Development and Validation of a County-Level Social Determinants of Health Risk Assessment Tool for Cardiovascular Disease

Young-Rock Hong, PhD, MPH[1]

Arch G. Mainous III, PhD[1,2]

[1]Department of Health Services Research, Management and Policy, University of Florida, Gainesville, Florida

[2]Department of Community Health and Family Medicine, University of Florida, Gainesville, Florida

MORE ONLINE
www.annfammed.org

ABSTRACT

PURPOSE Social determinants of health (SDoH) have been linked to a variety of health conditions, but there are no multivariate measures of these determinants to estimate the risk of morbidity or mortality in a community. We developed a score derived from multivariate measures of SDoH that predicts county-level cardiovascular disease (CVD) mortality.

METHODS Using county-level data from 3,026 US counties, we developed a score considering variables of neighborhood socioeconomic status, food/lifestyle environment, and health care resource availability and accessibility to predict the 3-year average (2015-2017) age-adjusted county-level mortality rate for all CVD. We used one 50% random sample to develop the score and the other to validate the score. A Poisson regression model was developed to estimate parameters of variables while accounting for intrastate correlation.

RESULTS The index score was based on 7 SDoH factors: percentage of the population of minority (nonwhite) race, poverty rate, percentage of the population without a high school diploma, grocery store ratio, fast-food restaurant ratio, after-tax soda price, and primary care physician supply. The area under the curve for the development and validation groups was similar, 0.851 (95% CI, 0.829-0.872) and 0.840 (95% CI, 0.817-0.863), respectively, indicating excellent discriminative ability. The index had better predictive performance for CVD burden than other area-level indexes: poverty only (area under the curve = 0.808, P <.001); the Centers for Disease Control and Prevention's Social Vulnerability Index (CDC-SVI) (area under the curve = 0.786, P <.001); and the Agency for Healthcare Research and Quality's Socioeconomic Status (AHRQ-SES) index (area under the curve = 0.835, P = .03).

CONCLUSIONS Our validated multivariate SDoH index score accurately classifies counties with high CVD burden and therefore has the potential to improve CVD risk prediction for vulnerable populations and interventions for CVD at the county level.

INTRODUCTION

Where a person lives is now thought to be an essential determinant of health.[1] The social and economic opportunities along with the resources and supports available in our homes, neighborhoods, and communities are considered social determinants of health (SDoH).[1,2] Evidence suggests that these factors influence health status and mortality by shaping opportunities or barriers to health-related behaviors and access to care.[3-9] Consequently, considerable interest has focused on how to incorporate SDoH into our assessment of health risks for populations and even individuals.[10,11]

Cardiovascular disease (CVD) is the leading cause of death in the United States.[12] Clear evidence has established that adhering to a healthy lifestyle (eg, eating a healthy diet and getting physical activity), having good access to care, and living in a neighborhood with health-promoting resources (ie, infrastructure or built environment) can reduce the risk and

Conflicts of interest: authors report none.

CORRESPONDING AUTHOR

Young-Rock Hong, PhD, MPH
Department of Health Services
Research, Management and Policy
College of Public Health
and Health Professions
University of Florida
PO Box 100195
Gainesville, FL 32610
youngrock.h@phhp.ufl.edu

burden of CVD.[13-16] Some individual-level CVD risk scores have attempted in a general way to address SDoH by incorporating poverty into their scores.[17] Although poverty is one general assessment of SDoH, it fails to address the relative contribution of several associated variables on health outcomes. What is missing is a multivariate assessment of SDoH context for predicting CVD risk on a community level. Such an assessment would help plan for appropriate health services in higher-risk communities.

The goal of this study was to develop and validate a multivariate index score for community- (county-) level CVD mortality for the entire Unites States that incorporates various SDoH factors. Knowing the relevant attributes of SDoH and being able to categorize counties with higher CVD burden would be of substantial benefit for health promotion at the population level.

METHODS

Study Design, Data, and Sample

This study was a retrospective, cross-sectional, county-level effort to synthesize the effects of neighborhood factors on variations in total CVD mortality rate. To compile information available on US county-level characteristics, we used various secondary data sources including 2015-2017 Centers for Disease Control and Prevention's Wide-Ranging Online Data for Epidemiologic Research (CDC Wonder)[18]; 2014-2015 Food Environment Atlas Data File (Food Atlas) from the US Department of Agriculture (USDA) Economic Research Service[19]; and 2015-2017 Area Health Resource File (AHRF) data from the US Department of Health and Human Services.[20] CDC Wonder provides access to information on detailed causes of death based on death certificates for residents in the 50 states and District of Columbia. The Food Atlas contains geographically based information on food and lifestyle measures, including the community's ability to access healthy food, food prices, nutrition assistance programs, and other community resources such as store-restaurant mix and physical activity facilities.[19] The AHRF data are compiled from more than 50 databases including the US Census, American Community Survey, and American Medical Association's Masterfile to provide comprehensive area-level information on health facilities, health care professionals, health service use, and other economic and population characteristics.[20] All data were merged into a single database using Federal Information Processing Standards codes.

The county was the unit of analysis, and a total of 3,232 counties were included in the merged data. We excluded 118 counties in Alaska, Hawaii, and US territories (Guam, Puerto Rico, etc) because they are noncontiguous parts of the country and often have missing information, and 88 counties with missing or unreliable estimates on our outcome measure. Analyses were thus based on data from 3,026 counties. We used a 50% random sample (1,513 counties) to develop the index score and the other 50% random sample to validate it.

This study was deemed non–human subjects research and therefore exempted from review by the University of Florida Institutional Review Board.

Primary Outcome: CVD Mortality Rate

Our primary outcome was the 3-year average (2015-2017) age-adjusted county-level mortality rate for all CVD. In the CDC Wonder detailed mortality database, presumed underlying causes of death at the county level are categorized using *International Statistical Classification of Diseases and Related Health Problems, Tenth Edition* (ICD-10), and we included codes I00 through I99.[18] We opted to use the 3-year average mortality to provide stable estimates because of possible year-to-year variations in mortality rates at the county level, consistent with prior work on area-level health outcomes.[9,21,22] The CVD mortality rates expressed per 100,000 population were age-adjusted using the direct method with the 2015 Census estimates as the standard population, and with the following age categories: younger than 1 year, 1-4, 5-14, 15-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75-84, 85 years and older.[23,24]

Selection of SDoH Variables

We tried to capture all possible SDoH factors (ie, candidate variables) based on a review of evidence.[1-9,13-16] We constructed and summarized hypothesized pathways in directed acyclic graphs to identify measurable factors in given data sets (Supplemental Figure 1, available at https://www.AnnFamMed.org/content/18/4/318/suppl/DC1/). Three SDoH contextual domains eventually emerged: (1) socioeconomic status (SES), (2) food and lifestyle environment, and (3) health care resources. Detailed data sources used in this study are listed in Supplemental Table 1 (available at https://www.AnnFamMed.org/content/18/4/318/suppl/DC1/).

Socioeconomic Status

We considered county-level socioeconomic characteristics related to health outcomes[3,4,6,8] in the SES domain. Those included were percentage of the population made up by racial minorities (nonwhite), as a proxy of population diversity; percentage of residents aged 25 years and older without a high school diploma; percentage of residents with 4 years of college or more; percentage of families living below federal poverty level; median household income; percentage of residents aged 16 years and older without employment; and percentage of

residents covered by health insurance plans. The USDA's Rural-Urban Continuum Codes[19] were used to measure county urbanization (metropolitan vs nonmetropolitan).

Food and Lifestyle Environment

We defined the food and lifestyle environment as availability of facilities and area-level features affecting diet quality and physical activity. To capture these contextual influences, we included the density (count) of grocery stores, convenience stores, fast-food restaurants, full-service restaurants, and recreational and fitness facilities per 1,000 residents, as well as the after-tax soda prices (soda tax rates applied) at the county level from the Food Atlas data. According to the North American Industry Classification System,[19] a grocery store was defined as a store primarily retailing fresh fruits and vegetables, fresh and prepared meats and fish, and other general lines of food (eg, canned and frozen foods). Other stores retailing a limited line of goods and snacks were defined as convenience stores (including gas stations). Fast-food restaurants were defined as establishments providing food services without restaurant servers in which patrons generally pay before eating (ie, drive-in and fast-casual restaurants); full-service restaurants were those where patrons were served while seated and paid after eating (ie, diners and family restaurants). Recreation and fitness facilities comprised amenities and establishments for active physical activities and recreational sports, including exercise centers, sports courts, and other outdoor recreational complexes. After-tax soda prices were calculated by multiplying regional tax rates (percentage) on soda purchased by average price of carbonated and sweetened beverage. We calculated the local average price of sodas in actual dollars using the ratio of soda price to the national average price ($1.25 per liter in 2015).[25]

Health Care Resources

To account for health care resources (availability and accessibility), several measures of health professionals and facilities were selected from the AHRF data. We included density of primary care physicians (PCPs), specialists, nonphysician health care professionals (nurse practitioners, clinical nurses, physician assistants), any type of hospitals, and other community health facilities (federally qualified health centers, rural health clinics, community health centers). To be consistent with other density measures (calculated per 1,000 residents), each measure was calculated by dividing the number of professionals or facilities by the total population estimates (2015 Census estimates) and multiplied by 1,000.

Data Analyses

Construction of the Index Score

To create the Hong-Mainous CVD (HM-CVD) index,

we used an approach analogous to that used for the Reynolds risk score developed based on regression coefficients.[26] The rationale behind this approach was that categorizing county characteristics (eg, quartiles, quintiles) did not result in meaningful risk stratifications during our preliminary analysis using multivariate logistic regression analysis (converting significant odds ratios into integer scores). Although it may be relatively difficult to use the regression coefficient approach as compared with integer scoring, this approach is generally adopted and used in most risk assessment tool development[26-28] and likely yields more precise estimates with improved model fit.[28]

To develop our index score, we used a randomly selected 50% sample of counties (1,513 counties) with complete data for CVD mortality rates and other key SDoH variables. For main analysis, we used the food store ratio (grocery/convenience stores; for healthy, fresh foods availability) and restaurant ratio (fast-food/full-services restaurants; for consumption of energy-dense and nutrient-poor foods) to reflect their comparative values while avoiding possible multicollinearity, consistent with previous work.[29] The median household income was z-scored to be estimated with quantifiable coefficients.

We fit a Poisson regression model to estimate variable parameters and used robust standard errors clustered at the state level to account for intrastate correlation.[30,31] All candidate and outcome variables were included in the initial model; thereafter, variables were selected using stepwise backward elimination with county-level CVD mortality rates as the dependent variable. We confirmed the absence of multicollinearity among retained variables using the variance inflation factor (<4.0) and tolerance (>0.2).[32] Under the stepwise approach, the best-fitting predictive model included 7 variables: percentage of population of minority race, family poverty rate, percentage of population without high school diploma, grocery store ratio, fast-food restaurant ratio, after-tax soda price, and density of PCPs. We confirmed this set of variables using an iterative forward-selection procedure based on a correlation matrix of all candidate variables (Supplemental Table 2, available at https://www.AnnFamMed.org/content/18/4/318/suppl/DC1/). Model fit was assessed using quasi-Akaike information criteria (QAIC).[30,31]

On the basis of the final best-fitting predictive model, we calculated a score by summing of the estimated regression coefficients multiplied by values of the corresponding retained variables. We then standardized the computed index scores (range = 0 to 100; mean = 50; standard deviation = 10) to allow easier calculation and standardized comparison between counties; higher index scores suggest greater CVD burden.

Validation and Comparison With Other Measures

We validated the HM-CVD index using the other random 50% sample of counties (1,513 counties), and compared its performance with that of 3 other indicators frequently used for area-level assessment and risk adjustment: poverty rate, the AHRQ-SES index,[33] and the CDC-Social Vulnerability Index (CDC-SVI).[34] Descriptions and calculations of these measures are given in the Supplemental Appendix (available at https://www.AnnFamMed.org/content/18/4/318/suppl/DC1/) and in more detail elsewhere.[33,34] We estimated logistic models with a dichotomized CVD outcome variable: high CVD burden (score of 1) vs low-intermediate burden (score of 0). We categorized counties with age-adjusted CVD mortality rates above the 75th percentile among all US counties (>283.2 deaths per 100,000 population) as having high CVD burden, given that top-quartile level has been extensively used as a high-risk threshold or health outcome indicator.[35,36]

For validation, we applied the standardized scoring formula to each county in the validation group. Then, discrimination and calibration performance were assessed. Discrimination performance, ability to separate counties with vs without high CVD burden, was assessed with area under the curve (AUC; equivalent to the C statistic) in a receiver operating characteristic curve generated from logistic models.[28] Calibration performance was assessed with a calibration plot (whereby a calibration slope close to 1 indicates satisfactory calibration)[27] by comparing the predicted number of counties with high CVD burden with the observed number of counties in index score quintile groups (0-20, 21-40, 41-60, 61-80, and 81-100). To compare the predictive performance of the HM-CVD index with the 3 area-level indexes, we also computed AUCs in receiver operating characteristic curves for each of those models.

All the analyses were performed using SPSS Complex Survey 24 (IBM Corp) and SAS 9.4 (SAS Institute Inc). We used 2-sided *P* values, with significance set at <.05.

RESULTS

A total of 3,026 US counties were eligible. Distribution of the SDoH variables included in the index score was generally similar between the 1,513 counties in the development group and the 1,513 counties in the validation group (Table 1).

HM-CVD Index Score

The resulting HM-CVD index score was based on 7 SDoH factors: percentage of minority population, family poverty rate, percentage of population without a high school diploma, grocery store ratio, fast-food restaurant ratio, after-tax soda price, and density of PCPs. Table 2 presents the coefficient estimates for each

Table 1. County Variables in the Development and Validation Groups

Variable	Value, Median (IQR) or Mean (SD)	
	Development Group (n = 1,513)	Validation Group (n = 1,513)
Socioeconomic status		
Population of minority race,[a] %	10.6 (3.5-30.1)	10.7 (3.6-29.1)
Median annual income, $	46,340 (40,201-53,508)	47,107 (40,454-54,519)
Family poverty rate, %	11.8 (8.3-15.4)	11.2 (8.1-15.0)
Population without high school diploma, %	13.4 (9.9-19.3)	13.0 (9.4-18.5)
Population with ≥4 years of college, %	18.0 (14.0-23.8)	18.3 (14.3-24.1)
Population uninsured, %	21.4 (16.9-26.4)	21.4 (16.6-25.8)
Population unemployed, %	5.3 (4.2-6.6)	5.3 (4.2-6.5)
Metropolitan area[b]	0.38 (0.48)	0.40 (0.48)
Food/lifestyle environment		
Density of grocery stores[c]	0.20 (0.15-0.28)	0.20 (0.14-0.28)
Density of convenience stores[c]	0.55 (0.41-0.71)	0.55 (0.41-0.73)
Grocery store ratio[d]	0.36 (0.25-0.50)	0.37 (0.25-0.54)
Density of fast-food restaurants[c]	0.58 (0.42-0.73)	0.59 (0.43-0.75)
Density of full-service restaurants[c]	0.69 (0.50-0.90)	0.68 (0.50-0.89)
Fast-food restaurant ratio[e]	0.90 (0.59-1.19)	0.88 (0.57-1.17)
Soda price ratio[f]	0.97 (0.95-1.01)	0.97 (0.95-1.03)
Soda tax rate, %	5.00 (1.50-6.15)	5.00 (1.23-6.15)
Tax-adjusted soda price, $ per liter	1.03 (0.99-1.06)	1.02 (0.99-1.06)
Density of recreation facilities[c]	0.63 (0.00-1.13)	0.63 (0.00-1.15)
Health care resources		
Density of primary care physicians[c]	0.49 (0.31-0.72)	0.48 (0.30-0.71)
Density of specialists[c]	0.19 (0.06-0.45)	0.19 (0.05-0.49)
Density of nonphysician health care professionals[c]	1.13 (0.75-1.66)	1.12 (0.69-1.68)
Density of community health facilities[c]	0.06 (0.01-0.15)	0.05 (0.01-0.15)
Density of hospitals[c]	0.03 (0.01-0.06)	0.03 (0.01-0.06)

IQR = interquartile range.

[a] Race other than white.
[b] Based on Rural-Urban Continuum Codes; a dichotomous variable (1 = metropolitan area, 0 = nonmetropolitan area). Expressed as mean (SD).
[c] Per 1,000 residents.
[d] Ratio of the number of grocery stores to the number of convenience stores.
[e] Ratio of the number of fast-food restaurants to the number of full-service restaurants.
[f] Ratio of the local average price of soda to the national average price.

Table 2. Coefficients for the 7 Variables Included in the HM-CVD Index Score

Variable	Coefficient (SE)	P Value
Socioeconomic status		
Population of minority race	−0.0024 (0.0005)	<.001
Family poverty rate	0.0159 (0.0019)	<.001
Population without high school diploma	0.0076 (0.0016)	<.001
Food/lifestyle environment		
Grocery store ratio	−0.0389 (0.0151)	.01
Fast-food restaurant ratio	0.0498 (0.0135)	<.001
Tax-adjusted soda price	−0.4265 (0.1060)	<.001
Health care resources		
Density of primary care physicians	−0.0603 (0.0237)	.01

HM-CVD = Hong-Mainous cardiovascular disease; SE = standard error.

Note: See Table 1 for description and units of measure for each variable.

Figure 1a. ROC curve for the HM-CVD index score in the development group.

AUC = 0.851 (95% CI, 0.829-0.872)

AUC = area under the curve; HM-CVD = Hong-Mainous cardiovascular disease; ROC = receiver operating characteristic.

Figure 1b. ROC curve for the HM-CVD index score in the validation group.

AUC = 0.840 (95% CI, 0.817-0.863)

AUC = area under the curve; HM-CVD = Hong-Mainous cardiovascular disease; ROC = receiver operating characteristic.

factor from the final best-fitting model; the standardized HM-CVD index score was calculated by the following formula:

HM-CVD index score = 70.150 + % of minority residents × −0.165) % of residents without a high school diploma × 0.521) + (poverty rate × 1.090) + (grocery store ratio × −2.667) + (fast-food restaurant ratio × 3.414) + (after-tax soda price × −29.240) + (PCP density × −4.134)

Supplemental Figure 2 (available at https://www.AnnFamMed.org/content/18/4/318/suppl/DC1/) displays a heatmap of the HM-CVD index score for the United States, visually depicting the spatial distribution of predicted CVD mortality risk across counties.

The AUCs for the development and validation models were similar, 0.851 (95% CI, 0.829-0.872) and 0.840 (95% CI, 0.817-0.863), respectively (Figure 1a and Figure 1b), indicating excellent discriminative ability. When index-predicted and observed high CVD burden were compared, the calibration plot suggested the model relatively overestimated the proportion of counties with high CVD burden for the top-quintile group (index score range of 81-100), with a predicted value of 90.5% vs an observed value of 55.0% (Figure 2). Based on the calibration slope (0.998) and intercept (0.001), however, predicted and observed high CVD burden agreed well across the index score quintile groups; therefore, calibration of the model was satisfactory.
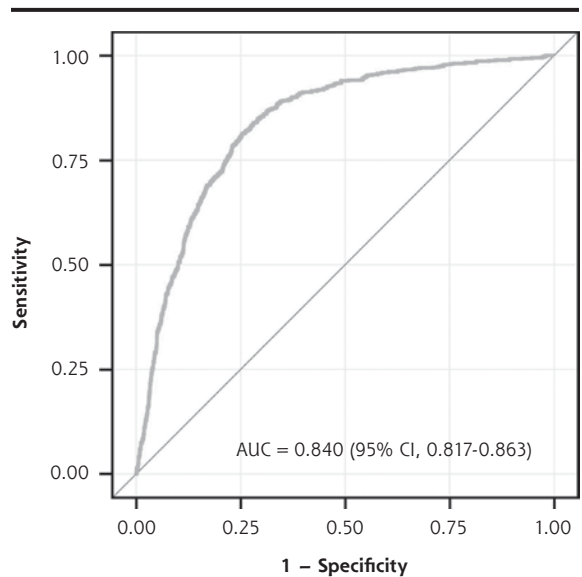
Supplemental Table 3 (available at https://www.AnnFamMed.org/content/18/4/318/suppl/DC1/) summarizes the performance of the HM-CVD index across different cut-point scores. Using higher cut-points tended to identify a smaller number of counties with high CVD burden (ie, lower sensitivity) but

resulted in higher specificity and positive predictive value. Negative predictive value was high (≥75%) across the cut-point scores. Given the highest positive predictive value and negative predictive value, using a cut-point score of 60 or greater seemed to perform best at identifying counties with high CVD burden.

Figure 2. Calibration plot for observed and predicted proportions of US counties with high CVD burden.
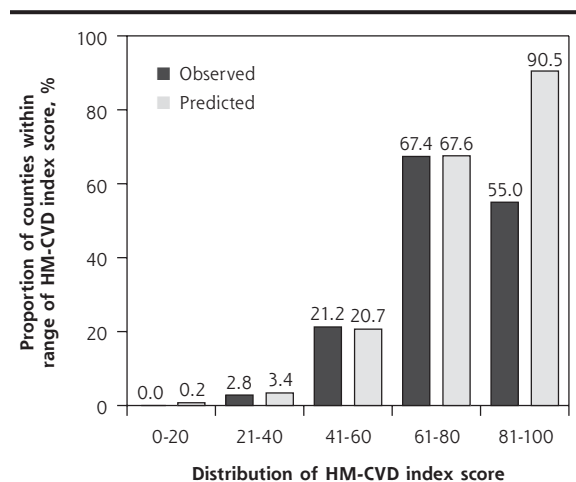
CVD = cardiovascular disease; HM-CVD = Hong-Mainous cardiovascular disease.

Note: intercept = .001; calibration slope = .998.

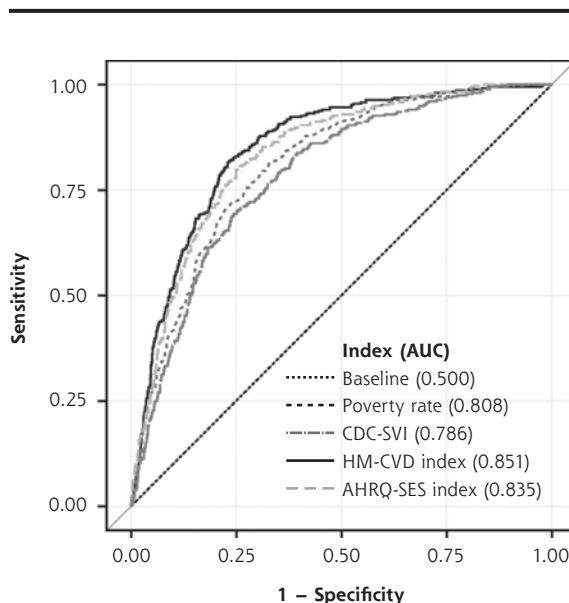**Figure 3. Comparison of predictive performance for counties with high CVD burden.**

AHRQ-SES = Agency for Healthcare Research and Quality's Socioeconomic Status; AUC = area under the curve. CDC-SVI = Centers for Disease Control and Prevention's Social Vulnerability Index; CVD = cardiovascular disease; HM-CVD = Hong-Mainous cardiovascular disease; ROC = receiver operating characteristic.

Comparison of HM-CVD With Other Measures

The AUC of the HM-CVD index was significantly higher than that of the poverty rate (0.851 vs 0.808, $P <.001$) and the CDC-SVI score (0.851 vs 0.786, $P <.001$) (Figure 3). The performance of the HM-CVD index appeared to be close to that of the AHRQ-SES index; however, the difference between the AUCs was significant (0.851 vs 0.835, $P = .03$).

DISCUSSION

We developed and validated the HM-CVD index, a county-level score using SDoH to identify counties having high CVD burden. After investigating a variety of SDoH candidate variables, we included 7 in the index: minority population, lack of high school education, poverty rate, grocery stores, fast-food restaurants, average price of soda, and PCP supply. The associations between these SDoH variables and CVD mortality observed herein are consistent with those seen in previous studies of area-level indicators and population health outcomes.[1-9,13-16] The HM-CVD index had both excellent discrimination and excellent calibration for predicting which counties had a high CVD burden.

Our index score can be easily calculated using publicly available data from the US Census Bureau and federal agencies (eg, USDA and Health Resources and Services Administration), suggesting its practicability in various settings for identifying communities or populations at high risk for CVD and in need of efficient delivery of interventions to improve neighborhood

conditions. Moreover, the index score has potential application as a computerized tool (eg, web-based or integrated into electronic health record system) for assessing county-level resources where patients reside, a SDoH screening tool for use in practice.[35] We provide a full list of HM-CVD index scores for US counties as of 2015 in Supplemental Table 3 (available at https://www. AnnFamMed.org/content/18/4/318/suppl/DC1/) for public use and plan to update them using census data.

The construct of SDoH captures neighborhood resources, including socioeconomic and structural conditions of a person's surroundings. Valid, accurate, and easily reproducible multivariate risk scores to predict morbidity and mortality related to different health conditions have been lacking. The HM-CVD index is therefore a step forward. Moreover, our results add evidence to the growing amount of literature on the importance of SDoH in CVD outcomes and to the general idea of health care outcome assessment.[10,11,35] Much evidence has demonstrated that area-level measures, independent of individual-level factors, are associated with CVD outcomes,[13-16] suggesting the importance of including SDoH data in assessing those health outcomes. Failure to adequately value the SDoH context may contribute to disproportionate distribution of resources in health care.[11,37] By using this index, for example, policy makers and researchers can prioritize

county-level interventions to address deficient neighborhood resources. In addition, CVD interventions can be improved based on the index score with an understanding of where a target population lives, availability of lifestyle resources, and access to care.[11,38]

One challenge to screening for patients' health-related SDoH is the complex nature of potential interrelationships among those factors (ie, beyond individual-level characteristics). Moreover, many validated tools for SDoH are based on SES factors, such as poverty, income, and education, and are generally designed for research purposes.[6,31,32] The HM-CVD index had better predictive utility than poverty rate and the CDC-SVI score, suggesting consideration of neighborhood characteristics other than SES may improve CVD risk assessment. Although performance of the AHRQ-SES index was numerically similar, the HM-CVD index still had significantly better performance, and we believe that our index well reflects the multidimensional character of the neighborhood SDoH status along with community infrastructure (ie, the built environment) and SES factors.

SDoH are important variables affecting the health of populations. As this was an initial effort, we focused on county-level characteristics; however, family physicians can use this score for planning purposes, by understanding where their patient population lives and whether they have a high CVD risk. This is population health at its most basic level in understanding one's patient population. The next step will be to evaluate whether the HM-CVD index has the potential to contribute more to both individual-level and population-level CVD risk assessment. In clinical settings, in particular, there has been increasing emphasis on screening for health-related socioecologic and environmental circumstances that can be modified for patient health-related behaviors and health outcomes.[39,40] Those factors are often overlooked, however, because of difficulties with assessment in clinical settings or a lack of validated multidimensional measures of SDoH.[35]

Our HM-CVD index has some limitations that may affect generalizability of the results. First, we used data limited to county-level information in the United States; thus, applying the scoring formula to other countries may require modifications of the measure. We note, however, that all of the SDoH factors included in the models have generally been considered and found to be associated with various health outcomes in other countries as well.[22,41,42] Second, it might be argued that an index score can be developed using factor analysis, as demonstrated in a previous effort.[39] The risk score approach, however, linking candidate variables with an outcome to be assessed or predicted, could be a more straightforward tool to

inform users of specific SDoH factors/items associated with the outcome. Third, although SDoH may affect multiple diseases and conditions, we focused on CVD. We believed that focusing on a single disease, the leading cause of death in the United States, was appropriate, useful, and consistent with the creation of many other risk scores. Our index demonstrated utility in predicting high CVD burden among US counties across the nation, suggesting broad generalizability. Further research could examine the value of the HM-CVD index for predicting and altering outcomes other than CVD mortality, such as health services use; for assessing the impact of multisite community-based interventions; and for performing risk adjustment when evaluating hospital performance.[11,43] Lastly, the index score was developed based on aggregate county-level data. Additional work should explore how the index affects clinical decision making and individual-level CVD risk assessment (especially for screening or targeted intervention), given the prior effort integrating SDoH factors into existing CVD risk scores.[17,44]

In conclusion, we developed the HM-CVD index, a multivariate score that systematically synthesizes the effects of SDoH factors associated with county-level CVD mortality. The score was validated and proven to be a better-discriminating tool than poverty rate and other area SES measures for predicting which counties have high CVD burden. Our index has the potential to improve CVD risk prediction and quality of intervention delivery at the county level.

References

1. Centers for Disease Control and Prevention. Social determinants of health: know what affects health. https://www.cdc.gov/socialdeterminants/. Published 2018. Accessed Jan 14, 2019.

2. Office of Disease Prevention and Health Promotion. HealthyPeople 2020: social determinants of health. https://www.healthypeople.gov/2020/topics-objectives/topic/social-determinants-of-health. Published 2019. Accessed Dec 12, 2019.

3. Boscoe FP, Johnson CJ, Sherman RL, Stinchcomb DG, Lin G, Henry KA. The relationship between area poverty rate and site-specific cancer incidence in the United States. *Cancer.* 2014;120(14):2191-2198.

4. Sallis JF, Saelens BE, Frank LD, et al. Neighborhood built environment and income: examining multiple health outcomes. *Soc Sci Med.* 2009;68(7):1285-1293.

5. Auchincloss AH, Diez Roux AV, Mujahid MS, Shen M, Bertoni AG, Carnethon MR. Neighborhood resources for physical activity and healthy foods and incidence of type 2 diabetes mellitus: the Multi-Ethnic study of Atherosclerosis. *Arch Intern Med.* 2009;169(18):1698-1704.

6. Kind AJH, 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.

7. Mason KE, Pearce N, Cummins S. Associations between fast food and physical activity environments and adiposity in mid-life: cross-sectional, observational evidence from UK Biobank. *Lancet Public Health.* 2018;3(1):e24-e33.

8. Richard L, Gauvin L, Kestens Y, et al. Neighborhood resources and social participation among older adults: results from the VoisiNuage study. *J Aging Health.* 2013;25(2):296-318.

9. Chan KS, Roberts E, McCleary R, Buttorff C, Gaskin DJ. Community characteristics and mortality: the relative strength of association of different community characteristics. *Am J Public Health.* 2014;104(9):1751-1758.

10. Fuchs VR. Social determinants of health: caveats and nuances. *JAMA.* 2017;317(1):25-26.

11. Ash AS, Mick EO, Ellis RP, Kiefe CI, Allison JJ, Clark MA. Social determinants of health in managed care payment formulas. *JAMA Intern Med.* 2017;177(10):1424-1430.

12. Roth GA, Johnson CO, Abate KH, et al; Global Burden of Cardiovascular Diseases Collaboration. The burden of cardiovascular diseases among US states, 1990-2016. *JAMA Cardiol.* 2018;3(5):375-389.

13. Rutten-Jacobs LCA, Larsson SC, Malik R, et al; MEGASTROKE consortium; International Stroke Genetics Consortium. Genetic risk, incident stroke, and the benefits of adhering to a healthy lifestyle: cohort study of 306 473 UK Biobank participants. *BMJ.* 2018;363:k4168.

14. Calling S, Li X, Kawakami N, Hamano T, Sundquist K. Impact of neighborhood resources on cardiovascular disease: a nationwide six-year follow-up. *BMC Public Health.* 2016;16(1):634.

15. Anthony S, Odgers T, Kelly W. Health promotion and health education about diabetes mellitus. *J R Soc Promot Health.* 2004;124(2):70-73.

16. Mainous AG III, Tanner RJ, Rahmanian KP, Jo A, Carek PJ. Effect of sedentary lifestyle on cardiovascular disease risk among healthy adults with body mass indexes 18.5 to 29.9 kg/m². *Am J Cardiol.* 2019;123(5):764-768.

17. Hippisley-Cox J, Coupland C, Brindle P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. *BMJ.* 2017;357(May):j2099.

18. Centers for Disease Control and Prevention. CDC Wonder. https://wonder.cdc.gov/. Published 2018. Accessed Aug 13, 2018.

19. US Department of Agriculture. Food environment atlas. https://www.ers.usda.gov/data-products/. Published 2018. Accessed Sep 25, 2018.

20. US Department of Health and Human Services. Area Health Resources File (AHRF). https://datawarehouse.hrsa.gov/data/data download.aspx#MainContent_ctl00_gvDD_lbl_dd_topic_ttl_0. Published 2018. Accessed Feb 5, 2019.

21. Ahern M, Brown C, Dukas S. A national study of the association between food environments and county-level health outcomes. *J Rural Health.* 2011;27(4):367-379.

22. Baker R, Honeyford K, Levene LS, et al. Population characteristics, mechanisms of primary care and premature mortality in England: a cross-sectional study. *BMJ Open.* 2016;6(2):e009981.

23. Colby SL, Ortman JM. *Projections of the Size and Composition of the US Population: 2014 to 2060: Population Estimates and Projections.* United States Census Bureau. https://census.gov/content/dam/Census/library/publications/2015/demo/p25-1143.pdf. Published 2015. Accessed Feb 5, 2019.

24. Klein R, Schoenborn C. *Age Adjustment Using the 2000 Projected U.S. Population.* Hyattsville, MD: US Department of Health and Human Services; 2001.

25. Statista. Market Directory: Soft Drinks. https://www.statista.com/outlook/20020000/104/soft-drinks/north-america. Published 2018. Accessed Jan 22, 2019.

26. Ridker PM, Buring JE, Rifai N, Cook NR. Development and validation of improved algorithms for the assessment of global cardiovascular risk in women: the Reynolds Risk Score. *JAMA.* 2007;297(6):611-619.

27. Pavlou M, Ambler G, Seaman SR, et al. How to develop a more accurate risk prediction model when there are few events. *BMJ.* 2015;351:h3868.

28. Koopman RJ, Mainous AG III. Evaluating multivariate risk scores for clinical decision making. *Fam Med.* 2008;40(6):412-416.

29. Mehta NK, Chang VW. Weight status and restaurant availability a multilevel analysis. *Am J Prev Med.* 2008;34(2):127-133.

30. Hanley JA, Negassa A, Edwardes MD, Forrester JE. Statistical analysis of correlated data using generalizing estimating equations: an orientation. *Am J Epidemiol.* 2003;157(4):364-375.

31. Grueber CE, Nakagawa S, Laws RJ, Jamieson IG. Multimodel inference in ecology and evolution: challenges and solutions. *J Evol Biol.* 2011;24(4):699-711.

32. O'Brien RM. A caution regarding rules of thumb for variance inflation factors. *Qual Quant.* 2007;41(5):673-690.

33. Agency for Healthcare Research and Quality. Chapter 1. Creation of new race-ethnicity codes and socioeconomic status (SES) indicators for Medicare beneficiaries. https://archive.ahrq.gov/research/findings/final-reports/medicareindicators/medicareindicators1.html. Published 2012. Accessed Jan 8, 2019.

34. Agency for Toxic Substances and Disease Registry. CDC's Social Vulnerability Index (SVI). https://svi.cdc.gov/. Published 2018. Accessed Jan 22, 2019.

35. Garg A, Boynton-Jarrett R, Dworkin PH. Avoiding the unintended consequences of screening for social determinants of health. *JAMA.* 2016;316(8):813-814.

36. Chen LM, Epstein AM, Orav EJ, Filice CE, Samson LW, Joynt Maddox KE. Association of practice-level social and medical risk with performance in the Medicare physician value-based payment modifier program. *JAMA.* 2017;318(5):453-461.

37. Harrison KM, Dean HD. Use of data systems to address social determinants of health: a need to do more. *Public Health Rep.* 2011;126(3_suppl):1-5.

38. Blum AB, Egorova NN, Sosunov EA, et al. Impact of socioeconomic status measures on hospital profiling in New York City. *Circ Cardiovasc Qual Outcomes.* 2014;7(3):391-397.

39. Singh GK. Area deprivation and widening inequalities in US mortality, 1969-1998. *Am J Public Health.* 2003;93(7):1137-1143.

40. McGinnis JM, Williams-Russo P, Knickman JR. The case for more active policy attention to health promotion. *Health Aff.* 2002;21(2):78-93.

41. Braveman P, Egerter S, Williams DR. The social determinants of health: coming of age. *Annu Rev Public Health.* 2011;32(1):381-398.

42. Marmot MG, Stansfeld S, Patel C, et al. Health inequalities among British civil servants: the Whitehall II study. *Lancet.* 1991;337(8754):1387-1393.

43. Sills MR, Hall M, Colvin JD, et al. Association of social determinants with children's hospitals preventable readmissions performance. *JAMA Pediatr.* 2016;170(4):350-358.

44. Molshatzki N, Drory Y, Myers V, et al. Role of socioeconomic status measures in long-term mortality risk prediction after myocardial infarction. *Med Care.* 2011;49(7):673-678.