



Identification of county-level health factors associated with COVID-19 mortality in the United States

Wei Pan^{1,2,✉}, Yasuo Miyazaki³, Hideyo Tsumura¹, Emi Miyazaki¹, Wei Yang⁴

¹School of Nursing, Duke University, Durham, NC 27710, USA;

²Department of Population Health Sciences, School of Medicine, Duke University, Durham, NC 27708, USA;

³School of Education, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA;

⁴School of Community Health Sciences, University of Nevada, Reno, NV 89557, USA.

Abstract

Many studies have investigated causes of COVID-19 and explored safety measures for preventing COVID-19 infections. Unfortunately, these studies fell short to address disparities in health status and resources among decentralized communities in the United States. In this study, we utilized an advanced modeling technique to examine complex associations of county-level health factors with COVID-19 mortality for all 3141 counties in the United States. Our results indicated that counties with more uninsured people, more housing problems, more urbanized areas, and longer commute are more likely to have higher COVID-19 mortality. Based on the nationwide population-based data, this study also echoed prior research that used local data, and confirmed that county-level sociodemographic factors, such as more Black, Hispanic, and older subpopulations, are attributed to high risk of COVID-19 mortality. We hope that these findings will help set up priorities on high risk communities and subpopulations in future for fighting the novel virus.

Keywords: COVID-19, mortality, health factors, health disparity, hierarchical generalized linear model

Introduction

COVID-19 is a global pandemic. As of the date of starting writing this manuscript (30 June, 2020), there were over 10.4 million total confirmed cases and 511 000 total deaths among 188 infected countries/regions, especially in the United States (US) whose numbers reached 2.6 million of total confirmed cases and 127 000 of total deaths^[1]. These numbers are continuing to grow. This public health crisis posed disproportional health outcomes related to COVID-19 across various racial/ethnic groups with unequal

socioeconomic status at different geographical locations in the US.

According to the County Health Rankings Model^[2], the contributing health factors associated with inequalities in COVID-19 outcomes are multidimensional: Health behaviors, clinical care, socioeconomics, and physical environment (**Fig. 1**). These contributing health factors can also exist at multiple levels within the US: individuals, communities/counties, and states. Many studies have investigated individual demographic and clinical factors of COVID-19 outcomes since the inception of

✉ Corresponding author: Wei Pan, Duke University, DUMC 3322, 307 Trent Dr., Durham, NC 27710, USA. Tel/Fax: +1-919-684-9324/+1-919-684-8899, E-mail: wei.pan@duke.edu.

Received 01 August 2020, Revised 28 August 2020, Accepted 07 September 2020, Epub 30 September 2020

CLC number: R511, Document code: A

The authors reported no conflict of interests.

This is an open access article under the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited.

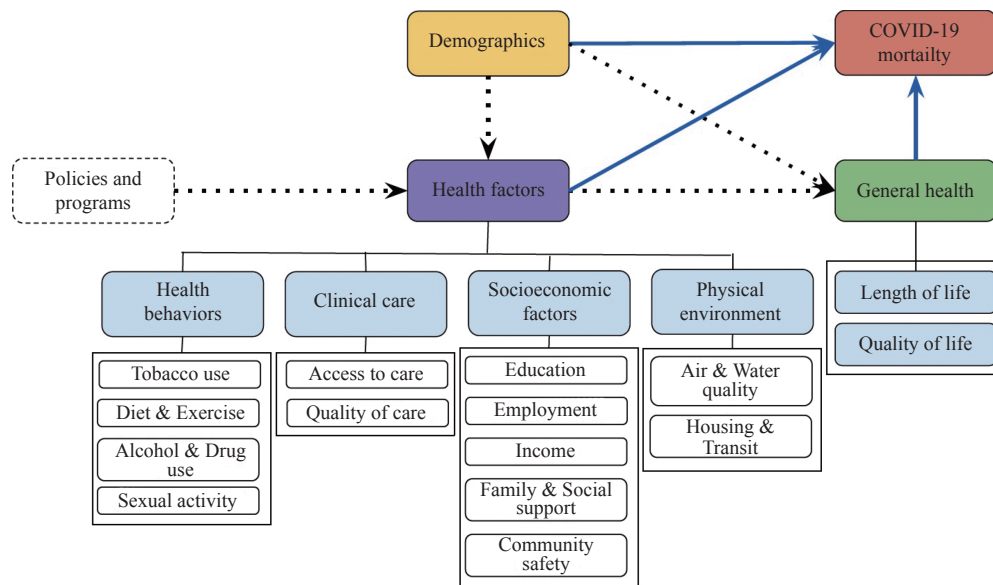


Fig. 1 A conceptual model adopted from the county health rankings model^[2]. The relationships indicated by dotted lines were not examined due to the scope of this study.

COVID-19 outbreaks in January 2020 in the US. For example, a review study by Tian and colleagues on 14 studies found that individuals with older age, male sex, and certain co-morbidities (e.g., hypertension, diabetes, and other chronic diseases) were at a high risk of dying from COVID-19^[3].

The aggregation of these contributing factors at higher levels (e.g., community/county-level and state-level) are also significantly associated with COVID-19 outcomes. For instance, community-level social determinants, such as severe overcrowding, lower educational status, higher unemployment rates, less access to healthcare, and more chronic diseases, were associated with a higher rate of COVID-19 infection among districts in New York City^[4] and more COVID-19 deaths among counties in Colorado^[5]. Cyrus and colleagues^[6], and Yancy^[7] also found that communities and counties with more underrepresented minorities, especially African Americans, had a higher prevalence of COVID-19. These findings were confirmed by many other studies at county-level on the same topics^[8–10]. Moving to a higher level, state-level, the conclusion is the same. For example, Oronce and colleagues found that state-level income inequality was positively related to COVID-19 cases and mortality^[11]. By adjusting 322 counties' racial composition, state-level racial/ethnic disparities still resulted in COVID-19 mortality 80% higher in African Americans than White^[12].

These consistent findings from the studies at all levels, especially at county- and state-levels, are not unexpected. COVID-19 responses and decision making in the US remain decentralized^[10]. As shown

in the conceptual model (**Fig. 1**), variations in health policies and programs at county- and state-levels may influence health factors and, in turn, general health at those levels; which will ultimately have various impacts on populations' risks and capability to respond to COVID-19. As a matter of fact, local governments and their demographic characteristics, especially at the county-level, had been widely examined before COVID-19 outbreaks for other health-related outcomes, such as infectious disease^[13], vaccine^[14], obesity^[15], opioid^[16], and hospital performance^[17].

Because COVID-19 is a novel virus, besides demographic and socioeconomic factors, other contextual factors, such as health behaviors, clinical care, and physical environment (**Fig. 1**), may provide a new insight or better understanding of the places and people that could be most affected by the COVID-19 pandemic so as to inform policy makers to take actions to improve community conditions^[2]. Taking the factors of clinical care as an example, access to care (e.g., ventilator availability) would make a difference in outcomes of this novel COVID-19^[18]. Because treatments and vaccines for COVID-19 will not be ready or available for a certain period of time^[19], exploring other means of preventing and controlling COVID-19 spreading among communities becomes ever important. Therefore, a comprehensive examination of various county-level health factors associated with COVID-19 outcomes is warranted and imperative.

Although some prior studies investigated the associations between county- or state-level characteristics and COVID-19 outcomes, they either

examined a small number of counties^[4-6,8,12]; analyzed only a small set of county- or state-level characteristics, mainly focused on racial compositions and socioeconomics^[6,8-9,11-12]; or modeled county- and state-level factors separately^[4-6,9-11]. In this study, we aimed to comprehensively identify multidimensional county-level health factors associated with COVID-19 mortality for all 3141 counties among all 50 states plus Washington, DC in the US, accounting for variations of both county- and state-level demographic characteristics and general health.

Materials and methods

Data source

The data for this study included measures of county-level health factors and general health as well as demographic characteristics at both county- and state-levels. Those health factors, general health, and demographic characteristics were extracted from the County Health Rankings^[2]. In addition, to more accurately measure the rurality of counties, a key sociodemographic variable, we supplemented the measure of the proportion of rural areas in the County Health Rankings with the Rural-Urban Continuum Codes from the US Department of Agriculture Economic Research Service^[20].

The study data also included the total numbers of the confirmed COVID-19 cases and deaths from the onset date to 31 May, 2020 for each county. These COVID-19 outcomes were computed based on New York Times' COVID-19 Data on GitHub^[21] and cross-validated against Johns Hopkins University's COVID-19 Case Tracker^[1] and the Centers for Disease Control and Prevention's COVID-19 Cases in the US^[22], as well as some Google searches if necessary. From the late May 2020, some states had just started lifting the stay-at-home order and reopening their economies in phases. Thus, our data would reveal a general picture of the COVID-19 pandemic before the US economy was reopened.

Variables

COVID-19 outcomes

The COVID-19 outcomes were the total numbers of the confirmed COVID-19 cases and deaths from the onset date to 31 May, 2020 for each county. Because the total number of COVID-19 cases tend to be underestimated due to various reasons, such as the availability of COVID-19 testing and asymptomatic cases, the COVID-19 mortality, defined as the total number of deaths from COVID-19 per 100 000

population, was analyzed as the outcome in the study.

Country- and state-level demographics

In this study, demographic variables at both county- and state-levels were the same. They were age composition (proportion of below 18 years of age and proportion of 65 and older), gender (proportion of female), racial/ethnic composition (proportions of non-Hispanic White, non-Hispanic Black, Hispanic, Asian, Indian & Alaska Native, and Native Hawaiian/Other Pacific Islander—those six proportions add up to 100%), and an index of the rurality of the country (proportion of rural).

In addition, country-level demographics were supplemented with the Rural-Urban Continuum Codes^[20] which captured not only the population size but also the proximity to large cities. The higher score of the rural-urban continuum code was, the more rural the county was.

County-level health factors and general health

As indicated in the conceptual model (**Fig. 1**), the county-level health factors considered in this study were health behaviors, including tobacco use (1 item), diet and exercise (4 items), alcohol and drug use (2 items), and sexual activity (2 items); clinical care, including access to care (3 items) and quality of care (3 items); socioeconomic factors, including education (2 items), employment (1 item), income (2 items), family and social support (2 items), and community safety (2 items); and physical environment, including air and water quality (2 items), and housing and transit (3 items). The general health analyzed in the study were length of life (4 items), and quality of life (4 items). See the County Health Rankings^[2] for more information about the item-level measures.

Statistical analysis

Descriptive statistics were first used to summarize the demographic characteristics of all 3141 counties across all 50 states plus Washington, DC in the US. Then, we utilized hierarchical generalized linear models (HGLM)^[23] to identify county-level health factors associated with COVID-19 mortality for the 3141 counties, accounting for county-level general health and both county- and state-level demographic characteristics.

The use of HGLM was determined based on the following two considerations. First, the data were in a nested structure (*i.e.*, counties were nested within states), and thus observations on counties were likely to be not independent. In this data situation, hierarchical linear models (HLM)^[23] (aka multilevel

modeling or mixed models) were appropriate. Second, the outcome variable in this study, COVID-19 mortality, was directly derived from a count variable, the number of deaths. In other words, it took values of non-negative integers. Thus, a Poisson distribution was appropriate for analyzing this type of outcome variables. Therefore, HGLM with Poisson distribution was chosen for this study.

It is worth noting that COVID-19 mortality also exhibited more variation than a regular Poisson distribution, whose variance equals its mean, could capture. To address this issue, an extra parameter, overdispersion, needed be considered to accommodate the inflated variability. Therefore, HGLM with Poisson distribution with overdispersion was finally chosen to analyze the data. In terms of the estimation method, the default option of the penalized restricted quasi-likelihood estimation was elected in the HLM/HGLM software^[24].

Results

County- and state-level demographic characteristics

The descriptive statistics of demographic characteristics at county- and state-levels are provided in **Table 1**. From **Table 1**, it can be observed that the

average population for counties was 104 159 ranging from as few as 88 people to more than 10 million. At the state-level, the average population was more than 6.4 million ranging from about 0.6 million to almost 40 million. Of those populations, on average, a typical county had more younger (≤ 18 years) population (22%) than older (≥ 65 years) population (19%) so did a typical state (22% vs. 16%); whereas both counties and states had an equal gender distribution. In terms of the racial/ethnic composition, the proportion of non-Hispanic White was the majority for both counties (76%) and states (68%), followed by Hispanic (10% for counties and 12% for states) and then non-Hispanic Black (9% and 11%, respectively). Other racial/ethnic groups (*i.e.*, Asian, American Indian & Alaska Native, and Native Hawaiian/Other Pacific Islander) made of the rest ($< 5\%$ for counties and $< 8\%$ for states). It is worth noting that in the HGLM analysis, we included the proportions of non-Hispanic Black and of Hispanic as the groups of interest and the combined remaining four groups as the reference category the majority ($\sim 90\%$) of which was non-Hispanic White.

As for rurality, a typical county in general had more than a half of its area (59%) as rural; whereas a typical state only had a little more than a quarter of its area (26%) as rural. The average score of the rural-urban

Table 1 Descriptive statistics of demographic characteristics and COVID-19 outcomes

Characteristic	County-level (n=3141)				State-level (n=51)			
	Min.	Max.	Mean	SD	Min.	Max.	Mean	SD
Population size	88	10 105 518	104 159.48	333 534.49	577 737	39 667 045	6 415 047.73	7 343 307.89
Proportion of below 18 years of age	<0.01	0.42	0.22	0.03	0.18	0.30	0.22	0.02
Proportion of 65 and older	0.05	0.58	0.19	0.05	0.11	0.21	0.16	0.02
Proportion of female	0.27	0.57	0.50	0.02	0.48	0.53	0.51	0.01
Proportion of non-Hispanic White	0.03	0.98	0.76	0.20	0.22	0.93	0.68	0.16
Proportion of non-Hispanic Black	<0.01	0.85	0.09	0.14	0.01	0.45	0.11	0.11
Proportion of Hispanic	0.01	0.96	0.10	0.14	0.02	0.49	0.12	0.10
Proportion of Asian	<0.01	0.43	0.02	0.03	0.01	0.38	0.05	0.06
Proportion of American Indian & Alaska native	<0.01	0.93	0.02	0.08	<0.01	0.15	0.02	0.03
Proportion of native Hawaiian/Other Pacific Islander	0	0.49	<0.01	0.01	<0.01	0.10	<0.01	0.14
Proportion of rural	<0.01	1.00	0.59	0.31	<0.01	0.61	0.26	0.15
Rural-urban continuum code	0	8	4.01	2.71	–	–	–	–
Number of COVID-19 deaths	0	6732	33.17	241.29	8	29 699	2043.75	4528.34
Number of COVID-19 deaths per 100 000 population (COVID-19 mortality)	0	311.98	12.85	28.48	1.08	151.97	26.20	33.61

SD: standard deviation.

continuum code was 4.01 for counties, suggesting that most of counties were "nonmetro counties with an urban population of 20 000 or more, adjacent to a metro area". The rural-urban continuum codes were not available at the state-level.

COVID-19 mortality

Table 1 also displays the summary statistics of COVID-19 outcomes. As of 31 May, 2020, among all 3141 counties, the total number of COVID-19 deaths was 104 183, and the average number of COVID-19 deaths was 33.17, ranging from 0 to 6732; while for all 50 states plus Washington, DC, the average number of COVID-19 deaths was 2043.75, ranging from 8 to 29 699. To make the numbers of COVID-19 deaths more comparable across counties and across states, both of which had various sizes of population, a standardized measure, COVID-19 mortality, defined as the number of COVID-19 deaths per 100 000 population, was computed. Among the counties, the average COVID-19 mortality was 12.85 per 100 000 population; while the average COVID-19 mortality for states was more than doubled as 26.20 per 100 000 population. **Fig. 2** also demonstrated that the distribution of COVID-19 mortality was widely spread with several areas of high mortality in various regions across the country.

Associations between county-level demographics and COVID-19 mortality

To examine the raw associations between county-

level demographics and COVID-19 mortality, we first established an HGLM base model which included demographic variables only. The base model set a stage for later identifying health factors uniquely associated with COVID-19 mortality. The results of the base model were presented in **Table 2**. As expected, the overdispersion parameter estimate (*i.e.*, county-level residual variance σ^2) was 18.90, meaning that the estimated within-states variance of COVID-19 mortality given the demographics was 18.90 times larger than the variance expected from a regular Poisson model, which justified the use of the overdispersion parameter. Further, the between-states variance (*i.e.*, state-level residual variance τ) was 0.75 which was statistically significant ($P < 0.001$), meaning that COVID-19 mortality varied significantly across states.

In order to compare the substantive importance of impacts of demographics, the estimated regression coefficients were converted to a half-standardized event rate ratio (ERR)^[25]. The half-standardized ERR is a measure of effect size, and its interpretation is similar to odds ratios. That is, if a half-standardized ERR is greater than 1, the expected death rate is higher in the group of interest than the reference group; otherwise, the expected death rate is equal or lower in the group of interest than the reference group. Then, its absolute percent (%) of change was ranked at the last column of **Table 2** for indicating the substantive importance of each demographic variable.

The rank ordering by the absolute % change

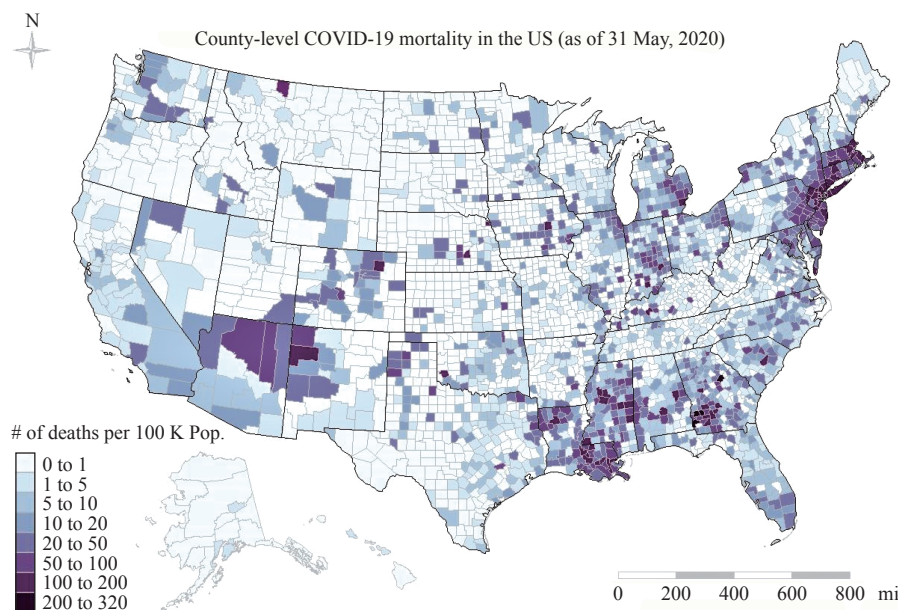


Fig. 2 The distribution of COVID-19 mortality across all 3141 counties in the United States (US). COVID-19 mortality was defined as the number of COVID-19 deaths per 100 000 population (Pop.) for the county.

Table 2 Results of HGLM base model^a for demographics

Demographic variable	Coeff.	Half-Std. Coeff.	Robust SE	t-Ratio	Approx. DF	P	ERR	Half-Std. ERR	% Change	Abs. % Change	Rank
(Intercept)	2.04		0.19	10.59	50	<0.001	7.68	–	–	–	–
Proportion of below 18 years of age	-0.38	-0.01	1.08	-0.35	3077	0.73	0.69	0.99	-1.3%	1.3%	7
Proportion of 65 and older	4.42	0.21	1.33	3.32	3077	<0.001	83.48	1.23	23.2%	23.2%	4
Proportion of non-Hispanic Black	2.10	0.30	0.44	4.73	3077	<0.001	8.14	1.35	35.0%	35.0%	1
Proportion of Hispanic	2.06	0.28	0.28	7.39	3077	<0.001	7.82	1.33	32.9%	32.9%	2
Proportion of female	2.33	0.05	3.22	0.72	3077	0.47	10.30	1.05	5.5%	5.5%	6
Proportion of rural	-1.22	-0.38	0.58	-2.09	3077	0.04	0.30	0.68	-31.9%	31.9%	3
Rural-urban continuum code	-0.08	-0.21	0.04	-1.87	3077	0.06	0.93	0.81	-18.9%	18.9%	5

^aCounty-level overdispersion parameter (σ^2)=18.90, and state-level residual variance (τ)=0.75 ($P<0.001$). SE: standard error; DF: degree of freedom; ERR: event rate ratio.

indicated that the proportion of non-Hispanic Black had the highest rank (#1) with a half-standardized ERR of 1.35, which means that for two counties that had one standard-deviation difference in the proportion of non-Hispanic Black ($SD=0.14$, **Table 1**), on average, the country that had a higher proportion of non-Hispanic Black had 35.0% more death than the country that had a lower proportion of non-Hispanic Black, controlling for other demographic variables in the model. Another example but a reverse case would be that if a county's proportion of rural went up one standard deviation ($SD=0.31$, **Table 1**), the COVID-19 mortality of that county would go down by 31.9% on average. It is also interesting to see that race/ethnicity measured as the proportions of non-Hispanic Black and Hispanic occupied the top two ranks, indicated that racial/ethnic minorities acted as the highest risk factors of COVID-19 mortality. On the other hand, rurality measured as the proportion of rural areas worked as a protective factor.

Identifying county-level health factors associated with COVID-19 mortality

To make a fair comparison for the mean event rate of COVID-19 mortality, we controlled demographic variables and general health by adding them to the HGLM model as covariates when identifying county-level health factors uniquely associated with COVID-19 mortality. To avoid potential model overfitting due to too many item-level health factors and general health (totally 77 items) in the County Health Rankings^[2], we took a two-step approach to investigating how each health factor was uniquely related to COVID-19 mortality. The first step was a preliminary round to select potentially highly influential factors. To do that, we added one item-level factor at a time to the HGLM Poisson regression in addition to the

demographic variables mentioned earlier. Using the criterion of half-standardized $ERR \geq 0.30$, which is considered to be a substantively moderate effect size^[26], we initially selected 12 item-level health factors and general health from the totally 77 items. Those selected item-level health factors and general health were listed in **Supplementary Table 1** (available online).

Furthermore, for highly correlated items, such as food environment index and limited access to healthy foods within the factor category of Diet & Exercise ($r=-0.77$), and proportions of uninsured, uninsured adults, and uninsured children in the category of *Access to Care* (all $r_s > 0.73$), we decided to create a composite score for the highly correlated items in the category by creating a z-score for each item, adding the z-scores up, and then re-standardizing them into a single overall z-score to represent each category in order to increase the interpretability of the results. For other items, we also created a z score for each item so that we could compare the relative influence of each factor on COVID-19 mortality.

At the second step of analysis, the outcome, COVID-19 mortality, was regressed on these nine z-scores in addition to the seven demographic variables described in the previous section using HGLM Poisson regression with overdispersion allowed. The results for this analysis were displayed in **Table 3**. From **Table 3** we can see that long commute driving alone, severe housing problems, and juvenile arrests rate were statistically significant health factors of COVID-19 mortality ($P < 0.05$); and the rates of suicide and uninsured were marginally statistically significant ($P < 0.10$). However, in terms of effect size measured by ERR, suicide rate was most influential (-28.1%), followed by uninsured rate (20.7%), long commute driving alone (19.9%), juvenile arrests rate (-15.0%),

Table 3 Results of HGLM Poisson regression with overdispersion^a for health factors and general health

Factor category	Item (in z-score)	Coeff.	Robust SE	t-Ratio	Approx. DF	P	ERR	Rank
Health factors								
Health behaviors								
Diet & exercise	Access to healthy food	0.11	0.15	0.69	1360	0.49	1.11	7
Clinical care								
Access to care	Uninsured rate	0.19	0.10	1.87	1360	0.06	1.21	2
Social & economic factors								
Family & social support	Social associations number rate	-0.02	0.08	-0.19	1360	0.85	0.98	9
Community safety	Suicides rate	-0.33	0.17	-1.90	1360	0.06	0.72	1
	Juvenile arrests rate	-0.16	0.07	-2.23	1360	0.03	0.85	4
Physical environment								
Housing & transit	Long commute driving alone	0.18	0.06	2.85	1360	<0.01	1.20	3
	Severe housing problems	0.13	0.05	2.70	1360	0.01	1.14	5
General health								
Length of life	Child mortality number rate	-0.10	0.11	-0.87	1360	0.39	0.91	8
	Average life expectancy	-0.14	0.10	-1.41	1360	0.16	0.87	6

^aThe seven demographic variables were controlled in the model but not shown in this table for clarity. The estimates of the intercept, the level-1 overdispersion parameter, and the level-2 residual variance are also not included in this table. SE: standard error; DF: degree of freedom; ERR: event rate ratio.

severe housing problem (13.6%), access to healthy food (11.2%), and social associations number rate (-1.5%). It should be noted that two general health items (child mortality number rate and average life expectancy) were neither statistically significant nor necessary to be interpreted as controlled covariates.

Discussion

Since the first confirmed case of COVID-19 in the US was reported on 20 January, 2020 in Snohomish County, Washington^[27], public health research on causes of COVID-19 and safety measures for preventing COVID-19 infections as well as biomedical and pharmacological research on COVID-19 vaccines and treatment medicines have exploded with about 2000 publications a week^[28-29]. The unit of analysis in most of public health research was states or the country as a whole, which failed to address the disparities in health resources and services among counties and the problems of the decentralized health administration in the US. This study investigated county-level health factors, frequently used in community public health research, that may impact or be associated with COVID-19 cases and deaths so that the study results can inform the local governments for making more targeted health policies to prevent and reduce future pandemics.

This study first confirmed the prior research

findings that most of the key county-level demographics are associated with COVID-19 mortality using the most recent available information about all counties in the U.S. Specifically, counties with a higher Black population, a higher Hispanic population, and a more older (≥ 65) population can have significantly higher COVID-19 mortality; whereas counties with more rural areas are significantly less likely to have higher COVID-19 mortality. Although these demographic characteristics are non-modifiable, local health officials and community health workers might need to have some targeted health prevention programs for those specific high-risk subpopulations and/or living in urbanized areas to reduce mortality from future outbreaks of novel viruses.

After controlling for these non-modifiable demographics as well general health, this study then further identified some county-level health factors that are significantly associated with COVID-19 mortality. For example, counties with a higher rate of uninsured population, more housing problems such as overcrowding, and longer commute driving alone are more likely to have significantly higher COVID-19 mortality; whereas counties with higher rates of suicides and juvenile arrests may have lower COVID-19 mortality. These health factors are modifiable; thus, more effective policies on health insurance and housing are needed to reduce future

mortality from novel viruses. Because of comprehensive nature of the data, these results can be treated as hard evidence to support the claims that have been frequently mentioned in the mass media.

It is interesting to find that the relationship between long commute driving alone and COVID-19 mortality is significantly positive. The finding seems counterintuitive but not surprising because the farther people commute by vehicle, the higher their blood pressure and body mass index and the less physical activity the individuals tend to participate in^[30]. It may result in higher likelihood of obesity^[31] and poorer mental health^[32-33]. Also, commutes by private motorized vehicle are associated with higher health risks such as cardiovascular diseases^[34]. All these health risks may eventually lead to higher COVID-19 mortality.

The findings that the rates of suicides and juvenile arrests have a significantly negative association with COVID-19 mortality also seem counterintuitive. One possible explanation is that those who died from suicide, especially for those who had high health risks, would contribute to a possible undercount of mortality from COVID-19. As for juvenile arrests, those juveniles would have less chance to contract COVID-19 due to the confinement. In addition, prior research found that the proportion of non-Hispanic White in the community is significantly associated with a lower rate of juvenile arrests^[35]. On the other hand, the present study found that, comparing with the reference group which has 90% of non-Hispanic White, counties with a higher proportion of non-Hispanic Black had higher COVID-19 mortality. Thus, it is possible that juvenile arrests appear to be negatively correlated with COVID-19 mortality after controlling for the racial/ethnic composition as we did in the analysis. Yet, these are only possible explanations. Because COVID-19 is novel, there is no established theoretical model or protocol to follow; thus, these counterintuitive findings are subject to future investigations.

Because there is no established theory to follow for studying this novel COVID-19, this study meant to be exploratory, and the findings are preliminary and open to discussion. Nevertheless, to our knowledge, this study was the first to examine such complex associations of a comprehensive set of health factors with COVID-19 outcomes for all the US counties, controlling for variations of all the US states, in a single modeling technique of HGLM Poisson regression with overdispersion. Such advanced technique allows disentangling county- and state-level effects on COVID-19 outcomes.

This study also has a few limitations subject to future research. For example, the COVID-19 outcomes might be underestimated due to undertesting and asymptomatic cases. Also, some time-varying events, such as availability of ventilators, state-to-state travel restrictions, social distancing practice, face mask requirement, stay-at-home order, economic reopen, school reopen, subsequent spikes, and so on, would be desirable factors along with state-level characteristics to be measured and integrated into analysis in future research.

The study data were collected up to 31 May, 2020, and the results reflected the nature of the country before reopening its economy gradually from the late May 2020. Unfortunately, after the country reopened its economy, the number of COVID-19 cases has skyrocketed. It is shocking to see that, as of the date of finishing writing this manuscript (31 July, 2020), the number of COVID-19 cases increased from 2.6 million, only a month ago, to more than 4.5 million with 25 000 more deaths nationally. Such alarmingly exponential increase makes public health research on COVID-19 prevention, along with biomedical and pharmacological research on COVID-19 vaccines and drugs, ever more important. It is hoped that the findings of this study can contribute to future more informed research and help local health administrators create more targeted health policies and programs so that people in the communities will be better off from the collective efforts from all sectors to flight COVID-19 pandemic as well as future potential virus outbreaks.

References

- [1] Johns Hopkins University & Medicine. COVID-19 case tracker 2020[EB/OL]. [2020-06-30]. <https://coronavirus.jhu.edu/accessed>.
- [2] University of Wisconsin Population Health Institute. County health rankings & roadmaps: building a culture of health, county by county 2020[EB/OL]. [2020-06-30]. <https://www.countyhealthrankings.org/accessed>.
- [3] Tian WJ, Jiang WL, Yao J, et al. Predictors of mortality in hospitalized COVID-19 patients: a systematic review and meta-analysis[J]. *J Med Virol*, 2020, 92(10): 1875–1883.
- [4] Harlem G. Descriptive analysis of social determinant factors in urban communities affected by COVID-19[J]. *J Public Health*, 2020, 42(3): 466–469.
- [5] Ramirez IJ, Lee J. COVID-19 emergence and social and health determinants in colorado: a rapid spatial analysis[J]. *Int J Environ Res Public Health*, 2020, 17(11): 3856.
- [6] Cyrus E, Clarke R, Hadley D, et al. The impact of COVID-19

- on African American communities in the United States [EB/OL]. [202-05-19]. <https://www.medrxiv.org/content/10.1101/2020.05.15.20096552v1>.
- [7] Yancy CW. COVID-19 and African Americans[J]. *JAMA*, 2020, 323(19): 1891–1892.
- [8] Nayak A, Islam SJ, Mehta A, et al. Impact of social vulnerability on COVID-19 incidence and outcomes in the United States[EB/OL]. [2020-04-17]. <https://www.medrxiv.org/content/10.1101/2020.04.10.20060962v2>.
- [9] Zhang CH, Schwartz GG. Spatial disparities in coronavirus incidence and mortality in the united states: an ecological analysis as of May 2020[J]. *J Rural Health*, 2020, 36(3): 433–445.
- [10] Chin T, Kahn R, Li RR, et al. U.S. county-level characteristics to inform equitable COVID-19 response[EB/OL]. [2020-04-11]. <https://www.medrxiv.org/content/10.1101/2020.04.08.20058248v1>.
- [11] Oronce CIA, Scannell CA, Kawachi I, et al. Association between state-level income inequality and COVID-19 cases and mortality in the USA[J]. *J Gen Intern Med*, 2020, 35(9): 2791–2793.
- [12] Goldstein JR, Atherwood S. Improved measurement of racial/ethnic disparities in COVID-19 mortality in the United States[EB/OL]. [2020-06-23]. <https://www.medrxiv.org/content/10.1101/2020.05.21.20109116v2>.
- [13] El Bcheraoui C, Mokdad AH, Dwyer-Lindgren L, et al. Trends and patterns of differences in infectious disease mortality among US counties, 1980-2014[J]. *JAMA*, 2018, 319(12): 1248–1260.
- [14] Lo NC, Hotez PJ. Public health and economic consequences of vaccine hesitancy for measles in the United States[J]. *JAMA Pediatr*, 2017, 171(9): 887–892.
- [15] Scheinker D, Valencia A, Rodriguez F. Identification of factors associated with variation in US county-level obesity prevalence rates using epidemiologic vs machine learning models[J]. *JAMA Netw Open*, 2019, 2(4): e192884.
- [16] Langabeer JR, Chambers KA, Cardenas-Turanzas M, et al. County-level factors underlying opioid mortality in the United States[J]. *Subst Abuse*, 2020, doi: 10.1080/08897077.2020.1740379. [Epub ahead of print].
- [17] Maraccini AM, Yang W, Slonim AD. "Top Performing" US hospitals and the health status of counties they serve[J]. *J Community Health*, 2018, 43(3): 477–487.
- [18] Richardson S, Hirsch JS, Narasimhan M, et al. Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City Area[J]. *JAMA*, 2020, 323(20): 2052–2059.
- [19] Corey L, Mascola JR, Fauci AS, et al. A strategic approach to COVID-19 vaccine R&D[J]. *Science*, 2020, 368(6494): 948–950.
- [20] US Department of Agriculture Economic Research Service. Rural-urban continuum codes Washington, DC: US Department of Agriculture 2019[EB/OL]. [2020-06-20]. <https://www.ers.usda.gov/data-products/rural-urban-continuum-codes/accessed>.
- [21] New York Times. Coronavirus (Covid-19) Data in the United States. New York Times[EB/OL]. [2020-05-31]. <https://aws.amazon.com/marketplace/pp/Coronavirus-COVID-19-Data-in-the-United-States-The/prodview-jmb464qw2yg74#overview>.
- [22] Centers for Disease Control and Prevention. Coronavirus Disease 2019 (COVID-19). U.S. Department of Health & Human Services[EB/OL]. [2020-05-31]. <https://search.cdc.gov/search/?query=Coronavirus+Disease+2019+%28COVID-19%29%3A+U.S.+Department+of+Health+%26+Human+Services&utf8=%E2%9C%93&affiliate=cdc-main>.
- [23] Raudenbush SW, Bryk AS. Hierarchical linear models: applications and data analysis methods[M]. 2nd ed. Thousand Oaks: Sage Publications, 2002.
- [24] Raudenbush SW, Bryk AS, Cheong YF, et al. HLM 7: hierarchical linear and nonlinear modeling[R]. Chicago: Scientific Software International, 2011.
- [25] Gregory C, Miyazaki Y. Multilevel analysis of student civics knowledge scores[J]. *J Educ Res*, 2018, 111(3): 295–309.
- [26] Azuero A. A note on the magnitude of hazard ratios[J]. *Cancer*, 2016, 122(8): 1298–1299.
- [27] Holshue ML, DeBolt C, Lindquist S, et al. First case of 2019 novel coronavirus in the united states[J]. *N Engl J Med*, 2020, 382(10): 929–936.
- [28] Lu Z, Chen Q, Allot A, et al. LitCovid Bethesda, MD: U.S. National Library of Medicine National Center for Biotechnology Information; 2020[EB/OL]. [2020-06-20]. <https://www.ncbi.nlm.nih.gov/research/coronavirus>.
- [29] Chen QY, Allot A, Lu ZY. Keep up with the latest coronavirus research[J]. *Nature*, 2020, 579(7798): 193.
- [30] Hoehner CM, Barlow CE, Allen P, et al. Commuting distance, cardiorespiratory fitness, and metabolic risk[J]. *Am J Prev Med*, 2012, 42(6): 571–578.
- [31] Frank LD, Andresen M, Schmid TL. Obesity relationships with community design, physical activity, and time spent in cars[J]. *Am J Prev Med*, 2004, 27(2): 87–96.
- [32] Künn-Nelen A. Does commuting affect health?[J]. *Health Econ*, 2016, 25(8): 984–1004.
- [33] Hilbrecht M, Smale B, Mock SE. Highway to health? Commute time and well-being among Canadian adults[J]. *World Leis J*, 2014, 56(2): 151–163.
- [34] Patterson R, Panter J, Vamos EP, et al. Associations between commute mode and cardiovascular disease, cancer, and all-cause mortality, and cancer incidence, using linked Census data over 25 years in England and Wales: a cohort study[J]. *Lancet Planet Health*, 2020, 4(5): e186–e194.
- [35] Gase LN, Glenn BA, Gomez LM, et al. Understanding racial and ethnic disparities in arrest: the role of individual, home, school, and community characteristics[J]. *Race Soc Probl*, 2016, 8(4): 296–312.