



Original Investigation | Public Health

# Quantification of Neighborhood-Level Social Determinants of Health in the Continental United States

Marynia Kolak, PhD, MFA, MS; Jay Bhatt, DO; Yoon Hong Park, MPP; Norma A. Padrón, PhD, MPH, MA; Ayrin Molefe, PhD

## Abstract

**IMPORTANCE** An association between social and neighborhood characteristics and health outcomes has been reported but remains poorly understood owing to complex multidimensional factors that vary across geographic space.

**OBJECTIVES** To quantify social determinants of health (SDOH) as multiple dimensions across the continental United States (the 48 contiguous states and the District of Columbia) at a small-area resolution and to examine the association of SDOH with premature mortality within Chicago, Illinois.

**DESIGN, SETTING, AND PARTICIPANTS** In this cross-sectional study, census tracts from the US Census Bureau from 2014 were used to develop multidimensional SDOH indices and a regional typology of the continental United States at a small-area level ( $n = 71\,901$  census tracts with approximately 312 million persons) using dimension reduction and clustering machine learning techniques (unsupervised algorithms used to reduce dimensions of multivariate data). The SDOH indices were used to estimate age-adjusted mortality rates in Chicago ( $n = 789$  census tracts with approximately 7.5 million persons) with a spatial regression for the same period, while controlling for violent crime.

**MAIN OUTCOMES AND MEASURES** Fifteen variables, measured as a 5-year mean, were selected to characterize SDOH as small-area variations for demographic characteristics of vulnerable groups, economic status, social and neighborhood characteristics, and housing and transportation availability at the census-tract level. This SDOH data matrix was reduced to 4 indices reflecting advantage, isolation, opportunity, and mixed immigrant cohesion and accessibility, which were then clustered into 7 distinct multidimensional neighborhood typologies. The association between SDOH indices and premature mortality (defined as death before age 75 years) in Chicago was measured by years of potential life lost and aggregated to a 5-year mean. Data analyses were conducted between July 1, 2018, and August 30, 2019.

**RESULTS** Among the 71 901 census tracts examined across the continental United States, a median (interquartile range) of 27.2% (47.1%) of residents had minority status, 12.1% (7.5%) had disabilities, 22.9% (7.6%) were 18 years and younger, and 13.6% (8.1%) were 65 years and older. Among the 789 census tracts examined in Chicago, a median (interquartile range) of 80.4% (56.3%) of residents had minority status, 10.2% (8.2%) had disabilities, 23.2% (10.9%) were 18 years and younger, and 9.5% (7.1%) were 65 years and older. Four SDOH indices accounted for 71% of the variance across all census tracts in the continental United States in 2014. The SDOH neighborhood typology of extreme poverty, which is of greatest concern to health care practitioners and policy advocates, comprised only 9.6% of all census tracts across the continental United States but characterized small areas of known public health crises. An association was observed between all SDOH indices and age-adjusted premature mortality rates in Chicago ( $R^2 = 0.63$ ;  $P < .001$ ), even after accounting for violent crime and spatial structures.

*(continued)*

**Open Access.** This is an open access article distributed under the terms of the CC-BY License.

## Key Points

**Question** How do social determinants of health vary across multiple dimensions and geographic space?

**Findings** In this cross-sectional study of 71 901 census tracts with approximately 312 million persons across the continental United States, multivariate social determinants of health measures were reduced to 4 indices reflecting advantage, isolation, opportunity, and mixed immigrant cohesion and accessibility and were clustered into 7 neighborhood typologies that included an extreme poverty group. Social determinants of health indices were associated with premature mortality rates in Chicago, Illinois.

**Meaning** The use of multidimensional geospatial approaches to quantify social determinants of health rather than the use of a singular deprivation index may better capture the complexity and spatial heterogeneity underlying these determinants.

## + Supplemental content

Author affiliations and article information are listed at the end of this article.

Abstract (continued)

**CONCLUSIONS AND RELEVANCE** The modeling of SDOH as multivariate indices rather than as a singular deprivation index may better capture the complexity and spatial heterogeneity underlying SDOH. During a time of increased attention to SDOH, this analysis may provide actionable information for key stakeholders with respect to the focus of interventions.

JAMA Network Open. 2020;3(1):e1919928. doi:10.1001/jamanetworkopen.2019.19928

## Introduction

The consequences of social determinants of health (SDOH) increasingly dominate public health discussions in the United States, as population health outcomes have not kept pace with those of other developed nations despite higher per-person spending for medical services.<sup>1-3</sup> An increased understanding of SDOH could be used to better connect patients with relevant social services in clinical contexts and could target vulnerable populations with health-improving social policies and programs while also addressing affordability.<sup>4,5</sup> Health policy frameworks that directly address the underlying social and behavioral determinants of health are now encouraged to promote improvements in population health outcomes and cost savings.<sup>4</sup> However, more finely quantifying these determinants at scale remains a challenge.

Social determinants of health are defined by the World Health Organization as the conditions in which people are born, grow, live, work, and age.<sup>6</sup> Despite this complex and nuanced description of SDOH, the phenomena are often represented solely by socioeconomic indicators, such as income and education. Social determinants of health indicators, such as income, are associated with greater life expectancy in the United States; however, these associations are complex and may change based on underlying area characteristics and health behaviors.<sup>7</sup> While these variables are associated with health outcomes, they are also likely to be associated with each other, resulting in issues of multicollinearity and presenting challenges for meaningful interpretation.<sup>8</sup>

Index-based approaches have also emerged to serve as proxies to assess the consequences of SDOH and have incorporated various methods. The area deprivation index (ADI) by Singh et al,<sup>9-11</sup> which was extended by Kind et al,<sup>12</sup> focused on socioeconomic disadvantage and the differing dimensions of poverty. The North Carolina Institute of Public Health and the Carolinas HealthCare System developed a conceptual model-driven SDOH index that includes additional dimensions of health-related social needs, such as food accessibility; however, this index has not yet been validated against actualized health outcomes.<sup>8</sup> Other conceptual frameworks incorporate a wide range of indicators with a broad range of applications for data collection, clinical use, and research. Proxies used to assess SDOH remain varied and unstandardized and may obfuscate the underlying factors involved in differential health outcomes. These approaches may also overlook the geographic heterogeneity of the US population or the ways in which uniquely characterized neighborhoods, such as retirement communities or diverse urban immigrant enclaves, complicate a unidimensional application of SDOH.

We addressed these challenges by implementing a cross-sectional multivariate analysis of SDOH components using a principal component and regionalization cluster analyses. In a cross-sectional view, SDOH may have more explanatory power for health outcomes than do the differences associated with changes in medical care or technology.<sup>8</sup> We hypothesized that actual SDOH outcomes may vary from existing conceptual models, although they should remain associated with health outcomes in meaningful ways. In the first phase, we developed a multidimensional SDOH data matrix for the continental United States (the 48 contiguous states and the District of Columbia) at the census-tract level for 2014. Census tracts are small-area geographic regions defined by the US Census Bureau for analyzing populations that approximate neighborhoods containing between 2500 and 8000 persons. We then developed a multidimensional index based on the results of a principal component analysis and imported the findings into spatially sensitive clusters or typologies

of SDOH across the continental United States. In the second phase, we associated the SDOH multidimensional index with age-adjusted premature mortality within Chicago, Illinois, a large metropolitan city with recognized challenges in health outcomes and social and economic disparities.

---

## Methods

### Population and Spatial Scale

In this cross-sectional multivariate analysis, the first phase of the study included all populated census tracts of the continental United States ( $n = 71\,901$ ), with a total observed population of approximately 312 million persons based on census estimates. To create the index, we used only the continental states to preserve neighboring relationships across tracts regardless of state boundary. In the second phase, we validated the index using a subset of Chicago with 789 census tracts that included approximately 7.5 million persons. We chose the census tract as the spatial scale of our analysis to estimate small-area variations at the neighborhood level. A total of 6908 of 71 901 census tracts (9.6%) were in the extreme poverty group (**Table 1**). A waiver of informed consent was granted by the institutional review board of the University of Chicago because this study used nonidentifiable public data sets that did not constitute human subjects research. Data analyses were conducted between July 1, 2018, and August 30, 2019 (eMethods in the [Supplement](#)). We used the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) checklist for cross-sectional analysis to guide the reporting of results.

### Data Definitions

#### Social Determinants of Health

Existing SDOH conceptual frameworks acknowledge the underlying indicators that are likely important factors in differential health outcomes. We reviewed multiple conceptual SDOH frameworks to identify indicators that could be measured in an empirical model at scale, including the Danaher, World Health Organization, and California Department of Public Health frameworks and the American Heart Association and American College of Cardiology SDOH guidelines and measures that are common in healthy places models.<sup>13-17</sup> Fifteen variables were chosen to capture small-area variations in economic status, social and neighborhood characteristics, housing and transportation availability, and demographic characteristics of vulnerable groups. We stratified these indicators across 3 broad topics that are common to multiple SDOH frameworks, representing social, economic, and physical environments. To facilitate more meaningful interpretation between these concepts, we also mapped each indicator selected onto the socioecological model of health,<sup>18</sup> which was adapted to include individual, interpersonal, organizational, and community spheres (**Figure 1**).

Demographic features of socioeconomically vulnerable groups include areas with high proportions of older adults (aged  $\geq 65$  years), persons with minority status, and persons with disabilities, which are all features associated with differentiating health outcomes and behaviors.<sup>18-20</sup> Additional social disadvantages, such as limited English language proficiency<sup>21,22</sup> and single-parent households,<sup>23,24</sup> are also associated with health outcomes. Vulnerable age groups, single-parent households, persons with limited English proficiency, and persons with minority status were mapped to the interpersonal sphere, which reflected localized social networks, family, and cultures. Disability status by neighborhood proportion was mapped across interpersonal and organizational spheres, reflecting complex interactions within social networks and organizations that are factors in social inclusion.<sup>25</sup>

We chose standard economic indicators with known associations with health outcomes, including poverty,<sup>15,16,26</sup> income,<sup>7,27</sup> unemployment,<sup>28,29</sup> education,<sup>4,27,30-32</sup> and health insurance status.<sup>33,34</sup> Educational level and health insurance status are included as service environments in some SDOH conceptual frameworks.<sup>15</sup> Educational level and health insurance status thus crossed both organizational and community spheres in our socioecological model adaptation. We included

poverty, income, and unemployment status in the community sphere, which reflected structural societal properties.

Table 1. Summary Statistics of Social Determinants of Health Typologies for United States and Chicago Tracts<sup>a</sup>

Variable	Median (IQR) [CV]													
	Rural Affordable		Vibrant Urban Core		Suburban Affordable		Extreme Poverty		Multilingual Working		Suburban Affluent		Sparse Areas	
	US	Chicago	US	Chicago	US	Chicago	US	Chicago	US	Chicago	US	Chicago	US	Chicago
Census tracts, No.	19 512	28	4619	230	14 017	33	6908	249	6352	141	17 811	102	2682	8
Population, thousands	76 282	106.4	77 530	841.9	73 473	140.2	22 662	787.4	31 304	594.4	19 602	403.7	11 131	32.1
Ethnic/racial minority, %	15.3 (27.2) [1.0]	97.0 (19.5) [0.2]	47.4 (44.6) [0.5]	40.5 (34.0) [0.5]	31.4 (36.8) [0.7]	68.0 (47.0) [0.5]	78.9 (43.6) [0.4]	98.0 (2.0) [0.1]	86.7 (20.1) [0.2]	89.0 (18.0) [0.1]	16.7 (21.9) [0.8]	33.0 (24.5) [0.5]	12.9 (19.8) [1.0]	28.5 (72.2) [0.8]
Aged ≥65 y, %	16.9 (5.7) [0.3]	19.9 (5.3) [0.2]	10.2 (7.9) [0.6]	8.7 (7.9) [0.7]	9.7 (4.9) [0.4]	8.9 (3.7) [0.3]	10.3 (5.8) [0.4]	10.6 (7.2) [0.5]	8.6 (5.3) [0.5]	7.5 (3.6) [0.4]	15.7 (6.2) [0.3]	12.4 (6.4) [0.4]	32.9 (15.5) [0.4]	32.2 (12.4) [0.3]
Aged ≤18 y, %	22.0 (4.9) [0.2]	20.6 (6.7) [0.3]	14.7 (11.8) [0.5]	13.7 (9.5) [0.5]	27.2 (5.3) [0.2]	26.7 (5.1) [0.1]	27.0 (8.2) [0.3]	27.8 (8.7) [0.2]	28.3 (8.4) [0.2]	28.4 (7.0) [0.2]	20.7 (5) [0.2]	21.0 (4.8) [0.2]	12.3 (6.9) [0.5]	14.1 (5.3) [0.4]
Disability, %	16.6 (5.1) [0.2]	16.0 (4.2) [0.2]	9.1 (6.2) [0.5]	7.0 (6.0) [0.6]	9.0 (4.4) [0.3]	8.0 (2.0) [0.2]	17.2 (7.0) [0.3]	15.0 (5) [0.3]	10.1 (5.0) [0.4]	8.0 (3.0) [0.3]	9.8 (4.1) [0.3]	9.0 (4.0) [0.4]	20.5 (8.9) [0.4]	22.5 (8.5) [0.3]
No high school diploma, %	14.5 (9.5) [0.4]	11.5 (12.2) [0.5]	10.9 (14.2) [0.8]	7.0 (12.0) [0.9]	7.8 (8.1) [0.7]	16.0 (16.0) [0.6]	21.9 (12.2) [0.4]	21.0 (12.0) [0.4]	35.0 (17.8) [0.3]	37.0 (17.0) [0.3]	5.8 (5.3) [0.7]	9.0 (9.0) [0.6]	10.4 (10.8) [0.8]	14.5 (7.8) [0.6]
Limited English proficiency, %	0.6 (2.1) [1.6]	0 (3.0) [1.6]	4.5 (10.1) [1.1]	2.5 (8.0) [1.2]	1.6 (3.6) [1.1]	8.0 (12.0) [0.8]	1.2 (4.7) [1.4]	0 (1.0) [2.1]	19.7 (12.0) [0.4]	25.0 (8.0) [0.3]	2.1 (2.4) [1.3]	5.0 (8.0) [0.9]	0.8 (2.7) [2.0]	15.0 (11.0) [1.1]
Single parent, %	8.5 (5.3) [0.4]	9.7 (4.2) [0.4]	5.7 (8.5) [0.9]	4.0 (6.0) [0.9]	10.0 (6.9) [0.5]	11.2 (6.9) [0.4]	19.8 (9.4) [0.4]	22.1 (11.4) [0.4]	14.2 (9.2) [0.5]	14.1 (8.7) [0.4]	5.3 (4.2) [0.6]	5.2 (4.5) [0.6]	3.4 (4.1) [0.8]	2.3 (3.1) [1.7]
Living in poverty, %	16.1 (10.4) [0.5]	17.5 (6.4) [0.3]	18.4 (18.0) [0.7]	16.4 (13.1) [0.5]	9.1 (9.7) [0.7]	13.2 (10.1) [0.6]	37.0 (15.9) [0.3]	38.5 (17.7) [0.3]	25.7 (16.6) [0.4]	24.7 (11.8) [0.4]	6.5 (6.3) [0.8]	8.1 (8.8) [0.6]	12.2 (12.1) [0.8]	12.8 (16.2) [0.7]
Per capita income, \$	22 382 (6273) [0.2]	21 773 (4874) [0.2]	28 959 (2433) [0.7]	41 546 (32 931) [0.5]	27 955 (11 403) [0.3]	23 302 (9323) [0.3]	14 545 (5512) [0.3]	14 210 (6398) [0.3]	15 648 (7082) [0.3]	15 108 (6075) [0.3]	36 383 (15 323) [0.4]	32 961 (17 268) [0.4]	28 383 (15 291) [0.5]	27 310 (2819) [0.6]
Unemployed, %	9.5 (5.7) [0.4]	18.0 (7.5) [0.3]	8.5 (6.4) [0.5]	7.0 (7.0) [0.6]	7.2 (5.1) [0.5]	11.0 (5.0) [0.3]	19.0 (10) [0.4]	26.0 (12.0) [0.3]	11.6 (6.7) [0.4]	12.0 (5.0) [0.3]	6.1 (3.9) [0.5]	8.0 (4.8) [0.4]	9.5 (7.3) [0.7]	10.5 (6.8) [0.6]
Uninsured, %	14.4 (8.2) [0.4]	15.2 (5.2) [0.3]	12.5 (11.2) [0.6]	12.2 (12.6) [0.6]	10.9 (9.3) [0.6]	13.7 (13.3) [0.5]	18.1 (9.5) [0.4]	19.7 (7.9) [0.3]	29.6 (12.1) [0.3]	29.9 (7.6) [0.2]	7.1 (6.3) [0.6]	11.2 (11.0) [0.6]	11.8 (10.4) [0.6]	7.0 (8.4) [0.8]
Renter, %	25.0 (8.9) [0.5]	26.0 (10.5) [0.5]	73.0 (25.0) [0.2]	63.0 (15.0) [0.2]	25.0 (25.0) [0.7]	19.0 (19.0) [0.6]	60.0 (24.0) [0.3]	67.0 (24.0) [0.3]	51.0 (32.0) [0.4]	57.0 (19.0) [0.3]	23.0 (24) [0.7]	35.0 (25.0) [0.5]	22.0 (23.0) [0.8]	48.5 (34.5) [0.5]
Rent burden, %	12.0 (8.9) [0.5]	13.5 (11.8) [0.6]	26.8 (27.9) [0.8]	26.0 (25.5) [0.7]	9.2 (10.7) [0.8]	11.0 (12.0) [0.7]	18.6 (18.1) [0.7]	23.0 (25.0) [0.8]	13.0 (16.1) [0.9]	20.0 (17.0) [0.6]	15.6 (13.1) [0.6]	18.5 (12.5) [0.5]	16.6 (14.1) [0.6]	31.5 (25.0) [0.6]
Crowded housing, %	1.6 (2.5) [1.0]	1.0 (3.0) [1.2]	3.4 (6.4) [1.1]	1.0 (3.0) [1.3]	1.8 (3.1) [1.1]	3.0 (4.0) [0.9]	3.0 (4.4) [1.0]	4.0 (4.0) [0.8]	13.1 (10.8) [0.6]	10.0 (9.0) [0.6]	0.7 (1.9) [1.4]	1.0 (2.0) [1.4]	0.6 (1.9) [1.5]	0.5 (2.0) [1.4]
No vehicle, %	5.5 (5.7) [0.7]	15.2 (8.6) [0.4]	27.8 (35.0) [0.7]	30.6 (19.3) [0.4]	2.6 (3.8) [1.0]	7.1 (6.8) [0.6]	21.7 (17.6) [0.6]	36.8 (17.7) [0.3]	9.3 (11.0) [0.9]	17.6 (9.3) [0.4]	3.8 (5.1) [0.9]	10.4 (6.2) [0.5]	6.4 (9.8) [1.2]	39.0 (24.0) [0.5]
Advantage score <sup>b</sup>	0.5 (1.6) [3.4]	-0.8 (0.9) [-0.9]	-0.8 (3.1) [-2.5]	0.3 (3.3) [-48.9]	0.9 (2.1) [2.3]	-0.7 (2.8) [-3.2]	-3.1 (2.3) [-0.5]	-4.4 (2.5) [-0.4]	-4.0 (2.8) [-0.5]	-4.4 (2.3) [-0.4]	2.1 (1.4) [0.6]	0.9 (2.2) [1.7]	1.7 (2.3) [1.3]	0.9 (2.7) [306.6]
Mobility score <sup>c</sup>	-0.7 (1.4) [-0.9]	-1.4 (1.2) [-0.8]	-0.2 (1.4) [-3.6]	0.2 (1.3) [-19.0]	1.1 (1.0) [0.6]	1.2 (0.5) [0.4]	-1.2 (1.6) [-0.9]	-1.6 (1.3) [-0.7]	1.6 (1.4) [0.6]	1.6 (1.1) [0.5]	0.4 (1.0) [1.9]	0.7 (0.8) [1.0]	-2.5 (1.6) [-0.6]	-3.1 (1.6) [-0.4]
Opportunity score <sup>d</sup>	-0.7 (0.7) [-0.8]	0 (0.8) [-13.1]	2.6 (1.5) [0.4]	2.8 (1.3) [0.3]	-0.6 (0.7) [-0.9]	-0.5 (0.6) [-1.0]	-0.4 (1.3) [-2.7]	0.1 (1.2) [4.6]	0.1 (1.4) [5.2]	0.7 (1.0) [1.0]	0.4 (0.9) [1.3]	1.0 (0.9) [0.7]	0.6 (1.5) [1.8]	1.9 (1.4) [0.4]
MICA score <sup>e</sup>	-0.4 (0.7) [1.4]	0 (0.7) [6.7]	0.70 (1.3) [-1.4]	1.00 (1.1) [-1.0]	0.5 (0.7) [-1.0]	0 (0.9) [-2.7]	1.3 (1.2) [-0.6]	2.2 (1.6) [-0.5]	-1.5 (1.4) [0.6]	-1.2 (1.0) [0.6]	0.1 (0.7) [-9.2]	0.2 (1.1) [-5.0]	-1.6 (1.0) [0.6]	-0.9 (0.8) [0.6]

Abbreviations: CV, coefficient of variation; IQR, interquartile range; MICA, mixed immigrant cohesion and accessibility index.

<sup>a</sup> Summary statistics are for each variable within each typology for US and Chicago tracts. Summaries for each typology can be considered with respect to the median estimates shown in Table 2.

<sup>b</sup> First primary component; Advantage indicates socioeconomic advantage index.

<sup>c</sup> Second primary component; Mobility indicates limited mobility index.

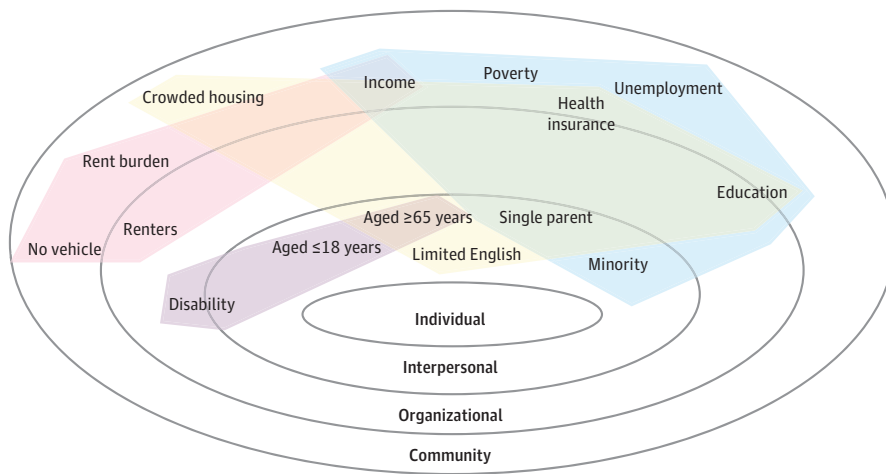
<sup>d</sup> Third primary component; Opportunity indicates urban core opportunity index.

<sup>e</sup> Fourth primary component.

The physical grouping included housing and transportation measures. Studies have also documented associations between housing status and health risks<sup>35-38</sup> as well as the consequences of vehicle access on health behaviors.<sup>39</sup> To serve as a proxy for housing characteristics, we included the proportion of renters, rent burden (ie, more than 30% of income used for rent payments), and crowded housing conditions (ie, occupied housing units consisting of more people than rooms). Crowded housing and rent burden were included in the community sphere, which reflected wider patterns of housing availability and cost, while the proportion of renters was identified across organizational and community spheres. To serve as a proxy for vehicle access, we included households with no vehicle in the community sphere as a structural component that may reflect wider geographic patterns.

We limited the analysis to data available as continuous variables across continental tracts in the United States for the period of interest. All variables were derived from the 2014 American Community Survey 5-year mean. The associated proxies and the data sources are presented in **Table 2**.

**Figure 1. Socioecological Model of Health and Social Determinants of Health Indices**



Social determinants of health indicators are mapped onto different spheres of the socioecological model of health adapted from Bronfenbrenner.<sup>18</sup> Principal components include indicators for each index. Blue shading indicates socioeconomic advantage index, purple shading indicates limited mobility index, red shading indicates urban core opportunity index, and yellow shading indicates mixed immigrant cohesion and accessibility index.

**Table 2. Social Determinants of Health Data<sup>a</sup>**

Variable	Median (IQR) [CV]	
	US (n = 71901)	Chicago (n = 789)
Ethnic/racial minority <sup>b</sup>	27. (247.1) [0.8]	80.4 (56.3) [0.4]
Aged ≥65 y	13.6 (8.1) [0.5]	9.5 (7.1) [0.6]
Aged ≤18 y	22.9 (7.6) [0.3]	23.2 (10.9) [0.4]
Disability <sup>c</sup>	12.1 (7.5) [0.5]	10.2 (8.2) [0.5]
No high school diploma <sup>d</sup>	11.4 (13.5) [0.8]	17.4 (18.7) [0.7]
Limited English proficiency <sup>e</sup>	1.3 (4.6) [1.7]	3.3 (13.5) [1.2]
Single parent <sup>f</sup>	8.4 (7.9) [0.7]	11.1 (14.1) [0.8]
Living in poverty	13.3 (15.7) [0.8]	21.9 (21.8) [0.6]
Per capita income, \$	25 174 (14 511) [0.5]	21 232 (19 127) [0.7]
Unemployed <sup>g</sup>	8.5 (6.7) [0.6]	12.9 (12.8) [0.7]
Uninsured <sup>h</sup>	12.6 (11.5) [0.6]	17.9 (12.8) [0.5]
Renter	31 (32) [0.6]	58 (27) [0.4]
Rent burden <sup>i</sup>	13.4 (13.3) [0.8]	21.7 (21.0) [0.7]
Crowded housing <sup>j</sup>	1.8 (3.7) [1.5]	3.3 5.9 [1.00]
No vehicle	5.5 (8.8) [1.3]	24.1 (22.9) [0.6]

Abbreviations: CV, coefficient of variation; IQR, interquartile range.

<sup>a</sup> Data are estimates from the 2014 American Community Survey 5-year mean.

<sup>b</sup> Defined as persons of all racial/ethnic ancestries with the exception of white, non-Hispanic ancestry.

<sup>c</sup> Persons in the civilian noninstitutionalized population.

<sup>d</sup> Persons aged 25 years or older.

<sup>e</sup> Persons aged 5 years or older.

<sup>f</sup> Households with children aged less than 18 years.

<sup>g</sup> Civilians aged 16 years and older.

<sup>h</sup> Persons in the total civilian noninstitutionalized population.

<sup>i</sup> Renters paying more than 30% of their household income for rent.

<sup>j</sup> Defined as occupied housing units consisting of more people than rooms.

## Health Outcomes and Violent Crime

To estimate associations between social determinants of health and health outcomes for a subset of data, the mortality rate at the census-tract level was used for the Chicago, for which we had sufficiently high-quality direct measurements of premature mortality. These data were obtained from the Chicago Department of Public Health. Premature mortality (ie, death before age 75 years) was measured by years of potential life lost, aggregated to a 5-year mean (2009-2013) and calculated as an age-adjusted rate at the census-tract level.

We accessed crime data from the Chicago data portal and obtained 2014 data from a subset of violent crime categories, including battery, assault, robbery, and homicide. We then used the available geocodes to translate crimes as spatial point data, and we spatially joined intersecting census tracts and aggregated data by tract using the *spdep* package (version 1.1-3) in the R software environment (R Foundation for Statistical Computing). To obtain a violent crime rate, we calculated the total number of violent crimes per population of associated census tract.

## Statistical Analysis

All statistical tests were 2-sided with a significance level of  $P = .05$ . All statistical analyses were performed using the R software environment (version 3.6.1), GeoDa (version 1.14.0), and GeoDaSpace (version 1.1) statistical software (Center for Spatial Data Science).

## Exploratory Factor Analysis

We implemented a principal component analysis of the SDOH matrix. In a principal component analysis, multiple variables are reduced to core components, with each component orthogonal (ie, not correlated) by construction to preserve the most information given by their variances. This approach is similar to the original construction of the Singh ADI, which used principal component analysis for index development.<sup>9</sup> We extended the Singh methodology with a more complex conceptual model of SDOH that incorporated additional variables reflecting multiple dimensions of health, and we further minimized variables that were likely to be associated with each other (eg, families classified as living in poverty and households with poverty levels of more than 150%). Because the first index approximated socioeconomic disadvantage, we examined its association with the Singh ADI index using a Spearman correlation coefficient. We also developed a composite SDOH index by summing principal component scores.

We implemented the principal component analysis using the singular-value decomposition method and initial variable standardization, and we used the Kaiser criterion (ie, components with eigenvalues of less than 1.0 were excluded) to determine the final number of components to retain. In accordance with literature standards, we used 0.30 variable loading for the component as the cutoff for the for determining the dominant variables within each principal component. Final components, which together accounted for most of the variance in all 15 SDOH variables, were visualized as SD maps.

## Regionalization Analysis

We conducted a dimension-reducing clustering analysis to decompose tracts into typologies that had similar SDOH characteristics. Similar to principal component analyses, this clustering analysis is a machine learning technique that uses unsupervised algorithms to reduce dimensions of multivariate data. To examine regional typologies, we collapsed the 4 dimensions associated with SDOH into a single dimension to give insight into the spatial heterogeneity of outcomes and examine the reasons that different results occurred in different places. The most important aspects of the construction were that each typology had similar attributes within its own grouping and that each typology was distinct. The attributes used for the analysis were distinct principal components that reflected unique phenomena (orthogonal by construction), which provided more meaningful results than would loading the algorithm with dozens of highly correlated variables. Region types are useful for a variety of methodological needs, such as the need to use 1 dimension of data to compare and contrast

neighborhoods across the continental United States or to identify similar areas for matching analyses in quasi-experimental studies.

Consistent with work that has used algorithms to define areas with both geographic and socioeconomic components,<sup>40-42</sup> we implemented multivariate regionalization techniques that joined areas based on attribute similarity and a minimal degree of spatial connectivity. We inputted dominant principal components into a k-means clustering analysis to identify k groups of tracts with similar SDOH characteristics. Because census tracts are nonphysical boundaries, in which residents are able to easily cross barriers and affect nearby tracts, it was important to account for spatial proximity between tracts. Geometric centroids of census tracts were thus given a 0.10 weighting within the k-means analysis to enforce a minimal but explicit spatial sensitivity. The k-means analysis was performed for 2 through 15 clusters for 71 901 tracts using the Arthur and Vassilvitskii<sup>43</sup> k-means++ procedure. The number of clusters that retained the lowest change in the proportion of between to total within-cluster sum of squares was chosen as an optimum. The typologies of the final clusters were summarized with descriptive statistics and visualized as an interactive map.

### Regression Analysis

We estimated associations between premature mortality rates in Chicago using the 4 indices derived from the dominant principal components while controlling for the violent crime rate. The indices were used as input to retain the greatest information rather than as k-means clusters because transforming the continuous principal components to discrete clusters invariably results in the loss of information. First, we estimated a linear regression model using ordinary least squares (OLS) estimation. A spatial weight matrix was constructed for the census tract data set using second-order queen contiguity, assigning bordering neighbors (up to 2 census tracts away) for each tract. We implemented a spatial autoregressive (SAR) model, incorporating spatially lagged mortality and 2-stage least squares estimation.<sup>44</sup>

---

## Results

Among the 71 901 census tracts (n = 312 million persons) examined across the continental United States, a median (interquartile range [IQR]) of 27.2% (47.1%) of residents had minority status, 12.1% (7.5%) had disabilities, 22.9% (7.6%) were 18 years and younger, and 13.6% (8.1%) were 65 years and older. Among the 789 census tracts (n = 7.5 million persons) examined in Chicago, a median (IQR) of 80.4% (56.3%) of residents had minority status, 10.2% (8.2%) had disabilities, 23.2% (10.9%) were 18 years and younger, and 9.5% (7.1%) were 65 years and older. Additional baseline and result summary statistics are shown in Table 1 and Table 2.

### Principal Component Analysis

Four principal components—the socioeconomic advantage index, the limited mobility index, the urban core opportunity index, and the mixed immigrant cohesion and accessibility index—met the Kaiser criterion for inclusion. Together, they accounted for 71% of the variance in the 15 SDOH variables across all census tracts in the continental United States. Principal component analysis scores had a mean of 0 by construction and were not further standardized to retain their original distributions (box and whisker plots of each component available in eFigure 1 in the [Supplement](#)).

The first principal component, the socioeconomic advantage index, accounted for 40.0% of the total variance and was dominated by socioeconomic status factors, including poverty, low high school graduation rates, minority status, proportion of uninsured persons, and number of single-parent households. We adjusted for cardinality so that areas with low proportions of socioeconomic disadvantage were positive and labeled this component the socioeconomic advantage index. This index could thus serve as a proxy for multidimensional poverty or socioeconomic disadvantage.

While other principal components reflected additional dimensions of socioeconomic disadvantage, the first index was principally characterized by the classic measures of socioeconomic

factors, such as poverty, minority status, and educational level. The variables of minority status (−0.32), no high school diploma (−0.34), living in poverty (−0.32), and uninsured (−0.31) were associated with each other, with higher absolute variable loadings in the same direction (eTable 1 and eTable 2 in the [Supplement](#)), thus reflecting a unique phenomenon. This index's association between minority status and poverty may reflect the role that racial segregation has played in perpetuating environments that are associated with health disparities.<sup>45,46</sup> Comparison with the ADI index at the tract level indicated a Spearman rank correlation of 0.59 between aggregated (as means) ADI national ranks and aggregated (as means) socioeconomic advantage index ranks. When identifying indicators that were important factors of this component in our adapted socioecological model, we found that this phenomenon was present across all spheres of health (interpersonal, organizational, and community), as mapped in Figure 1.

The second principal component, the limited mobility index, accounted for 13.4% of the variance and was dominated by areas with high proportions of older adults and persons with disabilities; these areas were negatively associated with the presence of children. Dominant principal component variable loadings for older adults (−0.41) and persons with disabilities (−0.58) were in opposite directions to children (0.34), as shown in eTable 1 in the [Supplement](#). Such neighborhoods may reflect uniquely vulnerable, isolated populations with limited mobility; thus, we labeled this component the limited mobility index. This index was focused within the inner scales of our adapted socioecological model of health, which was primarily associated with interpersonal and organizational dimensions.

We used the terms mobility and isolation to systematically reflect the complex interactions between aging, disability, and transportation. Older adults are more likely to remain in their homes, make fewer trips when they do leave home, and travel shorter distances, reflecting age-specific mobility issues even after controlling for confounding factors.<sup>47</sup> These reduced mobility issues are factors in both resource accessibility and social isolation needs. Older adults are also at greater risk of losing critical components of their social ties as they age, making access to social capital of great importance for their health—and in the United States, both social capital and community support of older adults have decreased over time.<sup>48</sup> The utility of this geographic approach is illustrated by ongoing research to identify areas with a high concentration of older populations to improve transportation policy, social service facilities, and planning.<sup>49,50</sup>

Neighborhoods with a greater number of older adults and persons with disabilities may be at greater risk of mobility issues and social isolation, which poses unique challenges when considering SDOH and health outcomes. For example, accessibility to health care and food resources in areas disproportionately populated by persons with less mobility owing to age and disabilities would require different assumptions about minimum distance traveled and transportation modality. These areas may represent neighborhoods with more senior housing or retirement and aging communities in disadvantaged areas that are experiencing economic decline. The limited mobility index is thus most useful when considered alongside the other indices. For example, low-income, socially isolated older adults are especially sensitive to environmental events, such as heat emergencies.

The third principal component, the urban core opportunity index, which accounted for 9.6% of the variance, included highly urbanized populations with high opportunity and corresponding high costs. High-scoring areas of this index, as defined by dominant variable loadings of the principal component analysis, were associated with high per capita income (0.36), high proportion of renters (0.38), high rent burden (0.38), and households without a vehicle (0.43; eTable 1 and eTable 2 in the [Supplement](#)). These areas also had fewer children. This index was focused on the outer spheres of our adapted socioecological model of health, which were associated with organizational and community dimensions. These areas were characterized by their compact geographies, dense urban centers, and strong economies, which were likely factors in the high local incomes.

The walkability and diversity of the areas further reflected characteristics of new urbanism and postmodern cities, which have been discussed in urban analysis literature.<sup>51,52</sup> Although these areas may attract residents who seek housing that is close to their jobs, the high rent burden may be a



disadvantage of that location choice. These areas thus have high job opportunities and well-established transportation infrastructures, but they also have a high cost of living that may have negative consequences for vulnerable individuals residing within them.

The fourth principal component, the mixed immigrant cohesion and accessibility index, which accounted for 8.1% of the variance, was characterized by mostly immigrant or multilingual groups with traditional family structures and multiple accessibility stressors. Areas with low scores on the mixed immigrant cohesion and accessibility index, as defined by dominant variable loadings of the principal component analysis, had high proportions of families with limited English proficiency (−0.41), older adults (−0.42), and crowded housing (−0.31) as well as a lack of health insurance (−0.30), lower high school graduation rates (−0.31), and fewer single parent households (0.32; eTable 1 and eTable 2 in the [Supplement](#)). Like the socioeconomic advantage index, this principal component was mapped across all spheres of health (interpersonal, organizational, and community), as shown in Figure 1.

This index also reflected dimensions of disadvantage, but it did so in a slightly different way than the socioeconomic advantage index. Vulnerable immigrant communities were represented, but the communities had unique patterns regarding accessibility and social cohesion. We viewed access as a complex topic with multiple dimensions.<sup>53,54</sup> Because it was not clear why these residents lacked these specific resources, we assumed it was related to a structural failure in accessibility, such as affordability, adequacy of services, or awareness of services. At the same time, the higher rates of older adults and children and the smaller number of single-family households suggested more cohesive family units, stronger interpersonal ties, and social connections that may have served as protective factors. Ongoing research seeks to better understand the role of language and social connection development, which are factors associated with immigrant health.<sup>55</sup>

The immigrant experience of acculturative stress and the decrease in family cohesion over time pose further challenges to achieving better health outcomes.<sup>56</sup> Scores on the mixed immigrant cohesion and accessibility index were negative and low in multilingual typologies (−1.50 in the United States and −1.20 in Chicago) that were also likely to have negative and low scores on the socioeconomic advantage index (−4.0 in the United States and −4.4 in Chicago), reflecting socioeconomic stressors (Table 1) (eTable 2 in the [Supplement](#)).

Mapped indices, using SD data classification, indicate spatially heterogeneous patterns across the continental United States (**Figure 2**) and Chicago (eFigure 2 in the [Supplement](#)). Areas of greatest socioeconomic disadvantage were found in small and dense tracts of cities throughout the country from Los Angeles, California, to Baltimore, Maryland, in addition to larger, more sparsely populated tracts across multiple Native American reservations across South Dakota, Wyoming, and Arizona. The southern border of Texas and a wide area through the southern United States were also likely to have more socioeconomic disadvantage, especially in comparison with the northern United States. A more heterogeneous pattern was found in the mobility-related isolation index; some areas with high proportions of older adults and persons with disabilities dominated the southern United States, Appalachia, southern and western sections of Chicago, and areas in parts of Florida and Phoenix, Arizona. The areas with the highest urban opportunity scores were limited to tightly bounded tracts in areas such as the centers of New York City and San Francisco, California, and the northern Chicago bordering the lake. Low outliers of the fourth principal component index reflected multilingual communities from southern Texas to the western and northwestern sides of Chicago, all with more traditional family structures and less advantageous accessibility characteristics.

### Regionalization Analysis

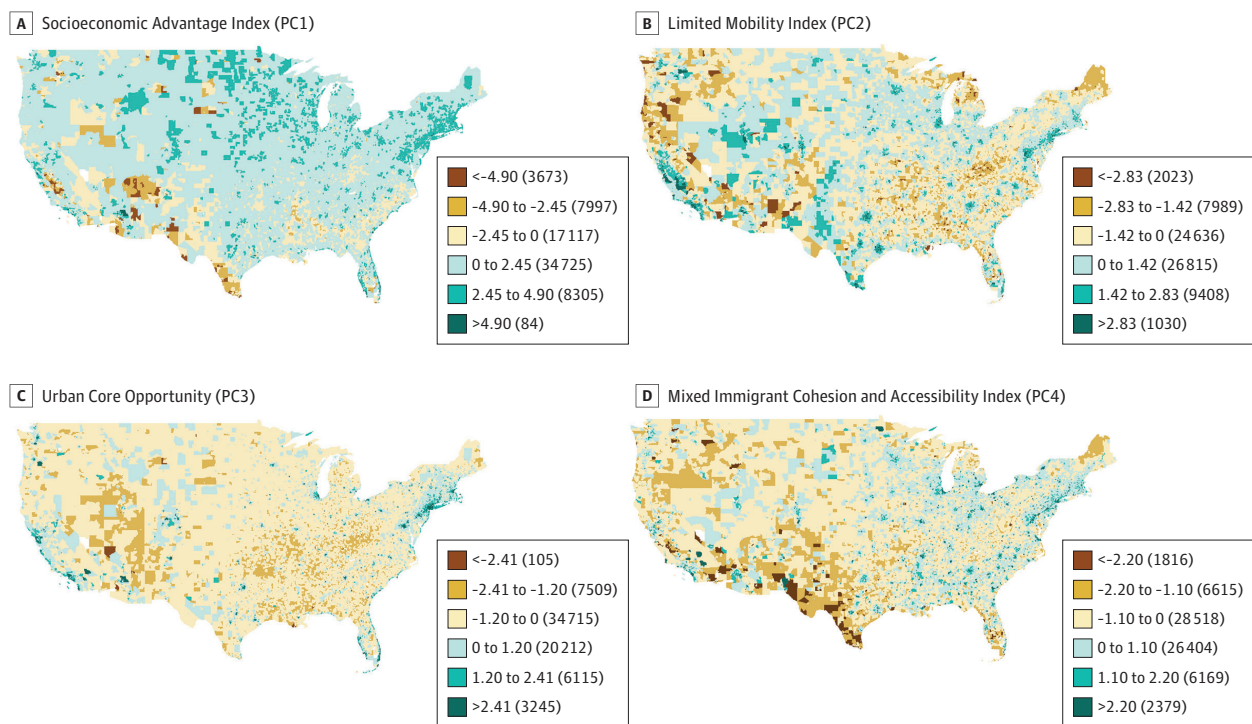
A cumulative SDOH index was calculated by adding all 4 component index scores, weighting each by their proportional variance from the principal component analysis. A cumulative SDOH index mapped across the continental United States (eFigure 3 in the [Supplement](#)) indicates a pattern similar to that of the socioeconomic advantage index, reflecting the dominant weighting of this index. In eFigure 4 in the [Supplement](#), a scatter plot of the cumulative index and premature mortality

rates in Chicago indicates that census tracts with higher mortality rates had lower index scores, reflecting higher vulnerability. To keep the maximum possible data, it is preferable to retain each index separately, although this singular index can be useful for exploratory analysis and comparison purposes.

The k-means analysis identified 7 optimal clusters, each corresponding to unique neighborhood typologies (eFigure 5 in the Supplement). We used both the summary statistics themes and the geographic patterns of the typologies to inform the names of the regional categories. The 7 optimal clusters of SDOH in the study area were categorized as (1) rural affordable, characterized by higher proportions of older adults and persons with disabilities, moderate income levels, and few persons living in crowded housing or without vehicles; (2) suburban affordable, characterized by high vehicle ownership, few renters, and a high proportion of children; (3) suburban affluent, characterized by high income, a high proportion of children, few persons without vehicles, and low poverty; (4) vibrant urban core, characterized by high income, a high proportion of renters, and few children; (5) extreme poverty, characterized by a high proportion of residents with minority status, low income, high poverty, and high unemployment; (6) multilingual working, characterized by a high proportion of residents with minority status, low English proficiency, low income, and low unemployment; and (7) sparse areas, characterized by a high proportion of older adults and generally located within national or state forests, parks, or other natural areas.

The regional typologies were complex. To impose as minimal a construct as possible, we used the minimum number of qualifiers to describe each category. We had not anticipated that these typologies would serve as proxies for different rural, suburban, and urban types, but we found this characteristic helpful for interpretation. However, we caution that these category names are approximations and generalizations. Social determinants of health clusters or typologies are

Figure 2. Social Determinants of Health Indices for Continental United States



A, Socioeconomic advantage index (PC1). B, Limited mobility index (PC2). C, Urban core opportunity index (PC3). D, Mixed immigrant cohesion and accessibility index (PC4). Index scores are provided for each interval, with total number of corresponding census tracts indicated in parentheses.

Data classification corresponds to SDs, with brown shades indicating deviations below the mean and blue shades indicating deviations above the mean. PC1 indicates first primary component; PC2, second primary component; PC3, third primary component; and PC4, fourth primary component.

visualized across the United States in eFigure 4 in the [Supplement](#), with summary statistics shown in Table 1. Results from an analysis of variance highlighted the significance of the cluster attribute value compared with all other tracts. As shown in eTable 3 in the [Supplement](#), each indicator mean for each typology census tract was significantly different from all other census tract means. Index and clustering results can be explored in the form of a table or an interactive map at the SDOH atlas website,<sup>57</sup> with the magnification tool enabled for meaningful exploration. To further facilitate interpretation of these results, we have provided eTable 2 in the [Supplement](#), which includes data regarding the direction and magnitude of indices across each regional typology for the continental United States and Chicago. For example, the socioeconomic advantage index was an order of magnitude lower in the extreme poverty and multilingual working group typologies compared with all other census tracts.

At a macro level, rural affordable areas were present across the country. Suburban affluent areas were clustered in the northeastern area of the United States and scattered across wider regions across the northern United States. Sparse areas with more natural, preserved environments were found in Maine and the southwestern, northwestern, and northern sections of Michigan and Wisconsin. Vibrant urban core areas were not visible at a macro level because they were generally small, densely populated census tracts contained within enclaves of large urban environments, such as New York City, Los Angeles, and San Francisco (SDOH typologies within Chicago are shown in eFigure 5 in the [Supplement](#)). Although clusters of extreme poverty were visible in the southern United States from Mississippi to Alabama, most of these census tracts were likewise smaller and visible when magnified to neighborhood levels.

When considering how typologies vary across space together, the continental United States can be viewed as a complex patchwork of nuanced SDOH characteristics. Some areas were characterized by their tremendous spatial heterogeneity across SDOH neighborhood typologies, such as Florida, which was not obviously dominated by any specific group. Other areas were characterized by their relative spatial homogeneity, such as Oklahoma, which mainly included rural and suburban affordable areas. Notably, all 7 geographic typologies were present in Chicago, which is a spatially heterogeneous city known for its neighborhoods and socioeconomic disparities (Table 1).

### Regression Analysis

We found that in Chicago, more than 60% of the variation in premature mortality at the neighborhood level was associated with SDOH dimensions alone, even after accounting for violent crime and underlying spatial structures. An association was observed between all SDOH indices and age-adjusted premature mortality in Chicago (OLS adjusted  $R^2 = 0.61$  and SAR adjusted  $R^2 = 0.63$ ;  $P < .001$ ). Increased rates of premature mortality were associated with the principal components of less advantage (OLS estimate,  $-0.18$ ;  $P < .001$  and SAR estimate,  $-0.12$ ;  $P < .001$ ), less mobility (OLS estimate,  $-0.49$ ;  $P < .001$  and SAR estimate,  $-0.37$ ;  $P < .001$ ), less opportunity (OLS estimate,  $-0.29$ ;  $P < .001$  and SAR estimate,  $-0.17$ ;  $P < .001$ ), lower scores on the mixed immigrant cohesion and accessibility index (OLS estimate,  $-0.09$ ;  $P = .003$  and SAR estimate,  $-0.07$ ;  $P = .01$ ), and higher crime (OLS estimate,  $0.50$ ;  $P = .005$  and SAR estimate,  $0.34$ ;  $P < .001$ ). Results are shown in **Table 3**, with software output included in eResults in the [Supplement](#). Although census tracts with high immigrant or multilingual populations were associated with better health outcomes in Chicago despite the city's more disadvantaged socioeconomic characteristics, this association was inverted from positive to negative when the limited mobility index was added to the regression analysis. The sensitivity of the mixed immigrant cohesion and accessibility index may have thus reflected a nuanced and complicated underlying pattern. Low accessibility measures may have outweighed the protective, socially cohesive factors characteristic of communities of recent immigrants that mitigated health outcomes (Table 3).

## Discussion

The design, implementation, and evaluation of effective policies at the local, state, and federal levels to improve health outcomes can be improved through a deeper understanding of the complexities of social and economic disparities. In this study, we aimed to decompose and disentangle these complexities using publicly available data, and we found that most variation in data across all census tracts in the continental United States can be quantified by 4 core components that reflect socioeconomic advantage, limited mobility, urban core opportunity, and mixed immigrant cohesion and accessibility. Notably, while socioeconomic disadvantage was the dominant factor in this variation, social and neighborhood environment characteristics together accounted for almost the same amount of variation, underscoring the complexity and nuance of place-based SDOH indicator differences. These multidimensional indices can be further summarized as 7 SDOH clusters or typologies that serve as proxies for urban, suburban, and rural neighborhood groups as well as areas of both extreme wealth and poverty. When considering multiple dimensions of health, the compounding of socioeconomic disadvantage, vulnerable population distribution, low opportunity, and low accessibility may increase disparities.

Areas of extreme poverty not only reflect disadvantage but may also signal wider patterns of community deinvestment and infrastructure breakdown that together suggest consequences for population-level health outcomes. The extreme poverty SDOH neighborhood typology of greatest concern to health care practitioners and policy advocates comprised only 9.6% of all census tracts across the continental United States but characterized small areas of known public health crises, including lead contamination of drinking water in Flint, Michigan,<sup>58,59</sup> lead exposure among children in some parts of Chicago,<sup>60</sup> and parasitic worm outbreaks associated with a lack of sanitation in rural regions of Alabama.<sup>61</sup> These public health crises have had disproportionate consequences in disadvantaged communities and have often resulted from breakdowns in the basic infrastructures necessary for health. Our SDOH neighborhood typology identified and differentiated these areas in a nationwide, standardized manner and thus may be useful for further research and investigation and may serve as an approach to prioritize resources, interventions, and future investments.

Our findings confirm and extend the literature associating SDOH indicators with health outcomes, including policy reviews based on both clinical research and population-based studies, which have hypothesized that 60% of early deaths in the United States are associated with

**Table 3. Results of Regression Analysis<sup>a</sup>**

Measure	OLS Estimate (SE)	P Value	SAR Estimate (SE)	P Value
Advantage <sup>b</sup>	-0.179 (0.033)	<.001	-0.118 (0.031)	<.001
Mobility <sup>c</sup>	-0.490 (0.030)	<.001	-0.370 (0.031)	<.001
Opportunity <sup>d</sup>	-0.289 (0.026)	<.001	-0.168 (0.028)	<.001
MICA <sup>e</sup>	-0.085 (0.030)	.003	-0.070 (0.027)	.01
Violent crime	0.496 (0.108)	.005	0.341 (0.088)	<.001
Spatially lagged YPLL	NA	NA	0.396 (0.150)	<.001
R <sup>2</sup>	0.614	NA	0.643	NA
Adjusted R <sup>2</sup>	0.612	NA	0.627	NA
Mean squared error	0.036	NA	0.032	NA

Abbreviations: Advantage, socioeconomic advantage index; MICA, mixed immigrant cohesion and accessibility index; Mobility, limited mobility index; NA, not applicable; OLS, ordinary least square; Opportunity, urban core opportunity index; SAR, spatial autoregressive model; YPLL, years of potential life lost.

<sup>a</sup> The regression analysis included 789 observations. The SAR model results are shown with maximum likelihood estimation. The generalized method of moments estimation results are available in eResults in the Supplement.

<sup>b</sup> First primary component.

<sup>c</sup> Second primary component.

<sup>d</sup> Third primary component.

<sup>e</sup> Fourth primary component.

nonclinical factors.<sup>2,30</sup> Our approach allowed us to present a framework for understanding the intrinsically complex associations between education, poverty, and health outcomes, and it is well-suited for policy design, as it makes explicit factors of neighborhood environment characteristics that may be important factors in all policies designed to address SDOH. Our Chicago case study findings are consistent with those of Singh et al,<sup>9-11</sup> which indicated that socioeconomic indicators could account for almost half of the total variance in health outcomes.<sup>7-9</sup> Educational level, another important factor in our first component index, has likewise been associated with almost half of the deaths in working-age populations in the United States.<sup>31</sup> Our analysis confirms that the use of intercorrelated indicators of SDOH likely introduces issues of endogeneity in analysis (as hypothesized by Fuchs<sup>8</sup>) and may obfuscate the underlying phenomenon of socioeconomic advantage and overlook the further complexity of neighborhood patterns. The ADI and similar approaches served as proxies for the first principal component (the socioeconomic advantage index) in our analysis but may have missed the additional social and neighborhood environment characteristics that together may be equally important. By using indicators of disadvantage alone, researchers may further underestimate the consequences of additional social and built environment characteristics.

Empirical assessment of the SDOH conceptual model is difficult because the complex causal pathways remain unknown, and challenges exist in identifying which factors are associated with worse health outcomes and in understanding the causes of the causes.<sup>62</sup> Our findings suggest that existing models, which focus on traditional measures of socioeconomic disadvantage (poverty, educational level, and minority status) that are generally associated with each other, may not account for additional population-level patterns that can interact with each other or serve as their own unique indicators. More specifically, we hope that by quantifying the distinct dimensions of SDOH, researchers can better understand why SDOH interacts with health outcomes differently in different places. The interactions between socioeconomic disadvantage and limited mobility areas provide an example, as areas with higher rates of poverty and larger numbers of older adults and persons with minority status are likely more vulnerable than areas with either dimension on its own.

## Limitations

Our study had several limitations. Notably, our analysis was an in-depth empirical exploration of SDOH indicators, their interactions, and their associations with premature mortality within a limited period. Because of the potentially unmeasured factors and unknown associations reported in SDOH literature, the results should be interpreted broadly and with caution. Multiple additional factors could and should be introduced in future work to determine the sensitivity of results, including an exploration of outcomes across different races and ethnicities to better approximate distinct associations. The data we used only allowed for cross-sectional analyses, which may have introduced the risk of missing changes (or the lack thereof) in socioeconomic patterns and associated health outcomes over time. Our analysis was performed at an aggregate level and reflected population means rather than individual phenomena. The documented problems of ecological fallacy<sup>63</sup> and modified areal units<sup>64</sup> suggest that caution should be used when interpreting or extrapolating our findings on an individual level.

In addition, because we wanted to retain patterns in both attribute and geographic data dimensions, more precise measurement of uncertainty across all variables in all census tracts was not straightforward. Regionalization has been used as a tool to reduce uncertainty in census data, as aggregating tracts to regions with similar characteristics may increase sampling sizes and reduce margins of error.<sup>65</sup> We attempted to minimize uncertainty by using tract-level estimates and clustering techniques. However, more measurement tool innovation is needed to account for uncertainty across the methods implemented in this study. Finally, owing to data availability, we limited our analysis to premature mortality as a health outcome. Social determinants of health indicators have different associations with differing health outcomes and should be further investigated at small-area and individual-level resolutions.

## Conclusions

We believe a multidimensional instrumentation for quantifying SDOH provides greater understanding of the ways in which underlying small-area structures are associated with health outcomes. Although socioeconomic advantage may be an important factor in the overall direction of health outcomes, those outcomes remain slightly varied across different populations in different places.<sup>4</sup> A place-based approach is focused on distilling factors associated with later-stage disease outcomes, which are far more costly, substantial, and unique. By focusing on these outcomes and working backward, factors can be identified that may lead to an understanding of why these diseases progress and what we can do to prevent them, a process we call the place-based framework.

Our empirical, exploratory approach sought to better understand and connect concepts across SDOH conceptual frameworks and the socioecological view of health, extending existing research and providing careful attention to the association between variables, geographic data science methods that are sensitive to spatial patterns, and the development of multidimensional indices that are useful for planning and practice. Further research is needed to estimate the consequences of the different dimensions of SDOH on health outcomes across multiple geographic places. Stakeholders can use these results, in conjunction with their own data, to assist in identifying areas of targeted intervention to improve health outcomes. Marginalized populations may experience place-based inequities that are associated with the occurrence of poor health. Given their high consequences on health care spending, these inequities may also be key to implementing substantial improvements in health care quality and value. Interventions targeting high-risk subpopulations may deliver direct benefits and produce large improvements at the local, regional, and national levels.

---

### ARTICLE INFORMATION

**Accepted for Publication:** November 30, 2019.

**Published:** January 29, 2020. doi:[10.1001/jamanetworkopen.2019.19928](https://doi.org/10.1001/jamanetworkopen.2019.19928)

**Open Access:** This is an open access article distributed under the terms of the [CC-BY License](https://creativecommons.org/licenses/by/4.0/). © 2020 Kolak M et al. *JAMA Network Open*.

**Corresponding Author:** Marynia Kolak, PhD, MFA, MS, Center for Spatial Data Science, Searle Chemistry Laboratory, University of Chicago, 5735 S. Ellis Ave., Room 232, Chicago, IL 60637 ([mkolak@uchicago.edu](mailto:mkolak@uchicago.edu)).

**Author Affiliations:** Center for Spatial Data Science, Searle Chemistry Laboratory, University of Chicago, Chicago, Illinois (Kolak, Park); Center for Health Innovation, American Hospital Association, Chicago, Illinois (Bhatt, Padrón, Molefe).

**Author Contributions:** Dr Kolak had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

*Concept and design:* Kolak, Bhatt, Padron.

*Acquisition, analysis, or interpretation of data:* Kolak, Park, Molefe.

*Drafting of the manuscript:* Kolak, Bhatt, Park, Padron.

*Critical revision of the manuscript for important intellectual content:* Kolak, Bhatt, Molefe.

*Statistical analysis:* Kolak, Park, Molefe.

*Administrative, technical, or material support:* Kolak, Bhatt, Padron.

*Supervision:* Kolak.

**Conflict of Interest Disclosures:** None reported.

**Meeting Presentation:** This article was presented at the 2019 meeting of the American Association of Geographers; April 10, 2019; Washington, DC.

### REFERENCES

1. Braveman P, Egerter S, Williams DR. The social determinants of health: coming of age. *Annu Rev Public Health*. 2011;32(1):381-398. doi:[10.1146/annurev-publhealth-031210-101218](https://doi.org/10.1146/annurev-publhealth-031210-101218)
2. McGinnis JM, Williams-Russo P, Knickman JR. The case for more active policy attention to health promotion. *Health Aff (Millwood)*. 2002;21(2):78-93. doi:[10.1377/hlthaff.21.2.78](https://doi.org/10.1377/hlthaff.21.2.78)

3. Murray CJ, Atkinson C, Bhalla K, et al; US Burden of Disease Collaborators. The state of US health, 1990-2010: burden of diseases, injuries, and risk factors. *JAMA*. 2013;310(6):591-608. doi:10.1001/jama.2013.13805
4. Adler NE, Cutler DM, Fielding JE, et al. *Addressing Social Determinants of Health and Health Disparities: A Vital Direction for Health and Health Care. NAM Perspectives Discussion Paper*. Washington, DC: National Academy of Medicine; 2016.
5. Garg A, Boynton-Jarrett R, Dworkin PH. Avoiding the unintended consequences of screening for social determinants of health. *JAMA*. 2016;316(8):813-814. doi:10.1001/jama.2016.9282
6. Office of Disease Prevention and Health Promotion. Healthy people 2020. <https://www.healthypeople.gov/>. Updated December 3, 2019. Accessed June 2019.
7. 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):1750-1766. doi:10.1001/jama.2016.4226
8. Fuchs VR. Social determinants of health: caveats and nuances. *JAMA*. 2017;317(1):25-26. doi:10.1001/jama.2016.17335
9. Singh GK, Siahpush M. Increasing inequalities in all-cause and cardiovascular mortality among US adults aged 25-64 years by area socioeconomic status, 1969-1998. *Int J Epidemiol*. 2002;31(3):600-613. doi:10.1093/ije/31.3.600
10. Singh GK. Area deprivation and widening inequalities in US mortality, 1969-1998. *Am J Public Health*. 2003;93(7):1137-1143. doi:10.2105/AJPH.93.7.1137
11. Singh GK, Azuine RE, Siahpush M, Kogan MD. All-cause and cause-specific mortality among US youth: socioeconomic and rural-urban disparities and international patterns. *J Urban Health*. 2013;90(3):388-405. doi:10.1007/s11524-012-9744-0
12. Kind AJ, Jencks S, Brock J, et al. Neighborhood socioeconomic disadvantage and 30-day rehospitalization: a retrospective cohort study. *Ann Intern Med*. 2014;161(11):765-774. doi:10.7326/M13-2946
13. Danaher A. Reducing disparities and improving population health: the role of a vibrant community sector. <http://www.wellesleyinstitute.com/wp-content/uploads/2011/10/Reducing-Disparities-and-Improving-Population-Health.pdf>. Published August 2011. Accessed June 2019.
14. Solar O, Irwin A. A conceptual framework for action on the social determinants of health. Social determinants of health discussion paper 2 (policy and practice). Geneva, Switzerland: World Health Organization. [https://www.who.int/sdhconference/resources/ConceptualframeworkforactiononSDH\\_eng.pdf](https://www.who.int/sdhconference/resources/ConceptualframeworkforactiononSDH_eng.pdf). Published October 2010. Accessed June 2019.
15. Bay Area Regional Health Inequities Initiative. A public health framework for reducing health inequities. <http://barhii.org/framework>. Published 2015. Accessed June 2019.
16. Hammond G, Joynt Maddox KE. A theoretical framework for clinical implementation of social determinants of health [published online October 9, 2019]. *JAMA Cardiol*. 2019. doi:10.1001/jamacardio.2019.3805
17. Koh HKA. A 2020 vision for healthy people. *N Engl J Med*. 2010;362(18):1653-1656. doi:10.1056/NEJMp1001601
18. Bronfenbrenner U. Ecological models of human development. In: Gauvain M, Cole M, eds. *Readings on the Development of Children*. 2nd ed. Basingstoke, UK: WH Freeman; 1997:37-43.
19. Majer IM, Nusselder WJ, Mackenbach JP, Klijs B, van Baal PH. Mortality risk associated with disability: a population-based record linkage study. *Am J Public Health*. 2011;101(12):e9-e15. doi:10.2105/AJPH.2011.300361
20. Forman-Hoffman VL, Ault KL, Anderson WL, et al. Disability status, mortality, and leading causes of death in the United States community population. *Med Care*. 2015;53(4):346-354. doi:10.1097/MLR.0000000000000321
21. Sudore RL, Yaffe K, Satterfield S, et al. Limited literacy and mortality in the elderly: the health, aging, and body composition study. *J Gen Intern Med*. 2006;21(8):806-812. doi:10.1111/j.1525-1497.2006.00539.x
22. Sentell T, Braun KL. Low health literacy, limited English proficiency, and health status in Asians, Latinos, and other racial/ethnic groups in California. *J Health Commun*. 2012;17(suppl 3):82-99. doi:10.1080/10810730.2012.712621
23. Blakely T, Atkinson J, Kiro C, Blaiklock A, D'Souza A. Child mortality, socioeconomic position, and one-parent families: independent associations and variation by age and cause of death. *Int J Epidemiol*. 2003;32(3):410-418. doi:10.1093/ije/dyg154
24. Amato PR, Patterson SE. Single-parent households and mortality among children and youth. *Soc Sci Res*. 2017;63:253-262. doi:10.1016/j.ssresearch.2016.09.017

25. Simplican SC, Leader G, Kosciulek J, Leahy M. Defining social inclusion of people with intellectual and developmental disabilities: an ecological model of social networks and community participation. *Res Dev Disabil*. 2015;38:18-29. doi:10.1016/j.ridd.2014.10.008
26. Yelin E, Yazdany J, Trupin L. Relationship between poverty and mortality in systemic lupus erythematosus. *Arthritis Care Res (Hoboken)*. 2018;70(7):1101-1106. doi:10.1002/acr.23428
27. Kawachi I, Kennedy BP. The relationship of income inequality to mortality: does the choice of indicator matter? *Soc Sci Med*. 1997;45(7):1121-1127. doi:10.1016/S0277-9536(97)00044-0
28. Gordon SH, Sommers BD. Recessions, poverty, and mortality in the United States: 1993-2012. *Am J Health Econ*. 2016;2(4):489-510. doi:10.1162/AJHE\_a\_00060
29. McLeod CB, Lavis JN, MacNab YC, Hertzman C. Unemployment and mortality: a comparative study of Germany and the United States. *Am J Public Health*. 2012;102(8):1542-1550. doi:10.2105/AJPH.2011.300475
30. Lleras-Muney A. The relationship between education and adult mortality in the United States. *Rev Econ Stud*. 2005;72(1):189-221. doi:10.1111/OO34-6527.00329
31. Hummer RA, Hernandez EM. The effect of educational attainment on adult mortality in the United States. *Popul Bull*. 2013;68(1):1-16.
32. Krueger PM, Tran MK, Hummer RA, Chang VW. Mortality attributable to low levels of education in the United States. *PLoS One*. 2015;10(7):e0131809. doi:10.1371/journal.pone.0131809
33. Franks P, Clancy CM, Gold MR. Health insurance and mortality: evidence from a national cohort. *JAMA*. 1993;270(6):737-741. doi:10.1001/jama.1993.03510060083037
34. Wilper AP, Woolhandler S, Lasser KE, McCormick D, Bor DH, Himmelstein DU. Health insurance and mortality in US adults. *Am J Public Health*. 2009;99(12):2289-2295. doi:10.2105/AJPH.2008.157685
35. Krieger J, Higgins DL. Housing and health: time again for public health action. *Am J Public Health*. 2002;92(5):758-768. doi:10.2105/AJPH.92.5.758
36. Bowen EA, Mitchell CG. Housing as a social determinant of health: exploring the relationship between rent burden and risk behaviors for single room occupancy building residents. *Soc Work Public Health*. 2016;31(5):387-397. doi:10.1080/19371918.2015.1137518
37. Mehdipanah R, Schulz AJ, Israel BA, et al. Neighborhood context, homeownership and home value: an ecological analysis of implications for health. *Int J Environ Res Public Health*. 2017;14(10):e1098. doi:10.3390/ijerph14101098
38. Seng JJB, Kwan YH, Goh H, Thumboo J, Low LL. Public rental housing and its association with mortality: a retrospective, cohort study. *BMC Public Health*. 2018;18(1):665. doi:10.1186/s12889-018-5583-6
39. Syed ST, Gerber BS, Sharp LK. Traveling towards disease: transportation barriers to health care access. *J Community Health*. 2013;38(5):976-993. doi:10.1007/s10900-013-9681-1
40. Rey SJ, Anselin L, Folch DC, Arribas-Bel D, Sastre Gutierrez ML, Interlante L. Measuring spatial dynamics in metropolitan areas. *Econ Dev Q*. 2011;25(1):54-64. doi:10.1177/0891242410383414
41. Spielman SE, Logan JR. Using high-resolution population data to identify neighborhoods and establish their boundaries. *Ann Assoc Am Geogr*. 2013;103(1):67-84. doi:10.1080/00045608.2012.685049
42. Guo D. Regionalization with dynamically constrained agglomerative clustering and partitioning (REDCAP). *Int J Geogr Inf Sci*. 2008;22(7):801-823. doi:10.1080/13658810701674970
43. Association for Computing Machinery. *SODA '07 Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*. Philadelphia, PA: Society for Industrial and Applied Mathematics; 2007.
44. Anselin L. Spatial econometrics. In: Baltagi BH, ed. *A Companion to Theoretical Econometrics*. Malden, MA: Blackwell Publishing; 2003:310-330. <http://doi.wiley.com/10.1002/9780470996249.ch15>. Published November 20, 2007. Accessed December 5, 2018.
45. Landrine H, Corral I. Separate and unequal: residential segregation and black health disparities. *Ethn Dis*. 2009;19(2):179-184.
46. Zenk SN, Powell LM. US secondary schools and food outlets. *Health Place*. 2008;14(2):336-346. doi:10.1016/j.healthplace.2007.08.003
47. Schwanen T, Paez A. The mobility of older people—an introduction. *J Transp Geogr*. 2010;18(5):591-595. doi:10.1016/j.jtrangeo.2010.06.001
48. Cannuscio C, Block J, Kawachi I. Social capital and successful aging: the role of senior housing. *Ann Intern Med*. 2003;139(5 Pt 2):395-399. doi:10.7326/0003-4819-139-5\_Part\_2-200309021-00003
49. Alsnih R, Hensher DA. The mobility and accessibility expectations of seniors in an aging population. *Transp Res Part A Policy Pract*. 2003;37(10):903-916. doi:10.1016/S0965-8564(03)00073-9



50. Kim S. Assessing mobility in an aging society: personal and built environment factors associated with older people's subjective transportation deficiency in the US. *Transp Res, Part F Traffic Psychol Behav*. 2011;14(5):422-429. doi:10.1016/j.trf.2011.04.011
51. Talen E, ed. *A Research Agenda for New Urbanism*. Cheltenham, UK: Edward Elgar Publishing; 2019. doi:10.4337/9781788118637
52. Thorns DC. *The Transformation of Cities: Urban Theory and Urban Life*. London: Macmillan Education UK; 2002. doi:10.1007/978-1-4039-9031-0
53. Penchansky R, Thomas JW. The concept of access: definition and relationship to consumer satisfaction. *Med Care*. 1981;19(2):127-140. doi:10.1097/00005650-198102000-00001
54. Saurman E. Improving access: modifying Penchansky and Thomas's theory of access. *J Health Serv Res Policy*. 2016;21(1):36-39. doi:10.1177/1355819615600001
55. Mulvaney-Day NE, Alegría M, Sribney W. Social cohesion, social support, and health among Latinos in the United States. *Soc Sci Med*. 2007;64(2):477-495. doi:10.1016/j.socscimed.2006.08.030
56. Dillon FR, De La Rosa M, Ibañez GE. Acculturative stress and diminishing family cohesion among recent Latino immigrants. *J Immigr Minor Health*. 2013;15(3):484-491. doi:10.1007/s10903-012-9678-3
57. Kolak M, Bhatt J, Park YH, Padron NA, Molefe A. SDOH atlas. <https://sdohatlas.github.io/>. Published 2019. Accessed January 3, 2020.
58. Hanna-Attisha M, LaChance J, Sadler RC, Champney Schnepf A. Elevated blood lead levels in children associated with the Flint drinking water crisis: a spatial analysis of risk and public health response. *Am J Public Health*. 2016;106(2):283-290. doi:10.2105/AJPH.2015.303003
59. Bellinger DC. Lead contamination in Flint—an abject failure to protect public health. *N Engl J Med*. 2016;374(12):1101-1103. doi:10.1056/NEJMp1601013
60. Muller C, Sampson RJ, Winter AS. Environmental inequality: the social causes and consequences of lead exposure. *Annu Rev Sociol*. 2018;44(1):263-282. doi:10.1146/annurev-soc-073117-041222
61. McKenna ML, McAtee S, Bryan PE, et al. Human intestinal parasite burden and poor sanitation in rural Alabama. *Am J Trop Med Hyg*. 2017;97(5):1623-1628. doi:10.4269/ajtmh.17-0396
62. Braveman P, Gottlieb L. The social determinants of health: it's time to consider the causes of the causes. *Public Health Rep*. 2014;129(suppl 2):19-31. doi:10.1177/003335491412915206
63. Piantadosi S, Byar DP, Green SB. The ecological fallacy. *Am J Epidemiol*. 1988;127(5):893-904. doi:10.1093/oxfordjournals.aje.a114892
64. Wong D. The modifiable areal unit problem (MAUP). In: Fotheringham AS, Rogerson PA, eds. *The SAGE Handbook of Spatial Analysis*. London: SAGE Publications; 2009:105-124. doi:10.4135/9780857020130.n7
65. Spielman SE, Folch DC. Reducing uncertainty in the American community survey through data-driven regionalization. *PLoS One*. 2015;10(2):e0115626. doi:10.1371/journal.pone.0115626

#### SUPPLEMENT.

**eMethods.** Spatial Scale, Data and Software, YPLL Data Definition, and Regional Typologies

**eResults.** OLS and SAR Estimations

**eFigure 1.** Raw Distributions of Principal Component Scores

**eFigure 2.** SDOH Indices for Chicago

**eFigure 3.** Cumulative SDOH Index Across Continental United States

**eFigure 4.** Association Between Cumulative SDOH Factors and Years of Potential Life Lost at Tract Level for Chicago

**eFigure 5.** SDOH Neighborhood Typology

**eTable 1.** Principal Component Analysis Overview

**eTable 2.** SDOH Indices, US Typology, and Chicago Typology

**eTable 3.** Average and Standard Deviation for Each Variable of Interest Across All US Tracts

**eReferences.**