DOI: 10.1111/1475-6773.14102

RESEARCH ARTICLE

HSR Health Services Research

The social vulnerability metric (SVM) as a new tool for public health

Loren Saulsberry PhD[1] | Ankur Bhargava MD, MPH[2] | Sharon Zeng BA[3] | Jason B. Gibbons PhD[4] | Cody Brannan MS[5] | Diane S. Lauderdale PhD[1] | Robert D. Gibbons PhD[1,5,6]

[1]Department of Public Health Sciences, The University of Chicago, Chicago, Illinois, USA

[2]Department of Pediatrics, The University of Chicago, Chicago, Illinois, USA

[3]Pritzker School of Medicine, The University of Chicago, Chicago, Illinois, USA

[4]Department of Health Policy and Management, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA

[5]Center for Health Statistics, The University of Chicago, Chicago, Illinois, USA

[6]Department of Medicine, The University of Chicago, Chicago, Illinois, USA

Correspondence

Loren Saulsberry, Department of Public Health Sciences, The University of Chicago, 5841 S. Maryland Ave., Chicago, IL 60637, USA.
Email: lsaulsberry@uchicago.edu

Funding information

National Human Genome Research Institute, Grant/Award Number: K08 HG011505; National Institute on Aging, Grant/Award Number: R56 AG066127

Abstract

Objective: To derive and validate a new ecological measure of the social determinants of health (SDoH), calculable at the zip code or county level.

Data Sources and Study Setting: The most recent releases of secondary, publicly available data were collected from national U.S. health agencies as well as state and city public health departments.

Study Design: The Social Vulnerability Metric (SVM) was constructed from U.S. zip-code level measures (2018) from survey data using multidimensional Item Response Theory and validated using outcomes including all-cause mortality (2016), COVID-19 vaccination (2021), and emergency department visits for asthma (2018). The SVM was also compared with the existing Centers for Disease Control and Prevention's Social Vulnerability Index (SVI) to determine convergent validity and differential predictive validity.

Data Collection/Extraction Methods: The data were collected directly from published files available to the public online from national U.S. health agencies as well as state and city public health departments.

Principal Findings: The correlation between SVM scores and national age-adjusted county all-cause mortality was $r = 0.68$. This correlation demonstrated the SVM's robust validity and outperformed the SVI with an almost four-fold increase in explained variance (46% vs. 12%). The SVM was also highly correlated ($r \geq 0.60$) to zip-code level health outcomes for the state of California and city of Chicago.

Conclusions: The SVM offers a measurement tool improving upon the performance of existing SDoH composite measures and has broad applicability to public health that may help in directing future policies and interventions. The SVM provides a single measure of SDoH that better quantifies associations with health outcomes.

KEYWORDS

biostatistical methods, determinants of health/population health/socioeconomic causes of health, health care disparities, health equity, health policy, social determinants of health

What is known on this topic

- Social determinants of health (SDoH) impact people's health and well-being.
- SDoH can contribute to health disparities and inequities.
- Valid measurement of SDoH can help accurately target interventions for communities facing the greatest social vulnerability.

What this study adds

- Introduces the Social Vulnerability Metric (SVM) as a new measure of social vulnerability.
- The SVM was derived from SDoH variables from multiple nationally representative public databases using multidimensional Item Response Theory.
- The SVM provides a higher level of precision than existing SDoH metrics in estimating a geographic index of social vulnerability.

1 | INTRODUCTION

1.1 | Importance of social determinants of health

Social determinants of health (SDoH) include the multifaceted set of social, economic, and demographic conditions that impact people's health and well-being.[1] These conditions include, for example, place of residence, education, economic stability, work opportunities, leisure activities, social interactions, health care access, housing, transportation, racism, language and literacy, and access to healthy foods and community resources.[2] At least one study has found that 80%–90% of the modifiable contributors to population health outcomes are related to socioeconomics, health-related behaviors, and environmental factors.[3] SDoH are shaped by the distribution of power and resources across multiple levels (global, national, and local) and contribute to health disparities and inequities.[4,5] Studies that measure the association between SDoH and health and wellbeing have shown that medical care alone cannot eliminate health disparities. Limited access to resources, whether due solely to economic factors or combinations of factors, is associated with adverse health outcomes. Resources that increase the quality of life can significantly improve population health outcomes.[6] Examples include access to safe and affordable housing, educational opportunity, public safety, healthy foods, and emergency health services.

1.2 | Previous composite measures from publicly available data

Valid measurement of SDoH can help policy makers accurately target their interventions and programs to communities facing the greatest social vulnerability and can help researchers better evaluate not only the role of SDoH but other risk factors in etiologic research. The Centers for Disease Control and Prevention (CDC) Social Vulnerability Index (SVI) is a widely used approach to measuring SDoH. The CDC's SVI is a composite measure that uses US Census data for 15 social factor variables to aid local officials in identifying communities at risk for the potential negative effects of external stresses on human health, such as natural/human-caused disasters or disease outbreaks.[7] The SVI relies on census tract level data to construct overall community rankings based on variables organized according to four themes: (1) "Socioeconomic Status" including percentages below poverty, unemployed, income, no high school diploma; (2) "Household Composition & Disability" including proportions aged 65 or older, aged 17 or younger, civilian with a disability, single-parent households; (3) "Minority Status & Language" including proportions minority and speaks English "less than well," and (4) "Housing Type & Transportation" including proportions multi-unit structures, mobile homes, crowding, no vehicle, and group quarters. As described by Flanagan and colleagues (2011), in order to construct the SVI, each of these variables was ranked from highest to lowest across all U.S. census tracts with the exception of per capita income.[8] Alternatively, per capita income was ranked from lowest to highest as higher values reflected less vulnerability. Then, a percentile rank was calculated over each of these variables for every census tract. A percentile rank at the tract-level was calculated for each of the four domains based on a sum of the percentile ranks of the variables that made up each domain. Finally, a composite percentile rank for each tract was determined by summation of all the domain percentile rankings. The process of percentile ranking described above was then repeated for individual states. While the SVI was originally designed to assist public health officials and emergency response planners identify and target communities that will most likely need support before, during, and after hazardous events, use of the SVI has become more widespread as researchers have started to apply the SVI to better understand the associations between SDoH factors and other salient public health outcomes like obesity.[9]

1.3 | A new methodological approach

This study introduces a new measure of social vulnerability, the Social Vulnerability Metric (SVM). The SVM was derived from a large set of SDoH variables from multiple nationally representative public use administrative databases, and the SVM was constructed using multidimensional Item Response Theory (MIRT), which is a statistical model of measurement which provides a higher level of precision than existing SDoH metrics in estimating a geographic index of social vulnerability, across the entire range of social vulnerability. It does this by estimating parameters related to each of the SDoH variables which (1) describe the strength of association between the specific variable or item, and the primary SDoH latent variable, which we call the SVM, (2) the relationship between each item and it's underlying subdomain (demographic, education, economic, physical infrastructure, and health care), and (3) the degree of vulnerability assessed by each item. This contrasts with the usual approach taken in constructing the CDC's SVI and related measures such as the Area Deprivation Index (ADI),[10] which give equal weight to all variables (items) regardless of whether they are actually related to the underlying social vulnerability variable, or the degree of vulnerability associated with the item. We expect that the MIRT approach will yield higher precision of measurement and as a consequence greater association with health outcomes than the traditional measures.

First, we describe the development of the SVM. Second, we compare the SVM and the SVI at the county level to each other, and then compare their predictive accuracy in terms of associations with all-cause age-adjusted mortality rates. Third, we examine the associations between the SVM and two additional independent public health data sources with zip-code level data: California COVID-19 vaccination rates and age-adjusted emergency department visits for asthma, and Chicago COVID-19 mortality and vaccination rates.

2 | METHODS

2.1 | Derivation of the SVM

To develop the SVM, we used the most recent data from 2018 in the federal, publicly available Agency for Healthcare Research and Quality (AHRQ) SDoH Database.[11] The AHRQ SDoH Database is a dataset of SDoH variables from data sources spanning 2009 to 2018, including over 200 SDoH variables derived from 17 publicly available data sources (e.g., American Community Survey). The data are available at the level of the Zip Code Tabulation Area (ZCTA). There are approximately 42,000 zip codes and 33,000 ZCTAs in the US. They differ because ZCTAs reflect only populated areas. For example, zip codes that only include PO boxes are not included in ZCTAs. Future releases of the AHRQ dataset, will include tabulations by county, ZCTA, zip code, and census tract block groups for 2020, and we will provide SVM scores for each of them based on our calibrated model described here. Full details of these 17 data sources and the methodology for their inclusion into the AHRQ SDoH Database is described in the AHRQ Data Source.[11] The large item bank in the AHRQ SDoH Database can be linked to other data sources by geography at both the county and zip-code levels. For this analysis, we selected variables corresponding to five key SDoH domains: (1) demographic (e.g., age and race/ethnicity), (2) education, (3) economic context (e.g., unemployment rate), (4) physical infrastructure (e.g., housing and transportation), and (5) health care (e.g., health insurance coverage). Following review, we reduced the 200+ SDoH item bank to 94 items that were relevant to these five key domains and provided quantifiable measures (e.g., percentage) of zip-code-level community characteristics (e.g., percentage with bachelor's degree, living below poverty level, disabled, vacant homes, etc.).

2.2 | Constructing the model for the SVM

We fit a bifactor MIRT model to our reviewed set of 94 variables in order to provide a single-value SVM score. Item response theory models attempt to explain the relationship between unobservable characteristics, or latent traits, and observed characteristics or "items." Rather than accounting for a single latent construct, MIRT permits evaluating multiple constructs relevant to the outcome(s). To accommodate the multidimensionality of SDoH, we used a full-information item bifactor model.[12–14] The bifactor model was the first confirmatory item factor analysis model, originally introduced for measurement data by Holzinger and Swineford (1937)[15] and adapted to item response theory (IRT) by Gibbons and Hedeker (1992).[12] To construct the SVM, we used a previously validated modeling approach that extends the bifactor model to the case of polytomous data with more than two distinct categories.[13] In contrast to the traditional unrestricted item factor analysis model where all items load on all dimensions, the bifactor model imposes restrictions on the traditional item factor analysis model of Bock and Aitkin (1981)[16] by requiring that each item loads on a primary dimension (e.g., SDoH) and only one subdomain (e.g., physical infrastructure). The subdomains used alongside the primary dimension must be prespecified, and this restriction on the item factor analysis model reduces the dimensionality of the integration problem to two, regardless of the number of subdomains, which is a major computational advantage, while at the same time, preserving the multidimensionality of the construct. In the bifactor model, the correlation between items is based on the items' correlation with all other items through the primary dimension and residual correlation among the items from that specific subdomain. Technical details of the bifactor model are provided in the Appendix S1.

We purposely did not include race/ethnicity as items in the derivation of the SVM. In this way, valid comparisons of SVM scores are possible between racial and ethnic groups, as are provided in the following, where we compared SVM distributions between communities with varying levels of White, Black, and Hispanic population proportions (Figure S1).

2.3 | Model calibration

Initially, the bifactor model was applied to data including 94 SDoH variables representative of 33,120 ZCTA codes. While the bifactor model can accommodate missing data, under the very general assumption of missing at random,[17] we decided to eliminate ZCTAs that were missing 44 or more items (i.e., 50 or more items are reported), to ensure that we could get precise SVM scores for all ZCTAs used in our analyses. This led to removal of approximately 1% of the ZCTAs. Items with loadings less than 0.4 on the primary dimension were removed to arrive at a final variable set consisting of 24 items described in Table 1 that were used for the final model calibration. The 0.4 threshold has previously been used in bifactor modeling.[18] The loadings are bounded by −1 and 1. The 24 items formed a bifactor pattern with high loadings on the primary dimension (>0.40). Items with positive loadings are associated with increased vulnerability, and items with negative loadings are associated with decreased vulnerability. A likelihood ratio chi-square statistic was used to assess improvement in fit of the bifactor model over a simple unidimensional IRT alternative.

2.4 | SVM score

In terms of scoring, interest is typically centered on the primary dimension designed to integrate all the subdomain information and preserve multidimensionality. An advantage of the bifactor model over both traditional factor analysis and unrestricted item factor analysis is that the bifactor loadings are rotationally invariant, greatly simplifying interpretability of the model estimates. Unlike traditional

TABLE 1 Bifactor model factor loading estimates for the Social Vulnerability Metric

Item	Primary	Social	Economic	Education	Infrastructure	Health Care
Percentage of civilian veterans with a disability (ages 18–64)	0.67	0.04				
Percentage of families with children that are single-parent families	0.52	−0.27				
Percentage of population divorced or separated (ages 15 and over)	0.49	−0.23				
Percentage of children living with a grandparent householder (ages 17 and under)	0.41	−0.14				
Percentage of households with any internet connection	−0.82	−0.46				
Percentage of households without a computer	0.77	0.53				
Percentage of households with a smartphone with no other type of computing device	0.6	−0.09				
Percentage of employed working in finance and insurance, real estate, and rental and leasing	−0.46		0.38			
Percentage of employed working in professional, scientific, management, administrative, and waste management services	−0.44		0.61			
Percentage of population with income to poverty ratio: 1.25–1.99	0.78		0.29			
Median household income (in dollars, inflation-adjusted to file data year)	−0.91		−0.06			
Percentage of population with a bachelor's degree (ages 25 and over)	−0.75			0.48		
Percentage of population with a master's or professional school degree or doctorate (ages 25 and over)	−0.64			0.48		
Percentage of population with only high school diploma (ages 25 and over)	0.54			−0.64		
Percentage of population with less than high school education (ages 25 and over)	0.75			0		
Median home value of owner-occupied housing units	−0.76				−0.1	
Percentage of housing units that are mobile homes	0.56				0.31	
Percentage of housing units vacant	0.45				0.46	
Convenience stores per 1000 people	0.49				0.7	
Percentage of population with any Medicaid/means-tested public health insurance coverage	0.74					0.52
Percentage of population with any private health insurance coverage	−0.83					−0.43
Percentage of population with employer-based health insurance	−0.75					−0.33
Percentage of population with Medicare, Medicaid, TRICARE/military, U.S. Department of Veterans Affairs (VA) coverage only	0.76					0.64
Percentage of population with no health insurance coverage	0.55					−0.03

Note: Agency for Healthcare Research and Quality Social Determinants of Health 2018 (24 measures). Loadings = correlations (−1 to 1), manifest variables, and the latent variable.

factor analysis of measurement data, the bifactor IRT model also includes *k-1* threshold parameters where *k* is the number of categories for the manifest item responses. This relaxes the assumption of an interval scale of measurement such that every additional unit/interval increase of the SVM score does not have to reflect a uniform magnitude of increased/decreased social vulnerability. While many of the SDoH variables are measured on a continuous scale, we transformed them to quintiles for analysis and scoring, similar to previous methods applying the bifactor model to a mix of clinical and biological variables.[19] The scores for the primary dimension of SDoH are Bayes expected a posteriori (EAP) estimates that are expressed on an underlying unit normal (N(0,1)) scale and can be transformed to percentiles using an inverse normal transformation. We computed the empirical reliability of the test as the ratio of the empirical variance of the test (expected value = 1.0, true score variance) to the total variance, or the empirical variance plus the posterior variance of the estimator (error variance). Empirical reliability more than 0.9 is considered excellent, providing a high level of confidence that the domain intended is actually being measured.[14] In the final SVM scale, higher scores represent increased social vulnerability.

2.5 | SVM validation

To validate the SVM and assess its performance, we conducted four analyses. First, we tested for convergent validity against the CDC's SVI, which is reported at the county-level. To permit the comparison, we aggregated the ZCTA SVM estimates to the county level using the United States Census Bureau Relationship Files ZCTA-County ND ZCTA-MSA crosswalks based on the 2010 census.[20] Aggregation was done by taking a population-weighted average of SVM scores across all ZCTAs assigned to a particular county. ZCTA population estimates were also obtained from US Census data.

Second, we obtained all-cause age-adjusted mortality data from the CDC Wide-ranging Online Data for Epidemiologic Research database for 2016 at the county level. Population weighted correlations between the SVM and the mortality data at the county level were computed, so that very small counties with higher uncertainty in the mortality rates and SVM scores would not unduly influence the overall correlation. We also repeated this analysis with the CDC's SVI[7] and compared the strength of association between the SVM and SVI with all-cause age-adjusted mortality. The SVI produces an overall ranking that is expressed as a percentile for each US county. For comparability, we transformed the SVM to a percentile. While it would also be of interest to compare the SVM to the ADI, the ADI is only available at the census tract block group, and the authors do not recommend aggregation which would permit comparison to our various health outcomes. To determine if our results for the total US population would be preserved in geographic areas with higher proportions of minority populations, we repeated the analysis in those counties with 25% or more Black residents.

Third, we obtained data from the Chicago Department of Public Health on COVID-19 mortality and first vaccination rates through June 19, 2021 (end of the alpha wave in Chicago). The data represent 58 zip codes in the Chicagoland area. COVID-19 mortality was expressed per 100,000 residents. Vaccination rates were cumulative percentages of residents with at least one vaccine. The associations between the SVM scores and COVID-19 mortality and vaccination rates were also assessed using population weighted correlations.

Fourth, we obtained state-wide data from the California Health and Human Services Agency (CHHS) on COVID-19 vaccination rates and emergency department visits for asthma. The data represent 1512 zip codes across the state of California (USA). Vaccination rates were cumulative percentages of residents with at least one vaccine or who completed the full course of vaccination through June 19, 2021, and emergency department visits for asthma were from 2018. The associations between the SVM scores and asthma emergency department visits and vaccination rates were assessed using population weighted correlations.

This study was reviewed by The University of Chicago Institutional Review Board and was determined to be exempt as it relied on publicly available data.

3 | RESULTS

The bifactor model significantly improved the fit over the unidimensional alternative (chi-square = 6455.87, df = 24, $p < 0.0001$). There was high empirical reliability of 0.93. The SVM demonstrated convergent validity against the existing CDC's SVI with a correlation between the SVM and the SVI of 0.68 (weighted by population).

3.1 | All-cause age-adjusted mortality

The SVM demonstrated predictive validity with a correlation 0.68 (weighted by population) with all-cause age-adjusted mortality. Rates ranged from 550/100,000 for the most socially advantaged counties to 1050/100,000 for the most socially vulnerable counties as represented by the SVM (Figure 1). The weighted correlation for mortality rates with the CDC's SVI score was 0.34. The SVM accounts for 46% ($r^2 = 0.46$) of the variability in all-cause age-adjusted mortality, whereas the CDC's SVI accounts for 12% ($r^2 = 0.12$) (Figure 1). The high correlation between the SVM and all-cause age adjusted mortality persisted when restricted to the 426 counties with 25% or more Black residents ($r = 0.64$).

3.2 | State of California COVID-19 vaccination rates and emergency department (ED) visits for asthma

In state-level data from California (USA), SVM correlations across the 1512 zip-codes with COVID-19 vaccination data (out of 1741 zip-codes total) were $r = -0.68$ for one or more COVID-19 vaccinations and $r = -0.70$ for completion of full vaccination (Figure 2, Panel A).

Across the social vulnerability gradient, vaccination rates ranged from just below 80% for those estimated with the SVM to be exposed to the least social vulnerability to slightly above 30% for those estimated to be exposed to the most social vulnerability. For age-adjusted ED visits for asthma, SVM correlations were r = 0.62 for California residents 0–17 years old and r = 0.60 for adults in California aged 18 or more years (Figure 2, Panel B). For the younger age group (0–17 years), ED visit rate ranged from about 30/10,000 for California residents of the most socially advantaged communities to more than 120/10,000 for California residents of the most socially vulnerable communities as represented with the SVM. For the older age group (18+ years), ED visit rate ranged from about 10/10,000 for California residents of the most socially advantaged communities to approximately 70/10,000 for California residents of the most socially vulnerable communities as represented with the SVM.

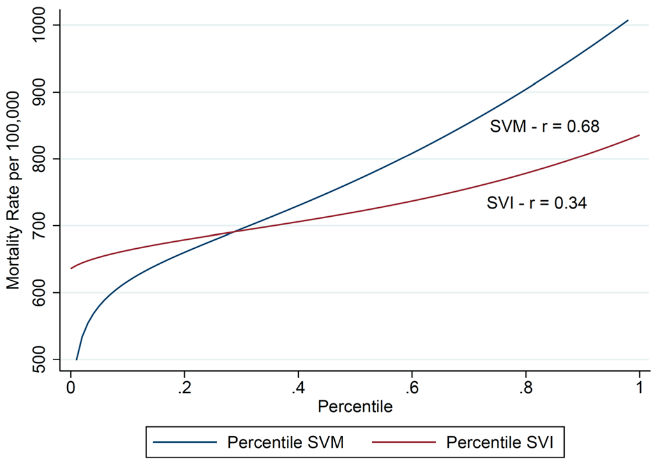

FIGURE 1 Social Vulnerability Metric (SVM) and the Center for Disease Control and Prevention Social Vulnerability Index correlations with all-cause age-adjusted mortality. r, correlation coefficient; SVI, Social Vulnerability Index; SVM, Social Vulnerability Metric. Centers for Disease Control and Prevention Wide-ranging Online Data for Epidemiologic Research data all-cause age-adjusted mortality. [Color figure can be viewed at wileyonlinelibrary.com]

3.3 | City of Chicago COVID-19 mortality and first vaccination rates

With respect to COVID-19 mortality and vaccination rates in Chicago, the population weighted SVM correlations across the 58 zip codes were r = 0.71 for COVID mortality and r = −0.85 for first vaccination rates (Figure 3). First vaccination rates ranged from 80% for those estimated with the SVM to be exposed to the least social vulnerability to 30% for those estimated to be exposed to the most social vulnerability. Conversely, COVID mortality rates varied from 50 per 100,000 for those estimated with the SVM to be exposed to the least social vulnerability to 300 per 100,000 for those estimated to be exposed to the most social vulnerability.

3.4 | Race and ethnicity

Finally, we examined the relationship between SVM and race/ethnicity. Figure S1 indicates a higher likelihood of experiencing social vulnerability for non-White populations, particularly African Americans

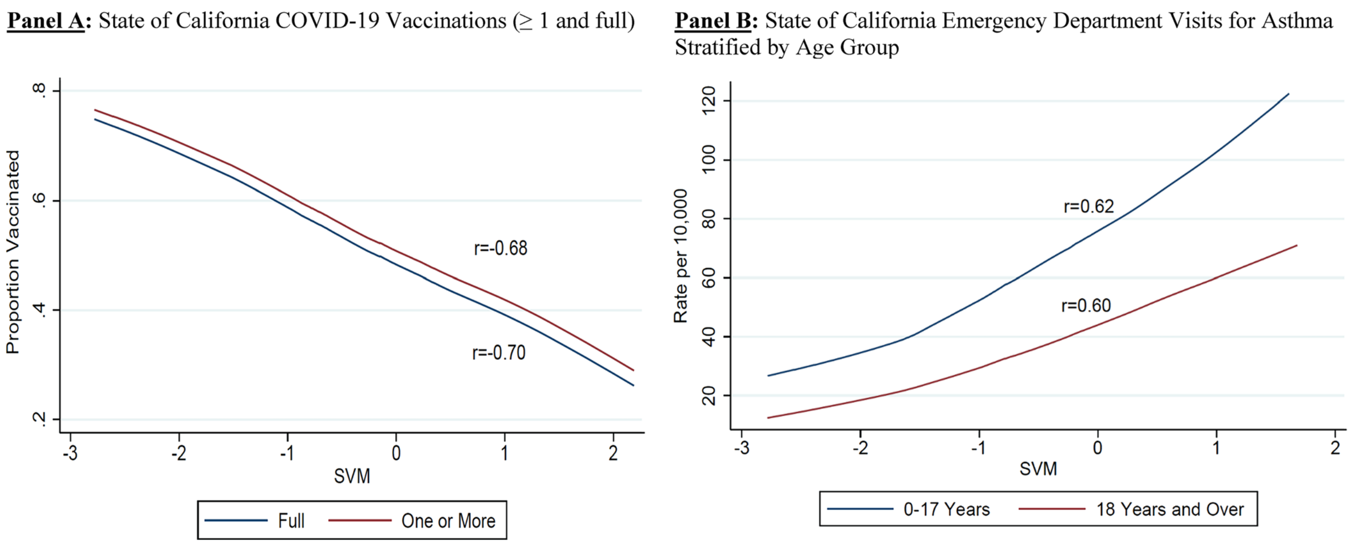

FIGURE 2 SVM versus COVID-19 vaccinations (Panel A) and age-adjusted emergency department visits for asthma (Panel B) in the state of California (USA). Lowess plots of the proportion of ≥1 and full COVID vaccination (Panel A) and the rate of age-adjusted emergency department visits for asthma in the state of California (Panel B). Data analyzed on COVID vaccinations extended through June 19, 2021. Data from 2018 were analyzed for emergency department visits for asthma. Panel B represents the visit rate per 10,000 individuals for age groups 0–17 years and adults 18+ years. r, correlation coefficient; SVM, Social Vulnerability Metric. Authors' analysis of zip-code level data from the California Health and Human Services on ≥1 and full COVID vaccinations and age-adjusted emergency department visits for asthma. [Color figure can be viewed at wileyonlinelibrary.com]

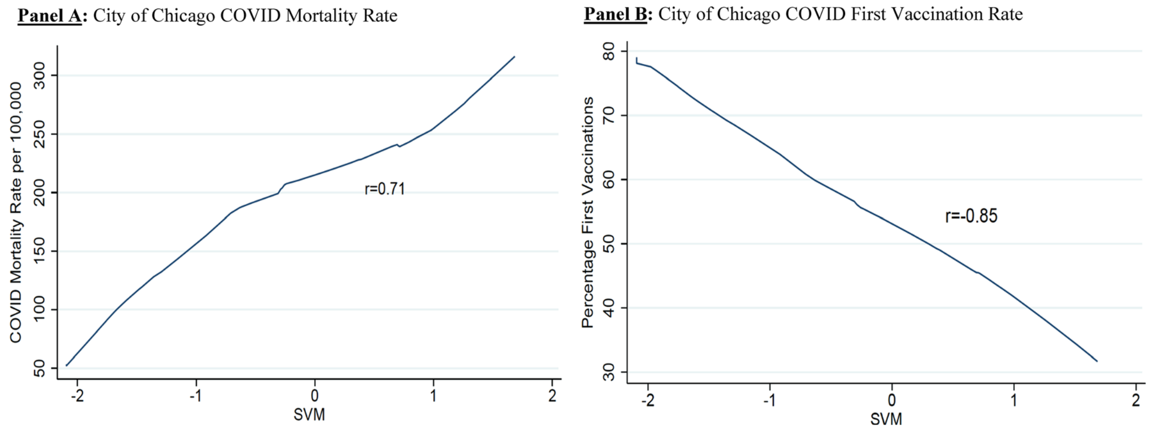

FIGURE 3 SVM versus COVID mortality rate (Panel A) and COVID first vaccination rate (Panel B) in the city of Chicago. Lowess plots of COVID mortality (Panel A) and proportion of COVID first vaccinations (Panel B). r, correlation coefficient, SVM, Social Vulnerability Metric. Authors' analysis of zip-code level data from the Chicago Department of Health on COVID mortality and vaccine initiation through June 19, 2021. Data are reflective of 58 Chicago, Illinois (USA) zip codes. [Color figure can be viewed at wileyonlinelibrary.com]

in the United States. Across the range of lower SVM scores, below zero, the average representation of African Americans is approximately 5%. However, the representation of African Americans increases from approximately 10% at a SVM score of zero to 60% for the highest SVM scores. (Figure S1).

4 | DISCUSSION

We present a new approach to estimating a composite SDoH metric for geographic areas using MIRT, which preserves the multidimensionality of the SDoH construct. We validated the SVM within three separate datasets collected by distinct national, state, and city public health agencies representing different levels of geographic scale. The SVM was strongly related ($r \geq 0.60$ in absolute value) to all outcomes assessed displaying its broad applicability to real-world public health data. At a national level, for all-cause age-adjusted mortality, we observed an almost two-fold increase in the mortality rate of 550/100,000–1020/100,000 across the SVM gradient. At a state level, in California, vaccination rates were almost three times higher in residents of the 10% least vulnerable areas relative to the 10% most vulnerable communities. Finally, at the city level, in the Chicagoland area, more socially vulnerable communities experienced six times the rate of COVID-19 mortality compared to less socially vulnerable communities (300/100,000 vs. 50/100,000 respectively) and a much lower first vaccination rate (32% vs. 79%, respectively). Although the SVM and SVI were strongly correlated ($r = 0.68$), the SVM outperformed the SVI in terms of predicting age-adjusted all-cause mortality, with an almost four-fold increase in explained variance (46% vs. 12%). With an individual's zip code or ZCTA, the SVM provides an estimate of SDoH affiliated with their local environment that can be determined as a single score and/or percentile.

There are of course limitations of the current study. First, the SVM items are based on national survey data from 2018 and as such will need to be updated as new data become available. Second, aggregation from zip code to county is complicated by borders that are not always aligned between zip codes, census tracts, and counties. The granularity of the geographic data analyzed will involve tradeoffs that may improve the estimate of an individual's environmental social vulnerability, or alternatively introduce unwarranted variability. For example, Cook County in Chicago represents a wide range of communities that have extremely heterogeneous levels of social vulnerability as expressed by wide variation in SVM scores. A county-level SVM score simply averages over this variability with corresponding loss of information and precision. New data from AHRQ will soon be released for 2020 and will directly include census tract block group, zip code, ZCTA and county level data, for which we will compute SVM scores and percentiles. These data will be made freely available at http://socialvulnerabilitymetric.com/. [Correction added on 7 December 2022, after first online publication: the URL 'www.adaptivetestingtechnologies.com' in the preceding sentence has been changed to 'http://socialvulnerabilitymetric.com/'.] Third, race and ethnicity are not a part of the SVM score. However, this is also a strength of the SVM when assessing SDoH. Given the high level of collinearity between health-related social needs and race/ethnicity, the SVM was developed to be a metric that could focus specifically on SDoH independently of race/ethnicity. There were racial and ethnic differences in SVM score distributions. The most socially vulnerable communities had a much higher percentage of African American residents and to a lesser extent Hispanic residents, whereas the least socially vulnerable communities were predominantly White. Future research is needed to further delineate the relationships between race/ethnicity and SDoH and their impact on health. Finally, a ZCTA SDoH measure is representative of a complex range of factors impacting both the individual and the larger community context. A person living in a socially vulnerable community, may not experience the negative effects of social vulnerability because of their own resources, other support systems, resilience, or other social and cultural factors. The community-level SVM represents both a community context and a noisy measure of SDoH for the individuals in the community.

In summary, we have developed a model-based measurement system for small area social vulnerability estimation, that allows us to precisely measure SDoH at the ZCTA level. Using similar features as extant SDoH measures (e.g., CDC's SVI), we are able to boost the strength of association with health outcomes, notably age-adjusted all-cause mortality by a factor of four in terms of explained variation. Although we began with a much larger item bank than the SVI, 200 items distilled to 94 items based on review, and then further distilled to 24 items based on bifactor model parameter estimates, ultimately there is a similar number of items (24 vs. 17) between the SVM and SVI. As such, it is not the number of items that drives the increased predictive accuracy, but rather the model-based measurement that improves precision and predictive accuracy. A further advantage of model-based measurement using MIRT, is that as new SDoH variables emerge, they can be calibrated with the existing item bank, and used to generate SVM scores and percentiles that can be interpreted in the same metric. The same is not true for the traditional approach to scoring based on classical test theory.

5 | POLICY AND PRACTICE IMPLICATIONS

Our SVM explains more variation in health outcomes than CDC's widely used SVI. Within public health policy, there is a recognized need for new methodological approaches to the measurement of SDoH.[21,22] Optimizing SDoH measurement for accuracy and the capability to reflect the real-world contexts in which individuals live is of extreme importance to: (1) understanding the extent to which SDoH are associated with health and health care, (2) developing programs to address disparities, and (3) improving adjustment for the SDoH when carrying out research examining other contextual or individual factors. Public health organizations (at local, state, and national levels) along with health systems delivering health care across the country are grappling with ways of assessing and intervening on SDoH to improve health outcomes. In the public sector, investments to generate and expand datasets such as the AHRQ SDoH Database[11] are ongoing, and several large commercial health insurers have invested in predictive analytics to expand their abilities to address SDoH.[23]

A systematic review of the literature describing the construction of geographic indices of socioeconomic disadvantage highlighted that the United States has lagged behind European nations where government agencies have leveraged health data to target high-needs populations for decades.[24] There is no consensus about how to best do this in the United States. An additional challenge has been that the metrics developed to date rely on different geographic scales. Such metrics are tied to the geographic scale inherent to their original design, which was influenced by data availability, quality, and the specific motivating application. In contrast, we designed the SVM to accommodate various geographic scales and produce reliable estimates across them. Therefore, the SVM is poised to evolve alongside emerging data sources and insights about what are the most appropriate levels for targeted health interventions.

To date, SDoH have not been integrated into health regulatory frameworks to improve health and tackle inequity.[25] A cultural shift has begun where sectors across the health care system contend with food insecurity, housing instability, access to transportation, etc., as critical challenges specifically termed "comorbidities."[25] Critical stakeholders including the Centers for Medicare and Medicaid Services, state policy makers, and private-sector entities require the development of measurement tools that are as dynamic as the societal contexts, policies, and demographic shifts they must adjust to, such as the development of better models for improving population health and delivering both quality-focused and equity-focused care.[25,26] In a country faced with rampant health disparities, supportive metrics like the SVM can assist with risk assessment, equitable allocation of resources, and improving health equity. Through model-based measurement, the SVM enhances existing data-informed approaches to incorporating SDoH that could lend insights into targeting interventions to the most socially vulnerable communities.[27]

ACKNOWLEDGMENTS

This project was supported by NIH/National Institute of Aging (NIA), R56 AG066127(Robert D. Gibbons and Diane S. Lauderdale) and NIH/National Human Genome Research Institute (NHGRI) K08 HG011505 (Loren Saulsberry).

CONFLICT OF INTEREST

Dr. Gibbons founded the company Adaptive Testing Technologies, which distributes mental health computerized adaptive tests. The SVM is being hosted by Adaptive Testing Technologies without charge. These activities have been reviewed and approved by the University of Chicago in accordance with its conflict of interest policies.

ORCID

Loren Saulsberry https://orcid.org/0000-0003-0325-2329

REFERENCES

1. World Health Organization. Social determinants of health. 2022. Accessed April 21, 2022. https://www.who.int/health-topics/social-determinants-of-health#tab=tab_1
2. US Department of Health and Human Services, Healthy People 2030. Social Determinants of Health. Accessed April 21, 2022. https://health.gov/healthypeople/objectives-and-data/social-determinants-health
3. Hood CM, Gennuso KP, Swain GR, Catlin BB. County health rankings: relationships between determinant factors and health outcomes. *Am J Prev Med.* 2016;50(2):129-135. doi:10.1016/j.amepre.2015.08.024
4. Institute of Medicine (US) Committee on Understanding and Eliminating Racial and Ethnic Disparities in Health Care. In: Smedley BD, Stith AY, Nelson AR, eds. *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care*. National Academies Press (US); 2003.
5. Magnan S. *Social Determinants of Health 101 for Health Care: Five plus Five*. NAM Perspectives. Discussion Paper, National Academy of Medicine; 2017 Accessed April 21, 2022. https://nam.edu/social-determinants-of-health-101-for-health-care-five-plus-five/

6. Centers for Disease Control and Prevention (CDC). 2021. About Social Determinants of Health (SDoH). Available at: https://www.cdc.gov/socialdeterminants/about.html.
7. Centers for Disease Control and Prevention/ Agency for Toxic Substances and Disease Registry/Geospatial Research, Analysis, and Services Program. CDC/ATSDR Social Vulnerability Index 2018 Database US. Accessed April 21, 2022. https://www.atsdr.cdc.gov/placeandhealth/svi/data_documentation_download.html
8. Flanagan B, Gregory E, Hallisey E, Heitgerd J, Lewis B. A social vulnerability Inde for disaster management. *J Homeland Secur Emergency Manage*. 2011;8(1):1-24. doi:10.2202/1547-7355.1792
9. An R, Xiang X. Social vulnerability and obesity among U.S. Adults. *Int J Health Sci*. 2015;3(3):7-12.
10. Area Deprivation Index (ADI). Neighborhood Atlas. Center for Health Disparities Research. University of Wisconsin School of Medicine and Public Health. Accessed August 22, 2022. https://www.neighborhoodatlas.medicine.wisc.edu/
11. Agency for Healthcare Research and Quality (AHRQ), Rockville, MD. Social Determinants of Health Database (Beta Version). Content last reviewed June 2021. Accessed August 22, 2022. https://www.ahrq.gov/SDoH/data-analytics/SDoH-data.html
12. Gibbons RD, Hedeker DR. Full-information item bi-factor analysis. *Psychometrika*. 1992;57:423-436. doi:10.1007/BF02295430
13. Gibbons RD, Bock RD, Hedeker D, et al. Full-information item Bifactor analysis of graded response data. *Appl Psychol Meas*. 2007;31(1): 4-19. doi:10.1177/0146621606289485
14. Bock RD, Gibbons RD. *Item Response Theory*. John Wiley & Sons, Inc.; 2021.
15. Holzinger KJ, Swineford F. The Bi-factor method. *Psychometrika*. 1937;2:41-54. doi:10.1007/BF02287965
16. Bock RD, Aitkin M. Marginal maximum likelihood estimation of item parameters: application of an EM algorithm. *Psychometrika*. 1981;46: 443-459. doi:10.1007/BF02293801
17. Little RJA, Rubin DB. *Statistical Analysis with Missing Data*. 2nd ed. Wiley (New York); 2002.
18. Gibbons RD, Weiss DJ, Pilkonis PA, et al. Development of a computerized adaptive test for depression. *Arch Gen Psychiatry*. 2012;69(11): 1104-1112. doi:10.1001/archgenpsychiatry.2012.14
19. Stan AD, Tamminga CA, Han K, et al. Associating psychotic symptoms with altered brain anatomy in psychotic disorders using multidimensional item response theory models. *Cereb Cortex*. 2020;30(5):2939-2947. doi:10.1093/cercor/bhz285
20. United States Census Bureau. Relationship Files. 2022. Accessed August 22, 2022. https://www.census.gov/geographies/reference-files/2010/geo/relationship-files.html#par_textimage_674173622
21. Duran DG, Pérez-Stable EJ. Science visioning to advance the next generation of health disparities research. *Am J Public Health*. 2019; 109(S1):S11-S13. doi:10.2105/AJPH.2018.304944
22. Palmer RC, Ismond D, Rodriquez EJ, Kaufman JS. Social determinants of health: future directions for health disparities research. *Am J Public Health*. 2019;109(S1):S70-S71. doi:10.2105/AJPH.2019.304964
23. United Health Group. UnitedHealthcare Introduces the Use of Predictive Analytics to Expand its Capabilities to Address Social Determinants of Health. 2021. Accessed April 21, 2022. https://www.unitedhealthgroup.com/newsroom/2021/2021-7-8-uhc-predictive-analytics-social-determinants-health.html
24. Buckingham WR, Bishop L, Hooper-Lane C, et al. A systematic review of geographic indices of disadvantage with implications for older adults. *JCI Insight*. 2021;6(20):e141664. doi:10.1172/jci.insight.141664
25. Agrawal S, Chen AH, Price G, Perla R. To Advance a National Health and Equity Infrastructure, measure drivers of health. *Health Affairs Forefront*. 2022. Accessed August 22, 2022. doi:10.1377/forefront.20220628.771279
26. Dzau VJ, Mate K, O'Kane M. Equity and quality—improving health care delivery requires both. *JAMA*. 2022;327(6):519-520. doi:10.1001/jama.2022.0283
27. Tipirneni R. A data-informed approach to targeting social determinants of health as the root causes of COVID-19 disparities. *Am J Public Health*. 2021;111(4):620-622. doi:10.2105/AJPH.2020.306085

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Saulsberry L, Bhargava A, Zeng S, et al. The social vulnerability metric (SVM) as a new tool for public health. *Health Serv Res*. 2022;1-9. doi:10.1111/1475-6773.14102