJ, et al; CDC COVID-19 Response Team. Characteristics of adult outpatients and inpatients with COVID-19—11 academic medical centers, United States, March-May 2020. MMWR Morb Mortal Wkly Rep **2020**; 69:841–6.
6. Azar KMJ, Shen Z, Romanelli RJ, et al. Disparities in outcomes among COVID-19 patients in a large health care system in California. Health Aff (Millwood) **2020**; 39:1253–62.
7. Bui DP, McCaffrey K, Friedrichs M, et al. Racial and ethnic disparities among COVID-19 cases in workplace outbreaks by industry sector—Utah, March 6-June 5, 2020. MMWR Morb Mortal Wkly Rep **2020**; 69:1133–8.
8. Thakur N, Lovinsky-Desir S, Bime C, et al. The structural and social determinants of the racial/ethnic disparities in the U.S. COVID-19 pandemic: what's our role? Am J Respir Crit Care Med **2020**; 202:943–9. doi:10.1164/rccm.202005-1523PP
9. Yancy CW. COVID-19 and African Americans. JAMA **2020**; 323:1891–2.
10. Yehia BR, Winegar A, Fogel R, et al. Association of race with mortality among patients hospitalized with coronavirus disease 2019 (COVID-19) at 92 US hospitals. JAMA Network Open **2020**; 3:e2018039-e.
11. Wortham JM, Lee JT, Althomsons S, et al. Characteristics of persons who died with COVID-19—United States, February 12-May 18, 2020. MMWR Morb Mortal Wkly Rep **2020**; 69:923–9.
12. Wu X, Nethery RC, Sabath BM, Braun D, Dominici F. Exposure to air pollution and COVID-19 mortality in the United States: a nationwide cross-sectional study. medRxiv **2020**: doi:2020.04.05.20054502.
13. Badr HS, Du H, Marshall M, Dong E, Squire MM, Gardner LM. Association between mobility patterns and COVID-19 transmission in the USA: a mathematical modelling study. Lancet Infect Dis **2020**; 20:1247–54.
14. Hamidi S, Sabouri S, Ewing R. Does density aggravate the COVID-19 pandemic? J Am Plann Assoc **2020**; 86:495–509. doi:10.1080/01944363.2020.1777891
15. Centers for Disease Control and Prevention Covid-19 Response Team. Preliminary estimates of the prevalence of selected underlying health conditions among patients with coronavirus disease 2019—United States, February 12-March 28, 2020. MMWR Morb Mortal Wkly Rep **2020**; 69:382–6.
16. Garg S, Kim L, Whitaker M, et al. Hospitalization rates and characteristics of patients hospitalized with laboratory-confirmed coronavirus disease 2019—COVID-NET, 14 States, March 1-30, 2020. MMWR Morb Mortal Wkly Rep **2020**; 69:458–64.
17. Suleyman G, Fadel RA, Malette KM, et al. Clinical characteristics and morbidity associated with coronavirus disease 2019 in a series of patients in metropolitan Detroit. JAMA Netw Open **2020**; 3:e2012270.
18. Stokes EK, Zambrano LD, Anderson KN, et al. Coronavirus disease 2019 case surveillance—United States, January 22-May 30, 2020. MMWR Morb Mortal Wkly Rep **2020**; 69:759–65.
19. Wallace M. Disparities in COVID-19 incidence, severity, and outcomes. Available at: https://www.cdc.gov/vaccines/acip/meetings/downloads/slides-2020-09/COVID-05-Wallace.pdf. Accessed 1 September 2020.
20. Blake K, Kellerson R, Simic A. Measuring overcrowding in housing. Fairfax, VA: U.S. Department of Housing and Urban Development, Office of Policy Development and Research September, **2007**.
21. County Health Rankings & Roadmaps. Residential segregation—white vs non-white. Available at: https://www.countyhealthrankings.org/explore-health-rankings/measures-data-sources/county-health-rankings-model/health-factors/social-and-economic-factors/family-social-support/residential-segregation-non-whitewhite. Accessed 1 September 2020.
22. County Health Rankings & Roadmaps. Income inequality. Available at: https://www.countyhealthrankings.org/explore-health-rankings/measures-data-sources/county-health-rankings-model/health-factors/social-and-economic-factors/income/income-inequality. Accessed 1 September 2020.
23. Balestri R, Magnano M, Rizzoli L, Infusino SD, Urbani F, Rech G. STIs and the COVID-19 pandemic: the lockdown does not stop sexual infections [manuscript published online ahead of print 11 July 2020]. J Eur Acad Dermatol Venereol **2020**; doi:10.1111/jdv.16808.
24. Google LLC. Google COVID-19 community mobility reports. Available at: https://www.google.com/covid19/mobility/. Accessed 7 September 2020.
25. Google LLC. Understand the data. Available at: https://support.google.com/covid19-mobility/answer/9825414?hl=en&ref_topic=9822927. Accessed 1 September 2020.
26. Zhang X, Mallick H, Tang Z, et al. Negative binomial mixed models for analyzing microbiome count data. BMC Bioinformatics **2017**; 18:4.
27. Parker C. Why Trousdale & Bledsoe counties have highest Covid-19 rates in the country. Available at: https://wdef.com/2020/05/12/trousdale-bledsoe-counties-highest-covid-19-rates-country/. Accessed 1 September 2020.
28. Hart A. COVID-19 swamps another rural Georgia county. Available at: https://www.ajc.com/news/state--regional/covid-swamps-another-rural-georgia-county/2KxYEndBNpQDEnvb1NSbSN/. Accessed 1 September 2020.
29. American Hospital Association. Task Force on Ensuring Access in Vulnerable Communities. Available at: https://www.aha.org/system/files/content/16/ensuring-access-taskforce-report.pdf. Accessed 8 October.
30. Rho HJ, Brown H, Fremstad S. A basic demographic profile of workers in frontline industries. Washington, DC: Center for Economic and Policy Research, **2020**.
31. Hawkins D. Differential occupational risk for COVID-19 and other infection exposure according to race and ethnicity. Am J Ind Med **2020**; 63:817–20.

32. Oliver SE. Epidemiology of COVID-19 in essential workers, including healthcare personnel. Available at: https://stacks.cdc.gov/view/cdc/91979. Accessed 1 September 2020.
33. National Institutes of Health. NIH to assess and expand COVID-19 testing for underserved communities. Available at: https://www.nih.gov/news-events/news-releases/nih-assess-expand-covid-19-testing-underserved-communities. Accessed 1 September 2020.
34. Hatfield KM, Reddy SC, Forsberg K, et al. Facility-wide testing for SARS-CoV-2 in nursing homes—seven U.S. jurisdictions, March-June 2020. MMWR Morb Mortal Wkly Rep **2020**; 69:1095–9.
35. Centers for Medicare & Medicaid Services. COVID-19 nursing home data, submitted data as of week ending: 10/18/2020. Available at: https://data.cms.gov/stories/s/COVID-19-Nursing-Home-Data/bkwz-xpvg/. Accessed 4 November 2020.
36. Lieberman-Cribbin W, Tuminello S, Flores RM, Taioli E. Disparities in COVID-19 testing and positivity in New York City. Am J Prev Med **2020**; 59:326–32.
37. Woolf SH, Braveman P. Where health disparities begin: the role of social and economic determinants—and why current policies may make matters worse. Health Aff (Millwood) **2011**; 30:1852–9.
38. Finn A, Malley R. A vaccine that stops Covid-19 won't be enough. Available at: https://www.nytimes.com/2020/08/24/opinion/coronavirus-vaccine-prevention.html. Accessed 1 September 2020.
39. Fried MW, Crawford JM, Mospan AR, et al. Patient characteristics and outcomes of 11 721 patients with COVID19 hospitalized across the United States. Clin Infect Dis **2020**; 72:e558–65.
40. Killerby ME, Link-Gelles R, Haight SC, et al; Centers for Disease Control and Prevention COVID-19 Response Clinical Team. Characteristics associated with hospitalization among patients with COVID-19—metropolitan Atlanta, Georgia, March-April 2020. MMWR Morb Mortal Wkly Rep **2020**; 69:790–4.
41. Lovett I, Frosch D, Overberg P. Covid-19 stalks large families in rural America: remote regions with crowded households have turned deadlier than some city blocks. Available at: https://www.wsj.com/articles/covid-19-households-spread-coronavirus-families-navajo-california-second-wave-11591553896. Accessed 1 September 2020.
42. Centers for Disease Control and Prevention. COVID-19 hospitalization and death by age. Available at: https://www.cdc.gov/coronavirus/2019-ncov/covid-data/investigations-discovery/hospitalization-death-by-age.html. Accessed 1 September 2020.
43. Griffith DM, Sharma G, Holliday CS, et al. Men and COVID-19: a biopsychosocial approach to understanding sex differences in mortality and recommendations for practice and policy interventions. Prev Chronic Dis **2020**; 17:E63.
44. Laxminarayan R, Wahl B, Dudala SR, et al. Epidemiology and transmission dynamics of COVID-19 in two Indian states. Science **2020**; 370:691–7. doi:10.1126/science.abd7672
45. Britton T, Ball F, Trapman P. A mathematical model reveals the influence of population heterogeneity on herd immunity to SARS-CoV-2. Science **2020**; 369:846–9.
46. Sette A, Crotty S. Pre-existing immunity to SARS-CoV-2: the knowns and unknowns. Nat Rev Immunol **2020**; 20:457–8.
47. Goldstein J. 1.5 million antibody tests show what parts of N.Y.C. were hit hardest. Available at: https://www.nytimes.com/2020/08/19/nyregion/new-york-city-antibody-test.html. Accessed 10 September 2020.
48. Burkom H, Kniss K, Meltzer M, et al. Investigating the effect of high spring incidence of pandemic influenza A(H1N1) on early autumn incidence. Epidemiol Infect **2012**; 140:2210–22.
49. Allison P. Logistic regression using SAS: theory and application, 2nd edition (Chapter 9). Thousand Oaks, CA: SAGE Publications, Inc., **2014**.
50. Lu PJ, O'Halloran A, Williams WW, Lindley MC, Farrall S, Bridges CB. Racial and ethnic disparities in vaccination coverage among adult populations in the U.S. Am J Prev Med **2015**; 49:S412–25.
51. McLaughlin JM, Swerdlow DL, Khan F, et al. Disparities in uptake of 13-valent pneumococcal conjugate vaccine among older adults in the United States. Hum Vaccin Immunother **2019**; 15:841–9.