

RESEARCH

Open Access



# Association between county-level risk groups and COVID-19 outcomes in the United States: a socioecological study

Sadiya S. Khan<sup>1,2\*</sup>, Amy E. Krefman<sup>2</sup>, Megan E. McCabe<sup>2</sup>, Lucia C. Petito<sup>2</sup>, Xiaoyun Yang<sup>2</sup>, Kiarri N. Kershaw<sup>2</sup>, Lindsay R. Pool<sup>2</sup> and Norrina B. Allen<sup>2</sup>

## Abstract

**Background:** Geographic heterogeneity in COVID-19 outcomes in the United States is well-documented and has been linked with factors at the county level, including sociodemographic and health factors. Whether an integrated measure of place-based risk can classify counties at high risk for COVID-19 outcomes is not known.

**Methods:** We conducted an ecological nationwide analysis of 2,701 US counties from 1/21/20 to 2/17/21. County-level characteristics across multiple domains, including demographic, socioeconomic, healthcare access, physical environment, and health factor prevalence were harmonized and linked from a variety of sources. We performed latent class analysis to identify distinct groups of counties based on multiple sociodemographic, health, and environmental domains and examined the association with COVID-19 cases and deaths per 100,000 population.

**Results:** Analysis of 25.9 million COVID-19 cases and 481,238 COVID-19 deaths revealed large between-county differences with widespread geographic dispersion, with the gap in cumulative cases and death rates between counties in the 90<sup>th</sup> and 10<sup>th</sup> percentile of 6,581 and 291 per 100,000, respectively. Counties from rural areas tended to cluster together compared with urban areas and were further stratified by social determinants of health factors that reflected high and low social vulnerability. Highest rates of cumulative COVID-19 cases (9,557 [2,520]) and deaths (210 [97]) per 100,000 occurred in the cluster comprised of rural disadvantaged counties.

**Conclusions:** County-level COVID-19 cases and deaths had substantial disparities with heterogeneous geographic spread across the US. The approach to county-level risk characterization used in this study has the potential to provide novel insights into communicable disease patterns and disparities at the local level.

**Keywords:** COVID-19, County-level health, Communicable diseases, Pandemic

## Introduction

Since the first case of the coronavirus disease 2019 (COVID-19) was confirmed in the United States (US) in Snohomish county, Washington, on January 21<sup>st</sup>, 2020 [1, 2], the US has experienced more cases (99,467 cases per

million population) and deaths (585,708) than any other country in the world (as of May 15, 2021) [3]. Clinical and epidemiologic data have identified several modifiable health factors (e.g., obesity [4], hypertension [5], cardiovascular disease [6]) as key risk factors of adverse outcomes such as hospitalization and mortality related to severe COVID-19 infection [7, 8]. However, assessment of individual-level health factors incompletely characterizes risk related to underlying social determinants of

\*Correspondence: s-khan-1@northwestern.edu

<sup>1</sup> Division of Cardiology, Department of Medicine, Northwestern University Feinberg School of Medicine, 680 N. Lake Shore Drive Ste. 1400, Chicago, IL 60611, USA

Full list of author information is available at the end of the article



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

health that may contribute to observed differences in COVID-19 outcomes [9, 10].

Understanding how place-based disadvantage has influenced COVID-19 case and death rates is critical for developing strategies to address long-term sequelae of COVID-19 [11] and to target vaccination efforts [12] at the local level. This is especially important given the significant geographic heterogeneity observed in the burden of COVID-19 with large differences in morbidity and mortality among counties within the US [13, 14]. Several studies have focused on single social and environmental factors (e.g., housing overcrowding [15], air pollution [16]) or grouped factors based on existing county-level rankings (e.g., social vulnerability index [17]) to examine differences at the county-level in COVID-19 outcomes. However, susceptibility to COVID-19 infection and mortality is likely due to interrelated biological, social, and environmental factors; to date, no studies have developed a comprehensive measure that integrates place-based factors across multiple domains to classify county-level risk for COVID-19 outcomes.

This study sought to use a novel method for clustering, latent class analysis, to identify meaningful groups of counties based on county-level characteristics across multiple sociodemographic, health, and environmental domains and to examine associations with COVID-19 cases and deaths.

## Methods

### Data Sources

We harmonized and linked a comprehensive set of county-level measures from a variety of resources including the U.S. Census Bureau, County Health Rankings (CHR), Centers for Disease Control and Prevention (CDC) Wide-ranging Online Data for Epidemiologic Research (WONDER) mortality database, Area Health Resources Files (AHRF), and the United States Department of Agriculture (USDA) across all available counties coded by Federal Information Processing Standard county codes for a total of 30 variables (Supplemental Table 1). First, from these various data sources, we obtained relevant measures to characterize counties' profiles according to the following categories based on pre-pandemic data on demographics, socioeconomic status, health status, healthcare access, and environmental factors. Next, we queried county-level COVID-19 data on 2/18/21 to include all deaths and cases from 1/21/20 to 2/17/21 from The New York Times GitHub where information on cases and fatalities is updated daily (<https://github.com/nytimes/COVID-19-data>) to link with our integrative county-level index. For a subset of counties, vaccination data (number of first dose and two doses administered) were manually abstracted from county and

state department of public health websites by SSK, AEK, LRP, and NBA. All data were deidentified and publicly available, and therefore deemed exempt from Northwestern University's Institutional Review Board.

### Measures

Counties were described using metrics relating to residents' demographics, socioeconomic status, health status, healthcare access, as well as environmental factors (Supplemental Table 1). Demographic factors included proportion of racial/ethnic minority and older residents in addition to population density and urbanization status. Socioeconomic metrics included rates of varying educational levels, poverty, and housing. Health status included quality of life and morbidity and mortality rates. Healthcare access reflected insurance, hospital beds per capita, and primary care physician density. Finally, the environmental factor was a measure of air pollution. The primary outcomes were cumulative deaths and cases per 100,000 population attributed to COVID-19 at the county-level. As an exploratory outcome, COVID-19 vaccine doses administered in a subset of counties was examined.

### Statistical Analysis

We included all counties for which we had complete ecological data across county-level metrics that span demographic, socioeconomic status, health status, healthcare access, and environmental factors and at least 5 COVID-19 deaths. We used latent class analysis (LCA) [18] to classify US counties with others who have similar socioecological profiles based on these included metrics. These groupings represent meaningful clusters defined by a combination of all the observed variables by using model-based predicted posterior probabilities. The LCA model incorporated binary and ordinal variables as-is, and continuous variables were categorized as tertiles before inclusion. The model was fit using R package *poLCA*, which estimates the latent class model by maximizing a log-likelihood function using the EM algorithm. We set the max iterations to 10000. To select the optimal number of classes (between 1-5) for the final LCA model, we imposed the following criteria based on previously published methods: (1) qualitative evaluation of distinct predicted classes or groups, (2) ensuring no group had fewer than 5% of counties, and (3) minimizing the Bayesian information criterion. Model fit statistics for models with 1-5 latent classes (Log-likelihood, Residual DF, BIC, aBIC, cAIC, and Likelihood-Ratio) in Supplemental Table 2 [19]. Conditional probabilities and a classification table for each of the county-level variables for the final model are shown in Supplemental Tables 3 and 4.

For the final identified groups, we described county-level characteristics using means, medians, and

proportions, as appropriate, and described crude rates of cumulative COVID-19 cases and deaths per 100,000 population. We additionally identified the ten counties in each group with the highest deaths per 100,000 population to describe dynamic trends in cumulative deaths since the beginning of the pandemic and cumulative cases, deaths, and vaccination rates per 100,000.

In the primary analysis, we calculated the association between each county-level risk group and each outcome (COVID-19 deaths and cases) using linear regression models. We fit two sets of models: 1) unadjusted and 2) adjusted for the time since the first confirmed case in the county, to allow comparison between counties at similar points in their outbreak. Point estimates were interpreted in context of their magnitude and Wald-style 95% confidence interval. We completed a sensitivity analysis including all counties reporting at least one COVID-19 death in order to examine the robustness of our estimates ( $N=2694$  counties). All analyses were completed using R version 3.6.3 or newer.

## Results

### County-Level Risk Groups

Data on the 30 county-level measures were available for 2,701 US counties. We identified four distinct groups of counties based on demographic, socioeconomic, health status, healthcare access, and environmental profiles (Table 1, Supplemental Figure 1): 1) diverse urban counties with greater social and health assets; 2) rural counties with social and health vulnerabilities; 3) older rural counties with greater social assets; and 4) urban counties with social vulnerabilities. Diverse urban counties had the highest population density (757 [2244]), greatest proportion of at least some college education (71% [7]), lowest rates of unemployment (3% [1]) and percent of children in poverty (11% [4]), and fewest absolute years of potential life lost (5756 [1127] years). By contrast, vulnerable rural counties had the highest proportion of non-Hispanic black residents (17% [19]), lowest proportion of some college education (48% [8]), highest rates of unemployment (5% [1]) and percent of children in poverty (30% [7]), and greatest absolute years of potential life lost (10,768 [2062] years).

Individual county-level metrics included in the latent class modeling were highly correlated with each other (Supplemental Figure 2). For example, favorable social and health assets tracked in counties with lower poverty metrics and higher median income; these counties also had more favorable health metrics such as proportions of smokers, people with obesity, and physically inactive people, as well as fewer years of potential life lost. Conversely, adverse social and health factors and a higher proportion of minority residents were observed in

vulnerable counties. These counties also had, on average, lower median income, greater proportion of households with overcrowding, and poorer air pollution rates. Counties within the four groups have widespread geographic dispersion with some regional clustering (Figure 1A). For example, rural counties with social and health vulnerabilities are predominantly, but not exclusively, localized in the southern US.

### Nationwide COVID-19 Deaths and Cases

As of February 17<sup>th</sup>, 2021, 2,621 counties had reported at least 1 confirmed COVID-19 death and 2,701 counties had at least 1 confirmed COVID-19 case to be included in the publicly available dataset hosted by the NY Times. A total of 481,238 COVID-19 deaths and 25.9 million COVID-19 cases were included from 2/29/20 to 2/17/21 in 2701 counties in the US. Large between-county differences were evident with widespread geographic dispersion, with a significant gap in cumulative COVID-19 cases and death rates between counties in the 90<sup>th</sup> and 10<sup>th</sup> percentile of 6,581 and 291 per 100,000, respectively. Across US census regions, counties in the Midwest had the greatest cumulative rate of COVID-19 cases (8825) and counties in the Northeast had the greatest cumulative rate of COVID-19 deaths (190) (Figures 1B and C). Yet, there was wide geographic dispersion with counties with high burden of COVID-19 distributed across the US. Qualitatively, counties with high COVID-19 cases also had high COVID-19 deaths.

### Association Between County-Level Risk Groups and COVID-19 Outcomes

A significant association was observed between county-level risk groups and COVID-19 deaths and cases per 100,000, even after adjusting for time since first death and case, respectively, within each county (Figure 2, Table 2, Supplemental Table 5). Highest rates of COVID-19 cases (9557 [2520]) and deaths (210 [97]) per 100,000 population occurred in the cluster comprised of rural disadvantaged counties ( $p<0.05$ ). Rates of COVID-19 deaths were nearly 2-fold higher in the rural disadvantaged cluster compared with the diverse urban counties with greater social assets.

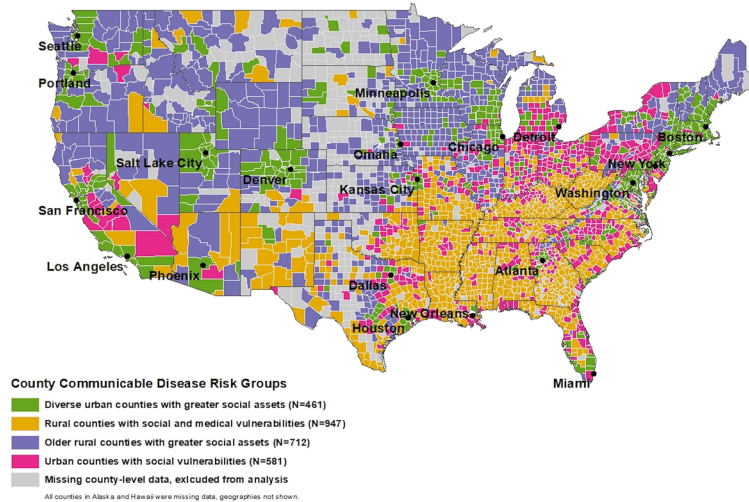
Within each county-level risk group, we next identified the ten counties with the highest cumulative COVID-19 deaths per 100,000 population and visualized the trajectory of COVID-19 mortality as a function of time since the 5<sup>th</sup> reported COVID-19 death in each county (Figure 3). Hancock county, Georgia, in the vulnerable rural county-level risk group experienced the highest burden of COVID-19 deaths (622 per 100,000) while Norton County, Kansas, in the older rural county-level risk group had the highest level of cases (22,357 per 100,000) across

**Table 1** County-level demographic, socioeconomic, health status, healthcare access, and other factors in each county-level risk group in the United States (N=2701 counties)

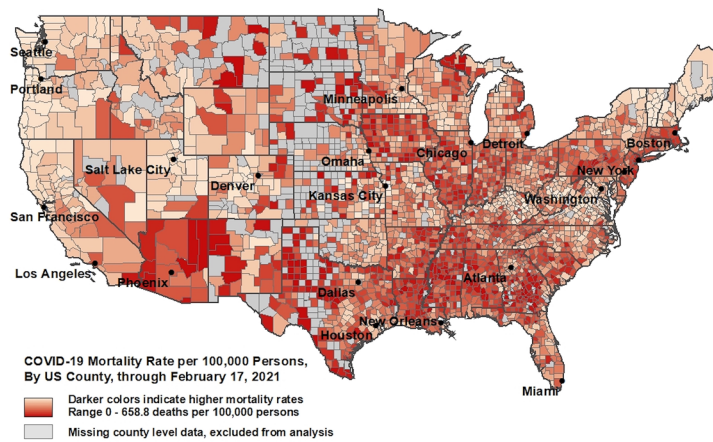
	<b>Diverse Urban counties with greater social assets</b> <i>n</i> =461	<b>Rural Counties with social and medical vulnerabilities</b> <i>n</i> =947	<b>Older Rural counties with greater social assets</b> <i>n</i> =712	<b>Urban Counties with social vulnerabilities</b> <i>n</i> =581
<b>Demographics</b>				
Population density	756.84 (2243.99)	92.35 (447.56)	29.20 (25.03)	378.82 (897.47)
Proportion male	0.50 (0.01)	0.50 (0.03)	0.51 (0.02)	0.50 (0.02)
Proportion non-Hispanic White	0.76 (0.16)	0.69 (0.23)	0.85 (0.14)	0.76 (0.17)
Proportion non-Hispanic Black	0.06 (0.07)	0.17 (0.19)	0.02 (0.03)	0.11 (0.11)
Proportion Hispanic	0.11 (0.10)	0.09 (0.16)	0.09 (0.13)	0.09 (0.12)
Proportion 60+ years old	0.23 (0.05)	0.26 (0.04)	0.29 (0.06)	0.24 (0.04)
Proportion 80+ years old	0.04 (0.01)	0.04 (0.01)	0.05 (0.01)	0.04 (0.01)
Rural-Urban Continuum Code (%)				
1	212 (46.0)	36 (3.8)	8 (1.1)	155 (26.7)
2	100 (21.7)	86 (9.1)	28 (3.9)	152 (26.2)
3	74 (16.1)	82 (8.7)	52 (7.3)	122 (21.0)
4	31 (6.7)	72 (7.6)	33 (4.6)	77 (13.3)
5	14 (3.0)	35 (3.7)	28 (3.9)	12 (2.1)
6	10 (2.2)	301 (31.8)	210 (29.5)	57 (9.8)
7	15 (3.3)	158 (16.7)	225 (31.6)	5 (0.9)
8	2 (0.4)	83 (8.8)	56 (7.9)	1 (0.2)
9	3 (0.7)	94 (9.9)	72 (10.1)	0 (0.0)
<b>Socioeconomic Status</b>				
High school graduation rate	88.70 (6.25)	87.97 (7.51)	89.14 (6.99)	88.70 (5.34)
Percent some college education	71.09 (7.23)	48.48 (8.24)	59.89 (9.31)	59.30 (7.61)
Percent unemployed	3.31 (0.74)	4.96 (1.44)	3.80 (1.44)	4.04 (0.91)
Percent children in poverty	11.27 (4.08)	29.85 (7.02)	17.47 (4.90)	19.16 (4.30)
Income ratio (80th to 20th percentile)	4.28 (0.66)	4.99 (0.78)	4.15 (0.43)	4.45 (0.59)
Percent households with overcrowding	2.22 (1.58)	2.75 (1.96)	2.02 (1.54)	2.22 (1.48)
Median income	71557.97 (14989.20)	40188.09 (6170.90)	52465.77 (6985.36)	53390.72 (6798.91)
<b>Health Status</b>				
Years of potential life lost rate	5755.71 (1127.19)	10767.91 (2061.60)	7410.08 (1550.12)	8334.47 (1203.43)
Percent reporting fair/poor health	13.56 (2.12)	22.78 (3.62)	15.11 (2.44)	17.86 (2.08)
Average physically unhealthy days/ month	3.36 (0.37)	4.70 (0.52)	3.61 (0.43)	4.03 (0.33)
Percent smoking	14.04 (2.12)	20.67 (2.99)	15.70 (1.81)	17.93 (2.01)
Percent obesity	27.92 (4.90)	35.94 (5.03)	32.06 (4.37)	33.85 (3.93)
Percent physically inactive	20.61 (3.88)	31.86 (4.86)	25.80 (4.02)	27.72 (3.68)
Cardiovascular disease mortality rate	339.42 (54.33)	516.66 (91.38)	397.78 (66.65)	432.29 (57.85)
Cancer mortality rate	157.83 (20.17)	202.21 (31.76)	177.72 (26.85)	185.79 (18.50)
<b>Healthcare Access</b>				
Percent uninsured	8.15 (3.63)	13.20 (4.77)	10.36 (5.11)	10.72 (4.33)
Primary care physician rate	79.86 (41.33)	41.38 (23.25)	56.43 (28.58)	54.56 (28.65)
Preventable hospitalization rate	3912.93 (1238.16)	5983.81 (1953.21)	4081.10 (1573.13)	4991.30 (1028.37)
Percent vaccinated	50.18 (5.64)	38.93 (7.77)	40.28 (10.21)	46.88 (5.49)
Hospital beds per 100,000 persons	229.62 (290.35)	281.63 (385.68)	306.94 (455.30)	264.30 (275.85)
<b>Environmental and other</b>				
Social association rate	9.81 (3.78)	10.73 (4.17)	14.07 (5.72)	10.96 (3.55)
Average daily PM2.5	8.98 (1.95)	9.73 (1.33)	8.06 (1.81)	10.42 (1.35)

All values are mean (standard deviation) except for rural-urban continuum codes that is *N* (%). PM represents particulate matter

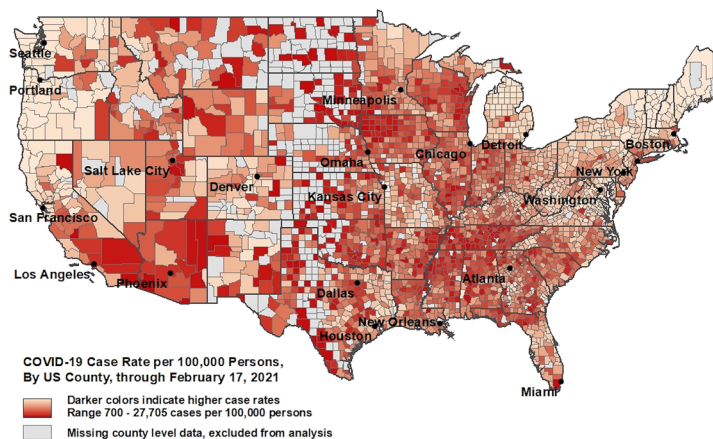
### A. County-Level Risk Groups



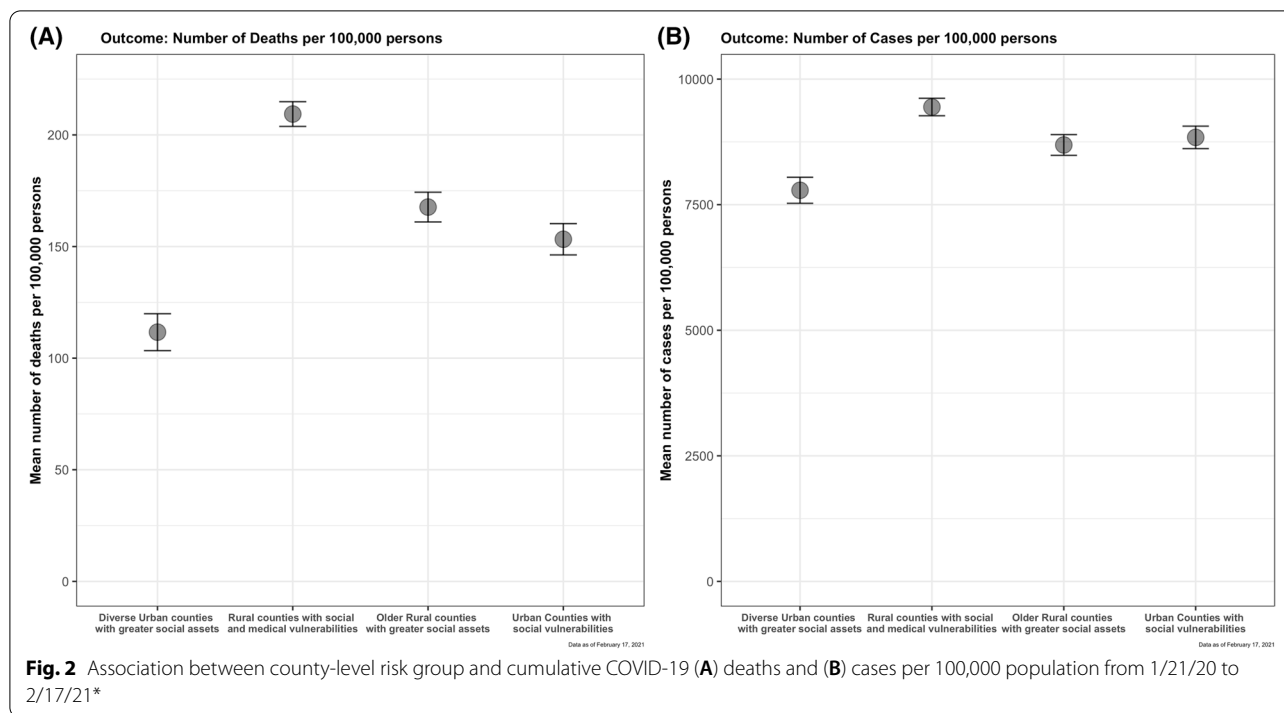
### B. COVID-19 Deaths (per 100,000 population)



### C. COVID-19 Cases (per 100,000 population)



**Fig. 1** Nationwide Distribution of (A) County-Level Risk Groups (B) COVID-19 deaths per 100,000 and (C) COVID-19 cases per 100,000 from 1/21/20 to 2/17/21



**Table 2** Summary of Covid-19 cases and deaths stratified by county-level risk groups from January 21<sup>st</sup> 2020 to February 17<sup>th</sup>, 2021

	Diverse Urban counties with greater social assets	Rural Counties with social and medical vulnerabilities	Older Rural counties with greater social assets	Urban Counties with social vulnerabilities
Counties, N	448	916	677	580
Covid-19 related cases, N	13,906,385	3,383,189	1,4092,18	8,606,695
Covid-19 related cases per 100,000 persons, Mean (SD)	7803 (2614.3)	9556.6 (2519.6)	8771.7 (3121.9)	8825.8 (2275.6)
Covid-19 related deaths, N	238,268	72,412	25,639	144,919
Covid-19 related deaths per 100,000 persons, Mean (SD)	109 (59)	210 (97.2)	169.6 (98.1)	152 (61.3)

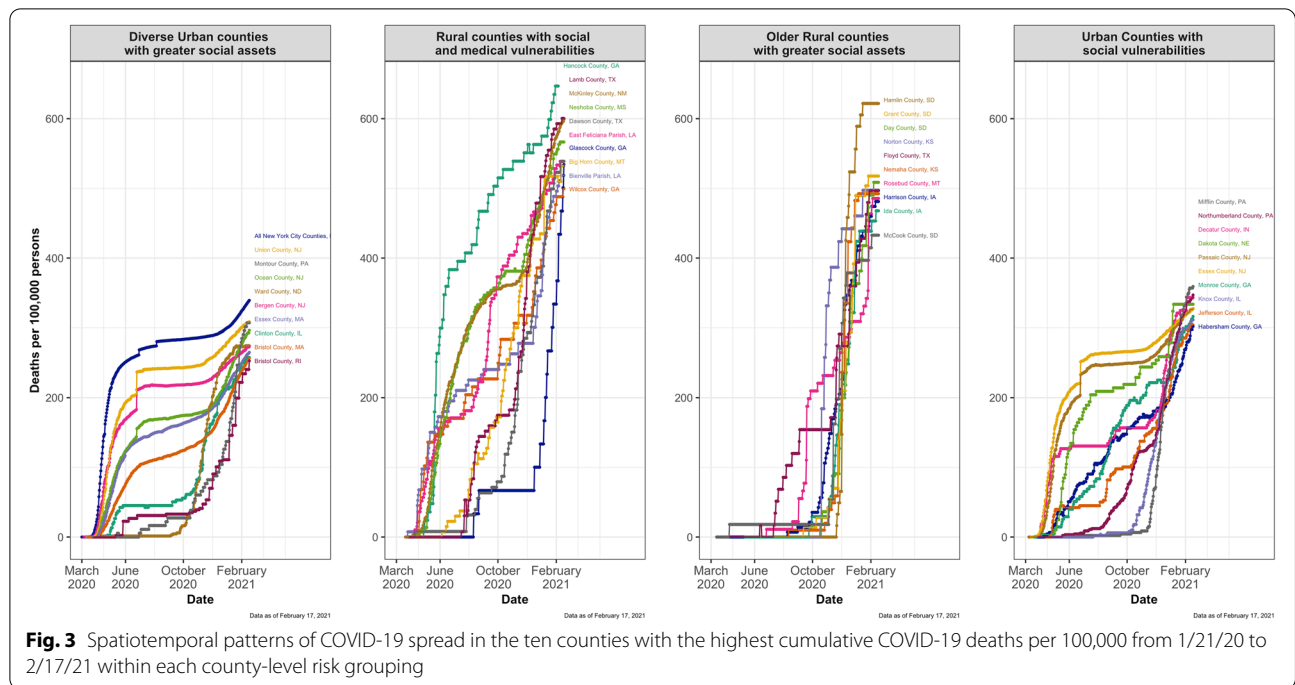
the US (Supplemental Table 6). Yet, both county-level risk groups that predominantly included rural counties had the lowest rate of vaccination per 100,00 population compared with the two county-level risk groups predominantly comprised of urban counties (most data available as of 2/23/21). Within and across groups, there was no correlation between COVID-19 case and death rate and vaccination rate (Supplemental Figure 3).

**Discussion**

This study describes a novel, data-driven, classification scheme that identified four distinct county-level risk groups, which were significantly associated with COVID-19 deaths and cases per 100,000 population from January 21, 2020 to February 17, 2021. The county-level risk group clustered by rural setting and social vulnerabilities had the greatest burden of COVID-19 deaths by nearly

2-fold, but the lowest documented rate of vaccination, to-date. This approach identified county-level vulnerability to COVID-19 deaths and cases by integrating characteristics across multiple key domains (sociodemographic, health-related, and environmental) implicated in the risk for COVID-19 infection and severity.

On a population level, targeting prevention of COVID-19 through vaccination, especially with resurgence in areas with novel variants, as well as directing resources to address long-term sequelae or disability due to COVID-19 infection requires identification of communities who have experienced greatest burden of COVID-19 infection and mortality [20, 21]. While there may be multiple mechanisms by which county-level factors may influence COVID-19 outcomes, classifying counties with similar socioecological profiles, which are implicated in COVID-19 infection rate and illness severity, will provide the



ability for public health responses targeted to the communities with the greatest need. This is especially important at this stage in the pandemic when novel variants are emerging that may be more infectious or be associated with greater infection severity [22]. Other tools, such as the social vulnerability index (SVI) have been applied and demonstrated significant association between greater social vulnerability and both COVID-19 and non-COVID-19 health outcomes [23]. However, the SVI [24] relies only on demographic variables from the US census and does not integrate health status and healthcare access for county’s residents or environmental measures at the county-level, such as air pollution. Therefore, this current study extends prior reports by developing a comprehensive and data-driven classification scheme to reflect social, health, and environmental vulnerability at the county-level.

These data build upon prior studies that have documented substantial county-level disparities in COVID-19 as well as a large body of literature demonstrating similar patterns in geographic disparities across the US in non-communicable diseases, such as cardiovascular disease [25, 26]. Indeed, COVID-19 infection susceptibility and mortality has further amplified pre-existing place-based health disadvantage due to interrelated sociodemographic [27], health [28], and environmental factors [29] at the county-level and expand growing awareness of a “rural mortality penalty” [30] in the US. Recent investigations have demonstrated widening

disparities in rural mortality rates due to noncommunicable diseases, such as chronic lung disease [31] and cardiovascular disease [32], compared with urban mortality rates. This likely reflects convergence of multiple ecological factors related to social determinants of overall health, such as economic prosperity [33] and education [27, 34]. Vulnerability as a result of underlying comorbidities may be one important factor in the observed place-based disparities, but does not account for intrinsic systemic and structural racism within and outside our healthcare system that may also be contributing [35]. Our study incorporates both sociodemographic factors and burden of chronic diseases at the county-level and provides a means to estimate susceptibility and burden of COVID-19 at the population level in hopes of mitigating avoidable health inequalities, especially in counties with limited resources.

Strengths of our study include use of innovative statistical methods to integrate 30 county-level characteristics that may be implicated in COVID-19 transmission and mortality with a latent-class modeling approach [18] that identified groups of counties with similar profiles of ecological risk factors. To accomplish this task, we harmonized and linked numerous publicly available county-level datasets to create a comprehensive classification scheme leveraged available data on county-level COVID-19 outcomes that represent cumulative cases and deaths from >12 months since the beginning of the pandemic.

This analysis has several limitations. First, there is the potential misclassification of COVID-19 cases and mortality due to under-ascertainment of COVID-19. Under-recognition of COVID-19 may have occurred due to limited availability of testing, especially earlier in the pandemic. Prior studies have demonstrated a strong correlation between testing and cases per 100,000 population and seroprevalence studies suggest that up to 10 times more COVID-19 infections occurred than number of cases [36, 37], which may reflect mild illness or no symptoms in addition to under-testing. Therefore, COVID-19 cases and deaths are unlikely to reflect the total burden of COVID-19. This has been demonstrated in analyses of excess deaths at the national level [38]. However, our analysis relies on the most comprehensive estimates of cumulative COVID-19 mortality and cases available at the county-level. Second, use of LCA assigns group membership for each county based on the highest probability, but there may be counties that may be equally likely to belong to another county-level risk group. However, examination of posterior probabilities revealed that no county had a probability <0.7 based on their assigned group membership and LCA facilitated the integration of 30 highly interrelated variables in one classification scheme in an unbiased data-driven approach. Third, our analysis does not account for individual-level characteristics that are associated with COVID-19 infection or disease severity. However, the current study design is ecological, and it is not meant to infer causation; rather it is meant to inform large-scale comparisons at the population-level to communicate with local health departments and leaders to inform decisions at the county-level [39]. Fourth, analysis at a smaller geographic scale, such as the neighborhood (census-tract) level was not possible in this nationwide study. Prior work in New York City [40], Milwaukee [41], and Chicago [42] have identified the neighborhood-level variation within a city in COVID-19 deaths that highlight structural inequities as root causes of these disparities and warrant further investigation. This analysis at the county-level is a meaningful geographic scale for changes in policy and public health response at the state- and national-level that can complement efforts within counties.

In conclusion, this study identified clustering of counties based on a variety of county-level factors that incorporate sociodemographic, health status and access, and environmental characteristics. The county-level risk index was significantly associated with COVID-19 cases and deaths, with the greatest burden for both occurring in the group with rural disadvantaged counties. Place-based health disadvantage for COVID-19 outcomes will require targeted and equitable allocation of resources for vaccination as well as management of long-term

sequelae and disability due to COVID-19. The findings from this current study provide unique insights into county-level characteristics that may inform the lack of reserve or resilience at a population-level when challenged by an unexpected stress of a highly infectious communicable disease that carries a significant risk of morbidity and mortality, such as COVID-19.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12889-021-12469-y>.

### Additional file 1.

### Acknowledgements

Not applicable

### Authors' contributions

SK, LP, KK, LP, and NA designed the research study. AK, MM, LP, XY performed data analysis and interpretation. SK drafted the manuscript. AK, MM, LP, XY, KK, LP and NA provided critical revisions to the manuscript. All authors read and approved the final manuscript.

### Funding

Research reported in this publication was supported, in part, by the National Institutes of Health's National Institute on Aging, Grant Number P30AG059988, to SSK. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The funding sponsor did not contribute to design and conduct of the study, collection, management, analysis, or interpretation of the data or preparation, review, or approval of the manuscript.

### Availability of data and materials

The datasets used and/or analyzed during the current study are available in the New York Times GitHub repository (<https://github.com/nytimes/COVID-19-data>). Additional datasets generated for analysis are included in this published article and its supplementary information files.

### Declarations

#### Ethics approval and consent to participate

All data were deidentified and publicly available, and therefore deemed exempt from Northwestern University's Institutional Review Board.

#### Consent for publication

Not applicable

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>Division of Cardiology, Department of Medicine, Northwestern University Feinberg School of Medicine, 680 N. Lake Shore Drive Ste. 1400, Chicago, IL 60611, USA. <sup>2</sup>Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA.

Received: 19 May 2021 Accepted: 20 December 2021

Published online: 13 January 2022

### References

- McMichael TM, Currie DW, Clark S, et al. Epidemiology of Covid-19 in a Long-Term Care Facility in King County, Washington. *N Engl J Med*. 2020;382(21):2005–11.



2. Yang BY, Barnard LM, Emert JM, et al. Clinical Characteristics of Patients With Coronavirus Disease 2019 (COVID-19) Receiving Emergency Medical Services in King County, Washington. *JAMA Netw Open*. 2020;3(7):e2014549.
3. Our World in Data. 2020; <https://ourworldindata.org>. Accessed 2/28/21, 2021.
4. Tartof SY, Qian L, Hong V, et al. Obesity and Mortality Among Patients Diagnosed With COVID-19: Results From an Integrated Health Care Organization. *Ann Intern Med*. 2020;173(10):773–81.
5. 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. *JAMA*. 2020;323(20):2052–9.
6. Guo T, Fan Y, Chen M, et al. Cardiovascular Implications of Fatal Outcomes of Patients With Coronavirus Disease 2019 (COVID-19). *JAMA Cardiol*. 2020;5(7):811–8.
7. Fisman DN, Bogoch I, Lapointe-Shaw L, McCready J, Tuite AR. Risk Factors Associated With Mortality Among Residents With Coronavirus Disease 2019 (COVID-19) in Long-term Care Facilities in Ontario, Canada. *JAMA Netw Open*. 2020;3(7):e2015957.
8. Rosenthal N, Cao Z, Gundrum J, Sianis J, Safa S. Risk Factors Associated With In-Hospital Mortality in a US National Sample of Patients With COVID-19. *JAMA Netw Open*. 2020;3(12):e2029058.
9. 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(7):e2016938.
10. Khazanchi R, Evans CT, Marcelin JR. Racism, Not Race, Drives Inequity Across the COVID-19 Continuum. *JAMA Netw Open*. 2020;3(9):e2019933.
11. Carfi A, Bernabei R, Landi F. Gemelli Against C-P-ACSG. Persistent Symptoms in Patients After Acute COVID-19. *JAMA*. 2020;324(6):603–5.
12. Moore JP. Approaches for Optimal Use of Different COVID-19 Vaccines: Issues of Viral Variants and Vaccine Efficacy. *JAMA*. 2021;325(13):1251–2. <https://doi.org/10.1001/jama.2021.3465>.
13. Fielding-Miller RK, Sundaram ME, Brouwer K. Social determinants of COVID-19 mortality at the county level. *PLoS One*. 2020;15(10):e0240151.
14. Vahabi N, Salehi M, Duarte JD, Mollalo A, Michailidis G. County-level longitudinal clustering of COVID-19 mortality to incidence ratio in the United States. *Sci Rep*. 2021;11(1):3088.
15. Ahmad K, Erqou S, Shah N, et al. Association of poor housing conditions with COVID-19 incidence and mortality across US counties. *PLoS One*. 2020;15(11):e0241327.
16. Wu X, Nethery RC, Sabath MB, Braun D, Dominici F. Air pollution and COVID-19 mortality in the United States: Strengths and limitations of an ecological regression analysis. *Sci Adv*. 2020;6(45):eabd4049. <https://doi.org/10.1126/sciadv.abd4049>.
17. Karmakar M, Lantz PM, Tipirneni R. Association of Social and Demographic Factors With COVID-19 Incidence and Death Rates in the US. *JAMA Netw Open*. 2021;4(1):e2036462.
18. Mori M, Krumholz HM, Allore HG. Using Latent Class Analysis to Identify Hidden Clinical Phenotypes. *JAMA*. 2020;324(7):700–1.
19. Weller BE, Bowen NK, Faubert SJ. Latent class analysis: a guide to best practice. *J Black Psychol*. 2020;46(4):287–311.
20. Pearce N, Vandenbroucke JP, VanderWeele TJ, Greenland S. Accurate Statistics on COVID-19 Are Essential for Policy Guidance and Decisions. *Am J Public Health*. 2020:e1–3.
21. Smith N, Fraser M. Straining the System: Novel Coronavirus (COVID-19) and Preparedness for Concomitant Disasters. *Am J Public Health*. 2020;110(5):648–9.
22. Chopra V, Toner E, Waldhorn R, Washer L. How Should U.S. Hospitals Prepare for Coronavirus Disease 2019 (COVID-19)? *Ann Intern Med*. 2020;172(9):621–2. <https://doi.org/10.7326/M20-0907>. Epub 2020 Mar 11.
23. Khazanchi R, Beiter ER, Gondi S, Beckman AL, Bilinski A, Ganguli I. County-Level Association of Social Vulnerability with COVID-19 Cases and Deaths in the USA. *J Gen Intern Med*. 2020;35(9):2784–7.
24. Wolkin A, Patterson JR, Harris S, et al. Reducing Public Health Risk During Disasters: Identifying Social Vulnerabilities. *J Homel Secur Emerg Manag*. 2015;12(4):809–22.
25. Roth GA, Dwyer-Lindgren L, Bertozzi-Villa A, et al. Trends and Patterns of Geographic Variation in Cardiovascular Mortality Among US Counties, 1980–2014. *JAMA*. 2017;317(19):1976–92.
26. Woolf SH, Schoemaker H. Life Expectancy and Mortality Rates in the United States, 1959–2017. *JAMA*. 2019;322(20):1996–2016.
27. Chetty R, Stepner M, Abraham S, et al. The Association Between Income and Life Expectancy in the United States, 2001–2014. *JAMA*. 2016;315(16):1750–66.
28. Peto R, Whitlock G, Jha P. Effects of obesity and smoking on U.S. life expectancy. *N Engl J Med*. 2010;362(9):855–6 author reply 856–857.
29. Pope CA 3rd, Ezzati M, Dockery DW. Fine-particulate air pollution and life expectancy in the United States. *N Engl J Med*. 2009;360(4):376–86.
30. Ferdows NB, Aranda MP, Baldwin JA, Baghban Ferdows S, Ahluwalia JS, Kumar A. Assessment of Racial Disparities in Mortality Rates Among Older Adults Living in US Rural vs Urban Counties From 1968 to 2016. *JAMA Netw Open*. 2020;3(8):e2012241.
31. Reyfman PA, Khuder B, Pierce JB, et al. Urban-Rural Mortality Disparities from Chronic Lower Respiratory Diseases in the US, 1999–2019. *Am J Respir Crit Care Med*. 2021;203(11):1435–7. <https://doi.org/10.1164/rccm.202011-4150LE>.
32. Shah NS, Carnethon M, Lloyd-Jones DM, Khan SS. Widening Rural-Urban Cardiometabolic Mortality Gap in the United States, 1999 to 2017. *J Am Coll Cardiol*. 2020;75(25):3187–8.
33. Khatana SAM, Venkataramani AS, Nathan AS, et al. Association Between County-Level Change in Economic Prosperity and Change in Cardiovascular Mortality Among Middle-aged US Adults. *JAMA*. 2021;325(5):445–53.
34. Sasson I, Hayward MD. Association Between Educational Attainment and Causes of Death Among White and Black US Adults, 2010–2017. *JAMA*. 2019;322(8):756–63.
35. Yancy CW. COVID-19 and African Americans. *JAMA*. 2020;323(19):1891–2. <https://doi.org/10.1001/jama.2020.6548>.
36. Moscola J, Sembajwe G, Jarrett M, et al. Prevalence of SARS-CoV-2 Antibodies in Health Care Personnel in the New York City Area. *JAMA*. 2020;324(9):893–5.
37. Sood N, Simon P, Ebner P, et al. Seroprevalence of SARS-CoV-2-Specific Antibodies Among Adults in Los Angeles County, California, on April 10–11, 2020. *JAMA*. 2020;323(23):2425–7.
38. Stokes AC, Lundberg DJ, Elo IT, Hempstead K, Bor J, Preston SH. COVID-19 and excess mortality in the United States: A county-level analysis. *PLoS Med*. 2021;18(5):e1003571.
39. Sedgwick P. Ecological studies: advantages and disadvantages. *BMJ*. 2014;348:g2979.
40. Wadhwa RK, Wadhwa P, Gaba P, et al. Variation in COVID-19 Hospitalizations and Deaths Across New York City Boroughs. *JAMA*. 2020;323(21):2192–5.
41. Munoz-Price LS, Nattinger AB, Rivera F, et al. Racial Disparities in Incidence and Outcomes Among Patients With COVID-19. *JAMA Netw Open*. 2020;3(9):e2021892.
42. Pierce JB, Harrington K, McCabe ME, et al. Racial/ethnic minority and neighborhood disadvantage leads to disproportionate mortality burden and years of potential life lost due to COVID-19 in Chicago, Illinois. *Health Place*. 2021;68:102540.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.