PREVENTING CHRONIC DISEASE

PUBLIC HEALTH RESEARCH, PRACTICE, AND POLICY

Population Health, Place, and Space: Spatial Perspectives in Chronic Disease Research and Practice

Table of Contents

About the Journal

Preventing Chronic Disease (PCD) is a peer-reviewed public health journal sponsored by the Centers for Disease Control and Prevention and authored by experts worldwide. PCD was established in 2004 by the National Center for Chronic Disease Prevention and Health Promotion with a mission to promote dialogue among researchers, practitioners, and policy makers worldwide on the integration and application of research findings and practical experience to improve population health.

PCD's vision is to serve as an influential journal in the dissemination of proven and promising public health findings, innovations, and practices with editorial content respected for its integrity and relevance to chronic disease prevention.

PCD Staff

Associate Editors

GUEST EDITORIAL

Population Health, Place, and Space: Spatial Perspectives in Chronic Disease Research and Practice

Michele Casper, PhD[1]; Michael R. Kramer, PhD[2]; James M. Peacock, PhD[3];

Adam S. Vaughan, PhD, MPH, MS[1]

Understanding the role of place and space in shaping the geographic distributions of chronic disease is critical to informing appropriate public health responses for chronic disease prevention and treatment. A geospatial perspective on chronic disease expands the focus of public health efforts beyond the individual, providing insights and guidance for action at the community, regional, and/or national levels. Accordingly, the articles in this special collection advance our understanding of population health dynamics and geospatial disparities for a wide range of chronic disease outcomes across 3 broad themes:

1. Examining connections between community-level characteristics and population health
2. Developing and applying spatial statistical methods and new geospatial tools
3. Using maps and geospatial results to guide program and policy decisions

Examining Connections Between Community-Level Characteristics and Population Health

Geospatial studies are uniquely designed to examine the contextual characteristics of communities that may affect opportunities for chronic disease prevention and treatment. The contextual characteristics addressed in this collection, *Population Health, Place, and Space: Spatial Perspectives in Chronic Disease Research and Practice*, range from underlying context (such as neighborhood deprivation [1], racial segregation [2], social capital [3], and resiliency [4]) to the built environment (walkability [5,6], park access [7], and healthy restaurants [8]) and environmental exposures (9). The study comparing cardiovascular disease–resilient neighborhoods with cardiovascular disease–at-risk neighborhoods examines the important, but understudied, concept of neighborhood resiliency as it affects black populations (4). The study of neighborhood risk and pediatric asthma provides additional evidence of the need for interventions that move beyond primary care or clinical settings (1). Through their maps and spatial analyses, these studies reinforce that chronic diseases are not randomly distributed across communities, emphasize that drivers of disease occur at multiple geographic levels, and stress the importance of developing and implementing programs and policies that address the relevant contextual characteristics.

Developing and Applying Spatial Statistical Methods and New Geospatial Tools

This is a time of great advances in the development and application of spatial statistics, spatial tools, spatially referenced data sets, and spatial data visualization — all of which enable public health professionals to more precisely understand and address existing inequities in chronic diseases. Many studies in this collection use state-of-the-art spatial statistics, including Bayesian spatial smoothing (10,11) and the spatial Durbin econometric model (3), along with other advanced spatial analytic techniques, such as hot spot analysis (12) and spatial scan statistics for spatial clustering (13), and trajectory analysis (14). Furthermore, the development of 2 spatial analysis tools is included in this collection – The Peel Walkability Composite Index (6) and the Rate Stabilizing Tool (RST) (11). The Peel Walkability Composite Index uses a diverse range of measures to construct a repeatable measure of neighborhood walkability. The RST responds to the demand for high-quality, local-level estimates of chronic disease, and enables users with

limited statistical expertise to generate reliable local-level, age-standardized, and spatially smoothed measures of chronic disease.

The rapid expansion of geo-referenced data sets is a critical driver of the increasing numbers and sophistication of geo-spatial studies. This collection includes the use of geo-referenced data from electronic health records (2), emergency medical services (EMS) (15), and market research (8). These large geo-referenced data sets have the potential to provide important insights into the geographic patterns and drivers of chronic diseases. One study demonstrates the novel application of a widely used, publicly available geo-referenced data source — Google Street View — for public health promotion (5).

A key element in conducting geospatial studies is striking the balance between the presentation of local-level data at the smallest appropriate geographic unit and the limitations of generating robust estimates in the presence of small population sizes and numbers of health outcomes. The studies in this collection have all successfully navigated this tension and present data across multiple geographic levels (census tract [6,9,15,16], county [10,14,17,18], and locally meaningful definitions of neighborhood [8,12]) with results that are statistically reliable and meaningful to stakeholders. One study developed a spatial statistical approach to overcoming some of the limitations of data that are highly censored for confidentiality reasons, thereby enabling state and local health departments to generate small area estimates using publicly available data (10).

Recognizing the potent communication capacity of maps, several articles in this collection explore novel geospatial visualizations that may supplement more commonly used maps and report data in an approachable and actionable format. For example, ring maps (19) allow the simultaneous visualization of multiple measures related to chronic diseases. Other studies include figures that demonstrate changes in hotspots over time, allowing a single figure to overcome the limitations of typical cross-sectional maps (12). Visualizing spatial data has also allowed first responders to identify patients at risk during a natural disaster (20) and allowed public institutions to collaborate with health systems, community organizations, and the public to use geospatial data to improve public health and address health equity in birth outcomes (20). Many of the studies published in this collection have also used the Chronic Disease GIS Snapshot article type, unique to *Preventing Chronic Disease* (21). GIS Snapshots are brief reports that focus on using maps to communicate the extent of geographic disparities in chronic disease–related outcomes and risk factors with an eye to providing information for guiding chronic disease prevention programs and policies.

Using Maps and Geospatial Results to Guide Program and Policy Decisions

Another key theme in this special collection is the use of geospatial data to inform programs and policies for chronic disease prevention and treatment. For example, the authors of a study about walkability state that, "Understanding the capacity of the built environment to facilitate walking for utilitarian purposes allows public health departments to advocate for strategic land use and infrastructure developments that promote an increase in population physical activity levels" (6). Several studies in this collection document geographic disparities in access to care (eg, for chronic disease management [22], blood pressure medication adherence [17], diabetes prevention programs [18], and asthma prevention programs [1,12]), providing compelling guidance about where facilities and services are needed. A unique study demonstrates the use of real-time GIS to develop and update emergency response for chronically ill veterans during Hurricane Irma (23). From an applied perspective, staff members from 4 health departments (Maine Center for Disease Control and Prevention, New Jersey Department of Health, New York State Department of Health, and Cuyahoga County, Ohio, Board of Health) describe the ways in which GIS has become a critical tool (24). Their article provides specific examples of how health departments use maps and spatial analyses to 1) communicate the burden of disease; 2) inform decisions about resource allocation, policy, and priority communities for intervention efforts; 3) develop culturally competent programs; and 4) assist with program planning, monitoring, and evaluation.

By embracing the benefits of GIS, increasing the volume of spatially referenced public health data, and applying a broad range of spatial statistical tools, public health practitioners and investigators are continually pushing the envelope for using geospatial data to inform surveillance, epidemiologic research, program evaluation, resource allocation, and communication for chronic disease prevention and treatment. We invite readers to engage deeply with the geospatial approaches presented in this special collection, to contemplate further advances in understanding how place and space shape the distribution of chronic diseases, and to apply a geospatial perspective to promote health equity and inform public health action for chronic disease prevention and treatment.

Author Information

Corresponding Author: Michele Casper, PhD, Division for Heart Disease and Stroke Prevention, Centers for Disease Control and Prevention, 1600 Clifton Rd, NE, MS S-107-1, Atlanta, Georgia 30330. Telephone: 770-488-2571. Email: mcasper@cdc.gov.

Author Affiliations: [1]Division for Heart Disease and Stroke Prevention, Centers for Disease Control and Prevention, Atlanta, Georgia. [2]Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, Georgia. [3]Cardiovascular Health Unit, Minnesota Department of Health, St. Paul, Minnesota.

References

1. Gjelsvik A, Rogers ML, Garro A, Sullivan A, Koinis-Mitchell D, McQuaid EL, et al. Neighborhood risk and hospital use for pediatric asthma, Rhode Island, 2005–2014. Prev Chronic Dis 2019;16:E68.

2. Bravo MA, Batch BC, Miranda ML. Residential racial isolation and spatial patterning of hypertension in Durham, North Carolina. Prev Chronic Dis 2019;16:E36.

3. Yang TC, Matthews SA, Sun F, Armendariz M. Modeling the importance of within- and between-county effects in an ecological study of the association between social capital and mental distress. Prev Chronic Dis 2019;16:E75.

4. Kim JH, Lewis TT, Topel ML, Mubasher M, Li C, Vaccarino V, et al. Identification of resilient and at-risk neighborhoods for cardiovascular disease among black residents: the Morehouse-Emory Cardiovascular (MECA) Center for Health Equity Study. Prev Chronic Dis 2019;16:E57.

5. Plascak JJ, Llanos AAM, Chavali LB, Xing CY, Shah NN, Stroup AM, et al. Sidewalk conditions in northern New Jersey: using Google Street View imagery and ordinary kriging to assess infrastructure for walking. Prev Chronic Dis 2019; 16:E60.

6. Mukhtar M, Guillette D, Lapos N, Fitzpatrick S, Jaros R. An objective walkability index for public health and planning in Peel Region, Ontario, Canada. Prev Chronic Dis 2019;16:E86.

7. Thomson JL, Goodman MH, Landry AS. Assessment of town and park characteristics related to physical activity in the lower Mississippi Delta. Prev Chronic Dis 2019;16:E35.

8. Gase LN, Green G, Montes C, Kuo T. Understanding the density and distribution of restaurants in Los Angeles County to inform local public health practice. Prev Chronic Dis 2019; 16:E06.

9. Maroko AR, Pavilonis BT. Occupational groups and environmental justice: a case study in the Bronx, New York. Prev Chronic Dis 2018;15:E139.

10. Quick H. Estimating county-level mortality rates using highly censored data from CDC WONDER. Prev Chronic Dis 2019; 16:E76.

11. Quick H, Tootoo J, Li R, Vaughan AS, Schieb L, Casper M, et al. The rate stabilizing tool: generating stable local-level measures of chronic disease. Prev Chronic Dis 2019;16:E38.

12. Byun J, McDonnell S, Robertson J. Using asthma-related housing complaints to target residents with uncontrolled asthma in Salt Lake County, Utah. Prev Chronic Dis 2019; 16:E63.

13. Kassem AM, Carter KK, Johnson CJ, Hahn CG. Spatial clustering of suicide and associated community characteristics, Idaho, 2010-2014. Prev Chronic Dis 2019;16:E37.

14. Baltrus P, Malhotra K, Rust G, Levine R, Li C, Gaglioti AH. Identifying county-level all-cause mortality rate trajectories and their spatial distribution across the United States. Prev Chronic Dis 2019;16:E55.

15. Byun J, Robertson J. Tracking senior fall and fall-related injury EMS calls to target fall prevention programs, Salt Lake County, Utah. Prev Chronic Dis 2019;16:E48.

16. Joiner J, Jordan M, Reid K, Kintziger K, Duclos C. Economic hardship and life expectancy in Nassau County, Florida. Prev Chronic Dis 2019;16:E27.

17. Pizzonia C, Pied D, Huston SL, Albert PF, Parent G, Morse N. Using local data on adults aged 18 to 64 to tailor interventions for blood pressure medication adherence in Maine. Prev Chronic Dis 2019;16:E80.

18. Zepka B, Anis M, Keith JD, Barksdale D, Rivera C. Using geographic information systems to highlight diabetes prevention program expansion areas in Pennsylvania. Prev Chronic Dis 2019;16:E41.

19. Lòpez-DeFede A, Stewart JE. Diagnosed diabetes prevalence and risk factor rankings, by state, 2014–2016: a ring map visualization. Prev Chronic Dis 2019;16:E44.

20. Jankowska MM, Yang J-A, Block J, Baer RJ, Jelliffe-Pawlowski LL, Flores S, et al.An online geographic data visualization tool to relate preterm births to environmental factors. Prev Chronic Dis 2019;16(E)

21. Centers for Disease Control and Prevention. GIS Snapshots 2019. https://www.cdc.gov/pcd/for_authors/types_of_articles.htm#GIS%20Snapshots. Accessed July 2, 2019.

22. Bobitt J, Aguayo L, Payne L, Jansen T, Schwingel A. Geographic and social factors associated with chronic disease self-management program participation: going the "extra-mile" for disease prevention. Prev Chronic Dis 2019;16:E25.

23. Katzburg J, Wilson D, Fickel J, Lind J, Cowper-Ripley D, Fleming M, et al.Ensuring the safety of chronically ill veterans enrolled in home-based primary care. Prev Chronic Dis 2019; 16(E)

24. Brissette I, Casper M, Huston SL, Jordan M, Karns B, Kippes C, et al. Application of geographic information systems to address chronic disease priorities: experiences in state and local health departments. Prev Chronic Dis 2019;16(E65).

ORIGINAL RESEARCH

Neighborhood Risk and Hospital Use for Pediatric Asthma, Rhode Island, 2005–2014

Annie Gjelsvik, PhD[1,2]; Michelle L. Rogers, PhD[2]; Aris Garro, MD, MPH[3,4]; Adam Sullivan, PhD[2,5];

Daphne Koinis-Mitchell, PhD[2,4,6,7]; Elizabeth L. McQuaid, PhD[2,6,7]; Raul Smego, MPH[2];

Patrick M. Vivier, MD, MPH[2,4,8]

PEER REVIEWED

Summary

What is already known on this topic?

Among children, exposure to acute and chronic stress is associated with worse asthma outcomes.

What is added by this report?

Our study shows that higher neighborhood risk is associated with higher odds of hospital reutilization, even after accounting for child-level factors.

What are the implications for public health practice?

Interventions and policies designed to address pediatric asthma should move beyond clinic settings and account for neighborhood context, with a careful evaluation of social context and environmental triggers to address the real-world challenges of managing asthma in high-risk environments.

Abstract

Introduction

Studies consistently show that children living in poor neighborhoods have worse asthma outcomes. The objective of our study was to assess the association between negative neighborhood factors (ie, neighborhood risk) and pediatric asthma hospital use.

Methods

This retrospective study used data from children aged 2 to 17 years in a statewide (Rhode Island) hospital network administrative database linked to US Census Bureau data. We defined an asthma visit as an *International Classification of Diseases, 9th Re-vision, Clinical Modification* (ICD-9-CM) code of 493 in any diagnosis field. We used 8 highly correlated measures for each census-block group to construct an index of neighborhood risk. We used maps and linear regression to assess the association of neighborhood risk with average annual census-block–group rates of asthma emergency department visits and hospitalizations. We used multivariable analyses to identify child characteristics and neighborhood risk associated with an asthma revisit, accounting for the child's sociodemographic information, season, and multiple measurements per child.

Results

From 2005 through 2014, we counted 359,195 visits for 146,889 children. Of these, 12,699 children (8.6%) had one or more asthma visits. Linear regression results showed 1.18 (95% confidence interval, 1.06–1.30) more average annual emergency departments visits per 100 children and 0.41 (95% confidence interval, 0.34–0.47) more average annual hospitalizations per 100 children in neighborhoods in the highest-risk index quintile than in neighborhoods in the lowest-risk index quintile.

Conclusion

Interventions to improve asthma outcomes among children should move beyond primary care or clinic settings and involve a careful evaluation of social context and environmental triggers.

Introduction

Asthma is a chronic illness of the airways and is one of the most common chronic conditions of childhood (1). In 2012, asthma affected 9% (6.8 million) of children aged 0 to 17 years living in the United States (2). Although potentially preventable hospitalizations among children for all diagnoses declined from 2000 to 2007, pediatric hospital stays for asthma may have increased from 2007 to 2009 during the recession (3). In 2013, there were 571,000 emergency department (ED) visits for asthma among children aged 0 to 14 years (4).

Among children, exposure to acute and chronic stress has been associated with increased odds of an asthma diagnosis (5) and increased asthma exacerbations (6). Most studies of neighborhood context demonstrated that negative neighborhood factors — such as poverty, low high school graduation rates, and low median housing prices — are associated with higher pediatric asthma prevalence (7) and risk of adverse outcomes such as hospital use and reutilization (8–16); one study found no association (17). Prior research had several limitations. For example, most studies used census tracts to measure neighborhoods (8–15), a few used census-block groups (7,16,17), and only some accounted for clus- tering of hospital use by children (7,10–12,14,15,17). Block groups are the smallest unit of geography for which the US Census Bureau publishes sample data; they are less heterogeneous than census tracts (18). Without accounting for clustering of hospital use, a study may erroneously identify an association that is due to clustering as one that is due to neighborhood factors.

An ED visit by a child for an asthma exacerbation is a major disruption for the child and family and costly for society (1). Racial/ethnic minority children and poor children are more likely to visit the ED than their non-Hispanic white and nonpoor counterparts (1), and evidence indicates that individual and contextual risks are cumulative (19). Most children who visit an ED do not return to the hospital for emergency care in the subsequent year. For those who do, repeated hospital use may indicate poor asthma management, severe asthma requiring close monitoring, or both (20). Identifying predictors of repeated hospital use for children with asthma has numerous clinical implications. Such predictors may help to 1) characterize social determinants of recurrent urgent health care use for asthma and 2) identify children in need of enhanced discharge services to prevent recurrent health care use and costly hospitalization. Decreasing asthma health disparities will require action on multiple levels, including social and environmental interventions (21). It is important to increase understanding of the role of neighborhood risk in asthma hospital use and revisits.

The objective of our study was to assess the association between negative neighborhood factors (ie, neighborhood risk) and pediatric asthma hospital use, specifically, ED revisits and rehospitalizations within the subsequent year. We conducted a statewide analysis in Rhode Island, where the rate of uninsured children (2% in 2017) has been among the lowest in the country (22); this low rate minimizes the effect of financial resources on health care coverage. We formulated the following hypotheses: 1) census- block–group asthma ED visit and hospitalization rates will be higher in neighborhoods with more risk indicators, 2) children who have an index ED visit and hospitalization and who live in a high-risk neighborhood will be more likely to have a revisit with- in the subsequent year, and 3) these differences will persist when accounting for child-level factors.

Methods

This retrospective study used data from a statewide hospital network administrative database in Rhode Island, the 2010–2014 American Community Survey (23), and the 2010 US Census (24). This hospital network provides approximately two-thirds of pediatric ED services and 90% of inpatient services for children living in the state (unpublished data from the 2014 Rhode Island State Emergency Department Database and 2014 Rhode Island State Inpatient Database) and includes the state's only children's hospital. Children aged 2 to 17 years, living in Rhode Island, with at least 1 asthma visit within this hospital network from January 1, 2005, through December 31, 2014, were identified by using the hospital network's information systems. An asthma ED visit or hospitalization was one in which *International Classification of Diseases, 9th Revision, Clinical Modification* (ICD-9-CM), code 493 was in any diagnosis field. The child's home address at the time of each visit was geocoded by using ArcGIS (Esri) to identify the census-block group in which the child lived (18). This study was approved by the institutional review board of the hospital network.

Measures

Child-level variables. Information on the child's age in years, sex (male, female), race/ethnicity (Hispanic, non-Hispanic black, non-Hispanic white, non-Hispanic other), and insurance coverage (private, public, self-pay/none) was recorded for each visit. We assigned season of visit according to the visit date (spring, March–May; summer, June–August; autumn, September–November; winter, December–February).

Census-block-group–level variables. For each census-block group, we constructed a neighborhood risk index by using 8 highly correlated census-block–group measures obtained from the 2010–2014 American Community Survey and the 2010 US Census: percentage of adults with no high school education, percentage of single-parent households, percentage of household crowding (>1 person per room), percentage of renter-occupied housing units, percentage of vacant homes (excluding vacation homes), percentage of families below 100% of the federal poverty level, percentage of nonwhite residents, and percentage of housing units built before 1950. We computed quintiles for each of the 8 measures and summed these, resulting in an index with a range of 8 to 40, with higher scores indicating greater neighborhood risk. We then computed quintiles for this index. We also dichotomized this index into high risk (at or above the 75th percentile, values of 30–40) and low risk (below the 75th percentile, values of 8–29). This dichotomization resulted in 25% (206 of 809) of census-block groups classified as high risk. These census-block groups accounted for 29% (N = 59,150) of the 2010 Census population count of children age 2 to 17 years in Rhode Island. We also cal-

culated average annual rates of ED visits and hospitalizations by dividing the average number of visits per year by the 2010 Census estimate of children aged 2 to 17 years living in each census-block group.

Index visit and revisits. We focused our analysis on high rates of hospital use, specifically, revisits for asthma after an initial ED visit. We retained data on all characteristics for the initial asthma ED visit or hospitalization (index visit) during the study period. For this analysis, to allow a full 365 days for a second visit to occur for all index visits, we included only index visits occurring before 2014. Visits occurring in 2014 were included only if they were a revisit to an index visit in 2013. If information on a characteristic was missing for the index visit, we used information from the next visit with valid information. We then examined all visits occurring between 8 and 365 days after the index visit. Revisits that occurred within a few days were considered a part of the same course of illness (20). If we found one or more asthma visits within 365 days, then we coded the child as having an asthma revisit. All others were coded as either having no revisit or having a nonasthma revisit. If we found multiple asthma revisits during the period, we used the first asthma revisit in the analyses. After coding the index visit, we then processed all additional asthma visits in the same manner (ie, the first asthma revisit became the second asthma index visit for the child and a new period of 8–365 days was assessed).

Statistical analysis

Census-block–group level. We first created choropleth maps to show the geographic distribution of the neighborhood risk index, the average annual rate of asthma ED visits, and the average annual rate of asthma hospitalization, by census-block group. We then used linear regression to assess the association of neighborhood risk with the average annual census-block–group rate of asthma ED visits and hospitalizations.

Visit level. We conducted analyses by using SAS version 9.4 (SAS Institute, Inc). We computed bivariate analyses to identify child characteristics associated with neighborhood risk at the index visit and with the occurrence of an asthma revisit. Because children can have multiple visits, we used generalized estimating equations with a repeated statement to account for the multiple measurements per child. We used an autoregressive order 1 correlation matrix to obtain dependence-corrected standard errors. The model included the child's sex, age, race/ethnicity, and insurance coverage at the time of the index visit, the season in which the index visit occurred, and the dichotomized neighborhood risk index at the time of the index visit.

Sensitivity analyses. To account for children moving from one neighborhood to another between visits and potentially changing neighborhood risk level, we used a cross-classification model in the next analysis (25), grouping on the neighborhood risk at the index visit (random effect), and we included the neighborhood risk at the follow-up visit (fixed effect). Because this model required 2 visits to assess potential moves, this analytic sample included only children with at least 2 visits. We also assessed whether results were sensitive to a stricter definition of asthma, because there is no gold-standard definition of asthma using ED or inpatient data (26). In the main analyses we defined an asthma ED visit or hospitalization as ICD-9-CM code 493 in any diagnosis field, and in sensitivity analyses we counted only ED visits or hospitalizations with ICD-9-CM code 493 in the primary (first) diagnosis field as an asthma visit.

Because hospital network coverage for ED visits is lower than coverage for hospitalizations, we also conducted analyses that excluded neighborhoods that were farthest from a network hospital. The results of these analyses were the same as those of the whole state; we therefore tabulated statewide results only.

Results

From 2005 through 2014, we counted 319,320 ED visits and 39,875 hospitalizations for 146,889 children aged 2 to 17 years in Rhode Island. Of these children, 12,699 (8.6%) had one or more asthma ED visits or hospitalizations (number of visits = 23,187). About 53% of visits were among children living in high-risk census-block groups.

Census-block–group level. The average annual count and rate per 100 children of pediatric asthma hospital use varied across census-block groups. For ED visits, the average annual count ranged from 0 to 161 (mean 18.3, median 11.0); the average annual rate per 100 children ranged from 0.0 to 5.5 (mean, 0.7; median, 0.5). For hospitalizations, the average annual count ranged from 0 to 134 (mean, 10.3; median, 8.0); the average annual rate per 100 children ranged from 0.0 to 4.0 (mean, 0.4; median, 0.3). Linear regression results showed 1.18 (95% confidence interval [CI], 1.06–1.30) more average annual emergency departments visits per 100 children and 0.41 (95% CI, 0.34–0.47) more average annual hospitalizations per 100 children in neighborhoods in the highest-risk index quintile than in neighborhoods in the lowest-risk index quintile. The highest-risk neighborhoods were concentrated in urban areas of Rhode Island, and distribution of neighborhoods with higher rates of ED use and hospitalizations was consistent with the distribution of higher-risk neighborhoods (Figure 1).
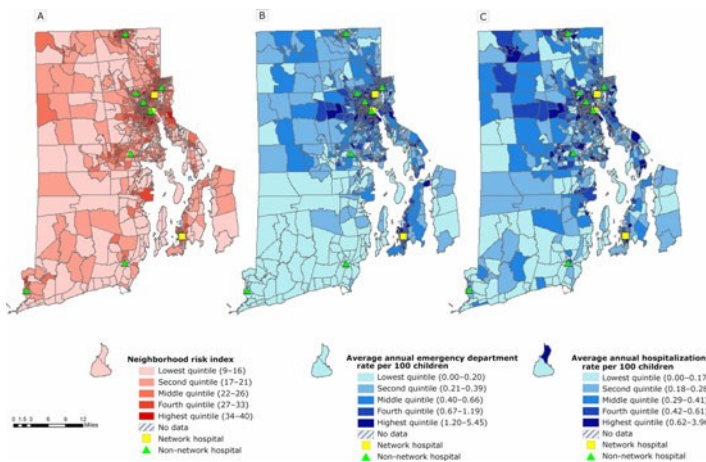
Figure 1. Distribution of quintiles of 3 measures used to assess the association between neighborhood risk and pediatric asthma hospital use among children aged 2 to 17 years (number of asthma emergency department [ED] visits or hospitalizations = 23,187), Rhode Island. A, Neighborhood risk index; the higher the index, the higher the prevalence of adverse socioeconomic and health-related factors, 2010–2014. B, Average annual emergency department visit rate per 100 children, 2005–2014. C, Average annual hospitalization rate per 100 children, 2005–2014. Data on neighborhood risk were collected from the 2010–2014 American Community Survey and the 2010 US Census. Data on emergency department visits and hospitalization were collected from a statewide hospital network administrative database, 2005–2014.

Figure 2. Pediatric asthma emergency department and hospitalization rates per 100 children by quintile of neighborhood risk among children aged 2 to 17 years, Rhode Island, 2005–2014. Data on neighborhood risk were collected from the 2010–2014 American Community Survey and the 2010 US Census. Data on emergency department visits and hospitalization were collected from a statewide hospital network administrative database. n's along the *x*-axis indicate the number of census block groups. Error bars indicate standard error.

Although the average annual rate per 100 children of both pediatric asthma ED visits and hospitalization increased as neighborhood risk quintile increased, the increase was greater for ED visits than for hospitalizations. In the lowest-risk neighborhoods, the ED visit rate (0.27 per 100 children) and the hospitalization rate (0.26 per 100 children) were similar (Figure 2). In the highest-risk neighborhoods, the ED visit rate was 1.45 per 100 children and the hospitalization rate was 0.66 per 100 children.

Visit-level analyses clustered on child

We counted 11,547 children with an index visit occurring during the period from January 1, 2005, through December 31, 2013. After excluding visits that occurred within 7 days of the index visit (n = 860), we counted 19,700 index visits. Of these visits, 28.9% (n = 5,703) had an asthma revisit between 8 and 365 days after the index visit. Compared to children with no asthma revisit, children with an asthma revisit were younger, more likely to have public insurance, be Hispanic or non-Hispanic black, and live in a high-risk neighborhood (Table 1). Hispanic children (adjusted odds ratio [OR] = 1.24; 95% CI, 1.03–1.50) and non-Hispanic black children (adjusted OR = 1.26; 95% CI, 1.02–1.56) had significantly higher odds than non-Hispanic white children of a revisit (Table 2), and children living in high-risk neighborhoods (adjusted OR = 1.22; 95% CI, 1.00–1.48) had significantly higher odds than children living in a low-risk neighborhood. We found no significant differences by age, sex, insurance coverage, or season.

The cross-classification model, which accounted for children moving from one neighborhood to another, showed that children living in high-risk neighborhoods did not have significantly higher odds of a revisit than children living in low-risk neighborhoods (adjusted OR = 1.14; 95% CI, 0.98–1.33) (Table 3). We found significantly lower odds of a revisit among older children than younger children (adjusted OR = 0.92; 95% CI, 0.91–0.93) and in-

dex visits that occurred in the summer than in the autumn (adjusted OR = 0.77; 95% CI, 0.69–0.86). We found significantly higher odds of a revisit among girls than among boys (adjusted OR = 1.15; 95% CI, 1.06–1.25), among children with public insurance than among children with private insurance (adjusted OR = 1.10; 95% CI, 1.00–1.21), and among children who were not non-Hispanic white than among children who were (Hispanic, adjusted OR = 1.13, 95% CI, 1.02–1.27; non-Hispanic black, adjusted OR = 1.43; 95% CI, 1.26–1.62; non-Hispanic other, adjusted OR = 1.30; 95% CI, 1.07–1.58). In the sensitivity analysis in which we counted only ED visits or hospitalizations with ICD-9-CM code 493 in the primary (first) diagnosis field as an asthma visit, the adjusted odds for neighborhood risk were similar to the odds produced in the main analysis. We found no significant results in the model accounting for clustering for children in high-risk neighborhoods, compared with children in low-risk neighborhoods (adjusted OR = 1.27; 95% CI, 0.95–1.70), or the model accounting for children moving (adjusted OR = 1.02; 95% CI, 0.83–1.26) for children in high-risk neighborhoods, compared with children in low risk neighborhoods.

Discussion

Our study demonstrated that increased neighborhood risks contribute to pediatric ED use and hospitalizations for asthma, confirming what has been documented in previous studies (8–16). Our results add to the growing body of evidence that diverse neighborhood factors (eg, crime rates [8], housing code violation density [9], pharmacy access [11], access to primary and specialty care [13], composites of variables specified by the US Census Bureau and the American Community Survey [10,12,14–16]) in various US cities, counties, and states affect pediatric outcomes such as ED visits and hospitalizations. Regardless of the components used to measure neighborhood risk or the level of geography, studies consistently show that children living in worse neighborhoods have higher risks of ED visits and hospitalizations.

In our study, higher levels of neighborhood risk were more strongly associated with pediatric asthma ED visits than with hospitalization rates. The risk for ED visits and hospitalizations was essentially the same in the lowest-risk neighborhoods, but the difference between the 2 types of visits was wide in the highest-risk neighborhoods. This pattern is notable because ED visits are expected to be highly correlated with hospitalizations (the former frequently leading to the latter). This discrepant pattern suggests that other factors may be driving recurrent ED visits in our group of patients, such as limited skills for acute disease management and reliance on the ED for ongoing asthma care. The discrepant pattern could also suggest that proximity to the ED is a factor in rates of ED use. In our study, children living in high-risk neighborhoods had an average distance to a network hospital of 3.5

miles, whereas children living in low-risk neighborhoods had an average distance of 8.6 miles.

Another finding was that higher neighborhood risk was associated with higher odds of a revisit. This association persisted even after accounting for child-level factors. One possible explanation is that children from high-risk neighborhoods may be more likely to return to chaotic and stressful home situations and/or poor housing, where amelioration of asthma triggers is challenging. For instance, it may be more difficult to avoid environmental tobacco smoke (27) or to actively manage asthma triggers such as dust mites and pest problems in publicly financed housing than in a private home (28).

To assess the association between neighborhood risk and the odds of revisit when a child moved from one neighborhood to another between visits (and possibly changing neighborhood risk level), we conducted sensitivity analyses in which we limited the analytic sample to visits among children who had at least one revisit (asthma-related or other). When the analytic sample was limited in this way, the adjusted regression results were similar. Although few children changed neighborhood risk level from index visit to revisit (4% moved from a high-risk neighborhood to a low-risk neighborhood, 3% moved from low-risk to high-risk), when cross-classification was accounted for in regression modeling, the neighborhood risk level of the index visit was no longer significant. Reasons for this could be that the decreased sample size resulted in less power or that residential mobility itself is contributing to stress for the child and family and interruptions in care for chronic illness (29). Similarly, when the definition of an asthma visit was limited to primary diagnosis only, the adjusted odds ratios were similar but not significant. One possible reason for this could be decreased sample size (23,187 visits with ICD-9-CM as any listed diagnosis versus 13,373 visits with ICD-9-CM as primary diagnosis). It is also possible that the broader definition (any listed diagnosis) identified some visits that were not caused by asthma, even though they were visits by children with asthma (30). We found that only 37% of visits with asthma as the secondary diagnosis had respiratory illness as the primary diagnosis. The top 5 diagnosis categories, representing 76% of visits with a secondary diagnosis of asthma, were respiratory disorders (37.0%); signs/symptoms/ill-defined conditions (16.1%); injury and poisoning (8.5%); digestive disorders (7.2%); and infectious and parasitic diseases (7.2%).

Our study had several limitations. One, we obtained data on race/ethnicity from medical records that may not be reliable for these data (31). Two, the neighborhood risk index measured cumulative risk and not individual neighborhood risks. Measurement of cumulative risk was necessary because neighborhood risks are highly correlated, but it did not permit us to disentangle how each risk

factor contributes to pediatric asthma hospital revisits. Three, our data were obtained from one hospital network and excluded children with out-of-network encounters. Thus, asthma-related encounters, especially ED visits, may have been underestimated. However, the hospital network represents approximately two-thirds of all pediatric ED visits and 90% of pediatric admissions and includes the only children's hospital in the state.

Our findings provide additional evidence that interventions and policies designed to address pediatric asthma need to account for neighborhood context (9,32). Interventions that move beyond primary care or clinic setting are required. A careful evaluation of social context (family strengths and supports, financial challenges) and environmental triggers (type of housing, exposure in home and school settings) is needed. Interventions need to address the real-world challenges of managing asthma in high-risk environments. For instance, in Rhode Island, the Rhode Island Asthma Integrated Response Program (RI-AIR) is implementing a comprehensive system for screening and intervention for pediatric asthma that includes school-based education, intensive home-based interventions, and coordination among parents, school nurses, and health care providers for children whose asthma is not well controlled. A health education intervention for parents that does not account for certain factors — whether families are living in older housing stock with mold or multi-unit dwellings with inadequate ventilation or whether children are chronically exposed to triggers in older school buildings — will be less effective than a health intervention that does.

Acknowledgments

Author Information

Corresponding Author: Annie Gjelsvik, PhD, Brown University, Box G-121S, Providence, RI 02912. Telephone: 401-863-2396. Email: Annie_gjelsvik@brown.edu.

Author Affiliations: [1]Department of Epidemiology, Brown University, Providence, Rhode Island. [2]Hassenfeld Child Health Innovation Institute, Providence, Rhode Island. [3]Department of Emergency Medicine, Brown University Warren Alpert Medical School, Providence, Rhode Island. [4]Department of Pediatrics, Brown University Warren Alpert Medical School, Providence, Rhode Island. [5]Department of Biostatistics, Brown University, Providence, Rhode Island. [6]Department of Psychiatry and Human Behavior, Brown University Warren Alpert Medical School, Providence, Rhode Island. [7]Department of Psychiatry, Rhode Island Hospital, Providence, Rhode Island. [8]Department of Health Services, Policy and Practice, Brown University, Providence, Rhode Island.

References

1. Akinbami LJ, Moorman JE, Garbe PL, Sondik EJ. Status of childhood asthma in the United States, 1980-2007. Pediatrics 2009;123(Suppl 3):S131–45.
2. Bloom B, Jones L, Freeman G. Summary health statistics for U.S. children: National Health Interview Survey, 2012. Hyattsville (MD): Centers for Disease Control and Prevention, National Center for Health Statistics, 2013.
3. Berdahl TA, Friedman BS, McCormick MC, Simpson L. Annual report on health care for children and youth in the United States: trends in racial/ethnic, income, and insurance disparities over time, 2002-2009. Acad Pediatr 2013; 13(3):191–203.
4. Centers for Disease Control and Prevention. National Hospital Ambulatory Medical Care Survey: 2013 Emergency department summary tables. 2017. https://www.cdc.gov/nchs/data/ahcd/nhamcs_emergency/2013_ed_web_tables.pdf. Accessed August 8, 2017.
5. Wing R, Gjelsvik A, Nocera M, McQuaid EL. Association between adverse childhood experiences in the home and pediatric asthma. Ann Allergy Asthma Immunol 2015; 114(5):379–84.
6. Sandberg S, Paton JY, Ahola S, McCann DC, McGuinness D, Hillary CR, et al. The role of acute and chronic stress in asthma attacks in children. Lancet 2000;356(9234):982–7.
7. Chang TS, Gangnon RE, Page CD, Buckingham WR, Tandias A, Cowan KJ, et al. Sparse modeling of spatial environmental variables associated with asthma. J Biomed Inform 2015; 53:320–9.
8. Beck AF, Huang B, Ryan PH, Sandel MT, Chen C, Kahn RS. Areas with high rates of police-reported violent crime have higher rates of childhood asthma morbidity. J Pediatr 2016; 173:175–182.e1.
9. Beck AF, Huang B, Chundur R, Kahn RS. Housing code violation density associated with emergency department and hospital use by children with asthma. Health Aff (Millwood) 2014;33(11):1993–2002.

10. Liu SY, Pearlman DN. Hospital readmissions for childhood asthma: the role of individual and neighborhood factors. Public Health Rep 2009;124(1):65–78.

11. Beck AF, Bradley CL, Huang B, Simmons JM, Heaton PC, Kahn RS. The pharmacy-level asthma medication ratio and population health. Pediatrics 2015;135(6):1009–17.

12. Beck AF, Simmons JM, Huang B, Kahn RS. Geomedicine: area-based socioeconomic measures for assessing risk of hospital reutilization among children admitted for asthma. Am J Public Health 2012;102(12):2308–14.

13. Garcia E, Serban N, Swann J, Fitzpatrick A. The effect of geographic access on severe health outcomes for pediatric asthma. J Allergy Clin Immunol 2015;136(3):610–8.

14. Beck AF, Moncrief T, Huang B, Simmons JM, Sauers H, Chen C, et al. Inequalities in neighborhood child asthma admission rates and underlying community characteristics in one US county. J Pediatr 2013;163(2):574–80.

15. Brewer M, Kimbro RT, Denney JT, Osiecki KM, Moffett B, Lopez K. Does neighborhood social and environmental context impact race/ethnic disparities in childhood asthma? Health Place 2017;44:86–93.

16. Nkoy FL, Stone BL, Knighton AJ, Fassl BA, Johnson JM, Maloney CG, et al. Neighborhood deprivation and childhood asthma outcomes, accounting for insurance coverage. Hosp Pediatr 2018;8(2):hpeds.2017-0032.

17. Saha C, Riner ME, Liu G. Individual and neighborhood-level factors in predicting asthma. Arch Pediatr Adolesc Med 2005;159(8):759–63.

18. US Department of Commerce. Geographic areas reference manual. Washington (DC): US Census Bureau; 1994.

19. Koinis-Mitchell D, McQuaid EL, Seifer R, Kopel SJ, Esteban C, Canino G, et al. Multiple urban and asthma-related risks and their association with asthma morbidity in children. J Pediatr Psychol 2007;32(5):582–95.

20. Alpern ER, Clark AE, Alessandrini EA, Gorelick MH, Kittick M, Stanley RM, et al.; Pediatric Emergency Care Applied Research Network (PECARN). Recurrent and high-frequency use of the emergency department by pediatric patients. Acad Emerg Med 2014;21(4):365–73.

21. Canino G, McQuaid EL, Rand CS. Addressing asthma health disparities: a multilevel challenge. J Allergy Clin Immunol 2009;123(6):1209–17, quiz 1218–9.

22. Kaiser Family Foundation. State health facts: health coverage and uninsured: health insurance coverage of children 0–18. https://www.kff.org/other/state-indicator/children-0-18. Accessed January 31, 2019.

23. US Census Bureau. American Community Survey summary file 2010– 2014. https://www2.census.gov/programs-surveys/acs/summary_file/2014/data. Accessed January 15, 2016.

24. US Census Bureau. 2010Census.https://www2.census.gov/census_2010/04-Summary_File_1/Rhode_Island.Accessed March 9, 2016.

25. Cafri G, Hedeker D, Aarons GA. An introduction and integration of cross-classified, multiple membership, and dynamic group random-effects models. Psychol Methods 2015;20(4):407–21.

26. Travers D, Lich KH, Lippmann SJ, Weinberger M, Yeatts KB, Liao W, et al. Defining emergency department asthma visits for public health surveillance, North Carolina, 2008-2009. Prev Chronic Dis 2014;11:E100.

27. Wilson KM, Torok M, McMillen R, Tanski S, Klein JD, Winickoff JP. Tobacco smoke incursions in multiunit housing. Am J Public Health 2014;104(8):1445–53.

28. Laster N, Holsey CN, Shendell DG, Mccarty FA, Celano M. Barriers to asthma management among urban families: caregiver and child perspectives. J Asthma 2009;46(7):731–9.

29. Busacker A, Kasehagen L. Association of residential mobility with child health: an analysis of the 2007 National Survey of Children's Health. Matern Child Health J 2012;16(1,Suppl 1):S78–87.

30. Akinbami LJ, Sullivan SD, Campbell JD, Grundmeier RW, Hartert TV, Lee TA, et al. Asthma outcomes: healthcare utilization and costs. J Allergy Clin Immunol 2012;129(3,Suppl):S49–64.

31. Zingmond DS, Parikh P, Louie R, Lichtensztajn DY, Ponce N, Hasnain-Wynia R, et al. Improving hospital reporting of patient race and ethnicity — approaches to data auditing. Health Serv Res 2015;50(Suppl 1):1372–89.

32. Kind AJH, Buckingham WR. Making neighborhood-disadvantage metrics accessible — the neighborhood atlas. N Engl J Med 2018;378(26):2456–8.

Tables

Table 1. Distribution of Child, Visit, and Neighborhood Characteristics for Pediatric Asthma Emergency Department Visits and Hospitalizations (N = 19,700) by Children Aged 2–17 Years in Rhode Island, 2005–2014[a]

Characteristics	No Asthma Revisit (n = 13,997)	Asthma Revisit (n = 5,703)	P Value
Age, mean (SD), y	8.1 (4.8)	7.4 (4.6)	<.001
Sex			
Male	59.4	59.2	.82
Female	40.6	40.8	
Insurance coverage			
Public	58.7	68.7	<.001
Private	37.5	27.7	
None/self-pay	3.8	3.6	
Season			
Winter	25.5	24.1	.18
Spring	26.5	27.3	
Summer	15.6	15.5	
Autumn	32.4	33.1	
Race/ethnicity			
Hispanic	31.4	38.0	<.001
Non-Hispanic black	15.3	21.5	
Non-Hispanic white	48.5	35.8	
Non-Hispanic other	4.8	4.8	
Neighborhood risk index[b]			
Low (<75th census-block–group percentile)	50.0	40.4	<.001
High (≥75th census-block–group percentile)	50.0	59.6	

[a] All values are percentages, unless otherwise indicated. Data on emergency department visits and hospitalization were collected from a statewide hospital network administrative database. We counted 11,547 children with an index visit occurring during the period from January 1, 2005, through December 31, 2013. After excluding visits that occurred within 7 days of the index visit (n = 860), we counted 19,700 index visits.

[b] Derived by using 8 measures from the 2010–2014 American Community Survey and 2010 US Census: percentage of adults with no high school education, percentage of single-parent households, percentage of household crowding (>1 person per room), percentage of renter-occupied housing units, percentage of vacant homes (excluding vacation homes), percentage of families below 100% of the federal poverty level, percentage of nonwhite residents, and percentage of housing units built before 1950. The index ranged in value from 8 to 40; high risk defined as an index of 30–40; low risk, 8–29.

Table 2. Child-Clustered Adjusted Regression Results for Pediatric Asthma Emergency Department Visits and Hospitalizations (N = 19,700) by Children Aged 2–17 Years in Rhode Island, 2005–2014[a]

Effect	Adjusted Odds Ratio (95% Confidence Interval)	P Value
Age	0.99 (0.98–1.01)	31
Sex		
Male	1 [Reference]	
Female	1.02 (0.93–1.11)	.72
Insurance coverage		
Private	1 [Reference]	
Public	1.20 (1.00–1.44)	.06
None/self-pay	1.31 (0.90–1.89)	.16
Season		
Winter	0.95 (0.79–1.15)	.60
Spring	1.02 (0.86–1.22)	.81
Summer	1.04 (0.84–1.29)	.71
Autumn	1 [Reference]	
Race/ethnicity		
Hispanic	1.24 (1.03–1.50)	.02
Non-Hispanic black	1.26 (1.02–1.56)	.04
Non-Hispanic white	1 [Reference]	
Non-Hispanic other	1.06 (0.75–1.50)	.74
Neighborhood risk index[b]		
Low (<75th percentile)	1 [Reference]	
High (≥75th percentile)	1.22 (1.00–1.48)	.04

[a] Data on emergency department visits and hospitalization were collected from a statewide hospital network administrative database. We counted 11,547 children with an index visit occurring during the period from January 1, 2005, through December 31, 2013. After excluding visits that occurred within 7 days of the index visit (n = 860), we counted 19,700 index visits. Multivariable model controlled for child and neighborhood characteristics.

[b] Derived by using 8 measures from the 2010–2014 American Community Survey and 2010 US Census: percentage of adults with no high school education, percentage of single-parent households, percentage of household crowding (>1 person per room), percentage of renter-occupied housing units, percentage of vacant homes (excluding vacation homes), percentage of families below 100% of the federal poverty level, percentage of nonwhite residents, and percentage of housing units built before 1950. The index ranged in value from 8 to 40; high risk defined as an index of 30–40; low risk, 8–29.

Table 3. Adjusted Cross-Classified Random-Effects Model, Clustered on Child and Neighborhood Risk, for Study on Pediatric Asthma Emergency Department Visits and Hospitalizations by Children Aged 2–17 Years in Rhode Island, 2005–2014[a]

Effect	Adjusted Odds Ratio (95% Confidence Interval)	P Value
Age	0.92 (0.91–0.93)	<.001
Sex		
Male	1 [Reference]	
Female	1.15 (1.06–1.25)	.001
Insurance coverage		
Private	1 [Reference]	
Public	1.10 (1.00–1.21)	.049
None/self-pay	0.94 (0.76–1.17)	.57
Season		
Winter	0.97 (0.88–1.06)	.49
Spring	1.01 (0.92–1.10)	.88
Summer	0.77 (0.69–0.86)	<.001
Autumn	1 [Reference]	
Race/ethnicity		
Hispanic	1.13 (1.02–1.27)	.02
Non-Hispanic black	1.43 (1.26–1.62)	<.001
Non-Hispanic white	1 [Reference]	
Non-Hispanic other	1.30 (1.07–1.58)	.008
Neighborhood risk index[b]		
Low (<75th percentile)	1 [Reference]	
High (≥75th percentile)	1.14 (0.98–1.33)	.08

[a] Data on emergency department visits and hospitalization were collected from a statewide hospital network administrative database. We counted 11,547 children with an index visit occurring during the period from January 1, 2005, through December 31, 2013. After excluding visits that occurred within 7 days of the index visit (n = 860), we counted 19,700 index visits. This cross-clarification model accounted for children who moved from their neighborhoods; after excluding these children, the number of visits was 15,156.

[b] Derived by using 8 measures from the 2010–2014 American Community Survey and 2010 US Census: percentage of adults with no high school education, percentage of single-parent households, percentage of household crowding (>1 person per room), percentage of renter-occupied housing units, percentage of vacant homes (excluding vacation homes), percentage of families below 100% of the federal poverty level, percentage of nonwhite residents, and percentage of housing units built before 1950. The index ranged in value from 8 to 40; high risk defined as an index of 30–40; low risk, 8–29.

ORIGINAL RESEARCH

Residential Racial Isolation and Spatial Patterning of Hypertension in Durham, North Carolina

Mercedes A. Bravo, PhD[1,2]; Bryan C. Batch, MD[3]; Marie Lynn Miranda, PhD[1,2]

PEER REVIEWED

> ### Summary
>
> #### What is already known on this topic?
>
> Neighborhood characteristics are possible underlying causes of health and health disparities among racial/ethnic groups in the United States.
>
> #### What is added by this report?
>
> Few studies have examined relationships between local measures of neighborhood racial isolation and hypertension, a prevalent chronic disease. We identified US Census blocks with the highest overall odds of hypertension in Durham, North Carolina, and estimated cross-sectional associations between racial isolation and odds of hypertension.
>
> #### What are the implications for public health practice?
>
> Information can be used to inform targeted interventions to address risk factors for developing hypertension or manage existing hypertension.

Abstract

Introduction

Neighborhood characteristics such as racial segregation may be associated with hypertension, but studies have not examined these relationships using spatial models appropriate for geographically patterned health outcomes. The objectives of our study were to 1) evaluate the geographic heterogeneity of hypertension; 2) describe whether and how patient-level risk factors and racial isolation relate to geographic heterogeneity in hypertension; and 3) examine cross-sectional associations of hypertension with racial isolation.

Methods

We obtained electronic health records from the Duke Medicine Enterprise Data Warehouse for 2007–2011. We linked patient data with data on racial isolation determined by census block of residence. We constructed a local spatial index of racial isolation for non-Hispanic black patients; the index is scaled from 0 to 1, with 1 indicating complete isolation. We used aspatial and spatial Bayesian models to assess spatial variation in hypertension and estimate associations with racial isolation.

Results

Racial isolation ranged from 0 (no isolation) to 1 (completely isolated). A 0.20-unit increase in racial isolation was associated with 1.06 (95% credible interval, 1.03–1.10) and 1.11 (95% credible interval, 1.07–1.16) increased odds of hypertension among non-Hispanic black and non-Hispanic white patients, respectively. Across Durham, census block-level odds of hypertension ranged from 0.62 to 1.88 among non-Hispanic black patients and from 0.32 to 2.41 among non-Hispanic white patients. Compared with spatial models that included patient age and sex, residual heterogeneity in spatial models that included age, sex, and block-level racial isolation was 33% lower for non-Hispanic black patients and 20% lower for non-Hispanic white patients.

Conclusion

Racial isolation of non-Hispanic black patients was associated with increased odds of hypertension among both non-Hispanic black and non-Hispanic white patients. Further research is needed to identify latent spatially patterned factors contributing to hypertension.

Introduction

Hypertension, a chronic health condition affecting approximately 1 in 3 US adults, increases the risk of myocardial infarction, stroke, heart failure, kidney disease, vision loss, and peripheral artery disease (1,2). In the United States, hypertension is most prevalent among black people (3), a disparity that persists even after adjustment for individual-level risk factors (4). Increasingly, neighborhood characteristics are implicated as possible underlying causes of health disparities observed across racial/ethnic groups. In the United States, place of residence is strongly patterned by race/ethnicity, and a growing body of evidence links

neighborhood environmental characteristics with a range of health outcomes. Nonetheless, only a few studies have examined relationships between neighborhood characteristics and hypertension.

Racial residential segregation is posited to be a fundamental cause of health disparities. Racial residential segregation of black people refers to the degree to which black people live separately from other racial/ethnic groups (5). Through the concentration of poverty and poor physical and social environments, racial residential segregation results in distinctive ecologic environments for black people that may underlie racial health disparities (6). Racial residential segregation is linked with various adverse health outcomes, including type 2 diabetes (7), preterm birth (8), infant mortality (9), and all-cause mortality (10).

Two studies of metropolitan-level segregation and hypertension found that adults residing in more segregated areas were more likely to be hypertensive than those living in less segregated areas (11,12); in one study this association was observed among black people but not white people (12). A study in New York City that used a local measure (as opposed to a city or metropolitan measure) of segregation found that non–US-born black people aged 65 or older residing in highly segregated neighborhoods were less likely to be hypertensive than their counterparts in neighborhoods with low levels of segregation, but this association was not observed among US-born black people aged 65 or older (13). In another study, black–white disparities in the prevalence of hypertension were attenuated in a racially integrated, low-income Baltimore neighborhood, suggesting that exposures associated with neighborhood environment explained some of the racial differences in hypertension observed in nationally representative samples (14). More recently, in 2017, a longitudinal study with follow-up over 25 years found that, among black people, moving from a less segregated to more segregated neighborhood was associated with a rise in systolic blood pressure (15).

Previous work examining cross-sectional associations of local measures of racial residential segregation with hypertension used aspatial statistical models that assumed independence among geographic units used to define a person's living space. Ignoring spatial dependency in a health outcome may lead to underestimation of standard errors, producing narrow confidence intervals and, potentially, incorrect inference (16).

The objectives of our study were to use aspatial and spatial regression techniques to 1) evaluate the geographic heterogeneity of hypertension; 2) describe whether and how patient-level risk factors and racial isolation relate to geographic heterogeneity in hypertension; and 3) examine cross-sectional associations of hypertension with racial isolation.

Methods

We used electronic health records from the Duke University Health System in Durham, North Carolina. We use a local spatial measure of racial isolation that represents 1 dimension of racial residential segregation and helps to overcome the shortfalls of simple measures of racial composition (eg, percentage of black residents) (17). We focused on the racial isolation of non-Hispanic black people because, compared with other dimensions of racial residential segregation (eg, evenness, the differential distribution of a population across a geographic unit), racial isolation may be more closely linked to health by serving as a proxy for the concentration of multiple disadvantages into a single ecological space (18).

The study area consisted of 5,029 census blocks composing Durham County, North Carolina. The Durham County population is 37.5% non-Hispanic black, 42.1% non-Hispanic white, and 13.5% Hispanic (19).

Patient data

We obtained electronic health records from the Duke Medicine Enterprise Data Warehouse for 377,556 unique persons who were patients of a Duke Medicine provider at any time from January 1, 2007, through December 31, 2011. Using ArcGIS software (Esri), we street-geocoded the residential address of each patient to link patients to a 2010 census block. Of 361,434 patients with valid addresses, 88% were geocoded. We restricted the geocoded data set to patients residing in Durham County (remaining n = 243,837) and removed data on patients whose records consisted only of laboratory test results (remaining n = 243,820). We excluded patients younger than 18 years or with missing information on age, race/ethnicity, or sex (remaining n = 171,520). We further restricted our analysis to patients who were either non-Hispanic black or non-Hispanic white (remaining n = 147,359) and resided in census blocks with a nonzero population (remaining n = 147,351).

Patients were defined as having hypertension on the basis of the following *International Classification of Diseases, Ninth Revision, Clinical Modification* (ICD-9) codes: 401.0, 401.9, 402.00, 402.01, 402.10, 402.11, 402.91, 403.00, 403.01, 403.10, 403.11, 404.00–404.03, 404.10–404.13, 404.90–404.93, 405.01, 405.09, 405.11, 405.19, 405.91, 405.99, and 437.2. We constructed a hypertensive status indicator equal to 1 if a patient ever received a positive diagnosis during the study period and 0 otherwise. We constructed maps to show, by quintile, the proportion of patients with a hypertension diagnosis during the study period in each census block.

This research was approved by the institutional review boards at Duke University and Rice University.

Racial isolation

Using 2010 census data, we calculated block-level racial isolation scores by accounting for the population composition in the index block along with adjacent blocks. We thus included neighboring blocks in surrounding counties in our adjacency structure.

The local spatial measure of racial isolation, described in detail elsewhere (8), ranges from 0 to 1: 0 indicates that the neighborhood environment is 100% non-black (no isolation), and 1 indicates that it is 100% black (complete isolation). We linked information on block-level racial isolation with patient data based on each patient's block of residence.

Statistical analysis

We computed descriptive statistics for the study sample. To inform the use of race-stratified models, we evaluated whether the racial isolation exposure distributions of non-Hispanic black patients and non-Hispanic white patients overlapped. To some degree, both populations had nonoverlapping neighborhoods (Appendix); that is, non-Hispanic black and non-Hispanic white patients tended to reside in different blocks and have different residential environments. Consequently, we chose to proceed with a race-stratified modeling approach.

Geographic heterogeneity of hypertension

We evaluated the geographic heterogeneity of a patient-level hypertension diagnosis by comparing 4 patient-level logistic regression models with the following: 1) no random effects (standard model); 2) unstructured block-level random effects only (v_j in Equation 1, the random-intercept model); 3) spatially structured block-level random effects only (υ_j in Equation 1, the spatially structured model); and 4) both structured and unstructured block-level random effects ($\upsilon_j + v_j$ in Equation 1, the convolution model). Thus, the convolution model was of the following form:

$$log\left(\frac{\hat{p}_{ij}}{1-\hat{p}_{ij}}\right) = \hat{\beta}_0 + \hat{\beta}_1 x_{ij} + \hat{\beta}_2 z_j + v_j + \upsilon_j \text{ (Equation 1)}$$

where $\hat{p}_{ij}$ is the fitted probability of patient i in block j having hypertension, x_{ij} is a vector of individual-level covariates (eg, age, sex) for patient i in block j; z_j is a block-level covariate for block j (eg, racial isolation); and v_j and υ_j are the unstructured and spatially structured block-specific random effects for block j, respectively.

Models with random effects are based on the hypothesis that patients in the same block share sources of unobserved variation in hypertension. The unstructured random effect assumes that blocks are independent across geographic space, whereas the spatially structured random effect assumes that hypertension in blocks nearer to each other is more similar. This term reflects sources of unobserved heterogeneity that vary locally ("clustering"). The unstructured random effects (v_j) are assigned a normal prior with unknown variance.

For the spatially structured block-level random effects (υ_j), we assumed a Besag–York–Mollie specification (20), modeled by using an intrinsic conditional autoregressive (iCAR) structure:

$$\upsilon_j \mid \upsilon_{k \neq j} \sim Normal(m_j, \vartheta_\upsilon^2/{\#N(j)}) \text{ (Equation 2)}$$

$$m_j = \frac{\sum_{k \in N(j)} \upsilon_j}{\#N(j)}$$

where m_j is the mean of the spatial random effects of blocks neighboring block j, and $\#N(j)$ is the number of blocks neighboring block j (21).

The variances of the unstructured and spatially structured random effects represent unknown hyperparameters, with priors for the precision taken from γ distributions with shape and scale equal to 1 and 0.0005, respectively. For all models, we assigned vague normal (0, 1000) priors to the parameters for patient risk factors and racial isolation.

We fit 3 model specifications, including a null model, a model adjusting for patient-level risk factors for hypertension (age, sex), and a model adjusting for patient-level risk factors and racial isolation. We used these model specifications to examine how residual geographic heterogeneity (ie, the variance of the block-level spatially structured random effects) in hypertension changes after the addition of patient-level risk factors and racial isolation. We conducted model selection by using the deviance information criterion (DIC) (22), with differences in DIC of 5 or less considered not meaningful.

We calculated the percentage change in residual geographic heterogeneity by sequentially comparing the null, patient-level risk factor, and patient-level risk factor plus racial isolation models.

Cross-sectional association of hypertension and racial isolation

We used the racial isolation index of non-Hispanic black patients in both the white and black race-stratified models and tabulated cross-sectional associations per 0.20-unit increase in racial isolation. We selected the regression model that included patient-level risk factors and racial isolation based on the DIC. We then applied a map decomposition technique (23) to explore the relative contribution of racial isolation versus the unstructured and spa-

tially structured random effects to odds of hypertension at the block level. For example, for a given block, the component odds for racial isolation is equal to the exponentiated fixed-effect estimate multiplied by the standardized racial isolation value for that block. This quantity represents the contribution of racial isolation to odds of hypertension for the average patient in the index block. Mapping the component odds enables visualization of the geographic distribution of odds of hypertension and the extent to which local odds may be driven by racial isolation versus unobserved sources reflected in the random effects.

Sensitivity analysis

We compared the Watanabe–Akaike information criterion (24) with the DIC to select our model. In place of the local spatial measure of racial isolation, we examined cross-sectional associations between hypertension and the block-level proportion of non-Hispanic black residents. Cross-sectional associations estimated between racial isolation and hypertension may be subject to confounding from factors for which we did not adjust, such as individual-level socioeconomic status (SES), which others have proxied by using insurance status. If individual-level SES acts as a confounder, not controlling for it may have biased the association estimated between racial isolation and hypertension. To explore this possibility, we used insurance status (private vs nonprivate) as a proxy for individual-level SES, then restricted the analysis to patients who were not missing information on insurance status, and fit race-stratified models with and without insurance as a covariate (revised n = 49,113 for non-Hispanic black patients and n = 52,556 for non-Hispanic white patients). Lastly, we compared cross-sectional associations for racial isolation (odds ratios and 95% credible intervals) from the model selected based on DIC with the remaining 3 models to investigate whether inference was sensitive to model assumptions.

All statistical analyses were performed by using R version 3.4.4 (The R Foundation). Models were fit by using integrated nested Laplace approximation (25).

Results

Patients resided in 3,439 (68%) blocks in the study area. More than half (56%) were non-Hispanic white. Approximately 38% of non-Hispanic black patients and 27% of non-Hispanic white patients had hypertension (Table 1 and Figure 1). In the study area, block-level racial isolation ranged from 0 (no isolation) to 1 (completely isolated), with a mean (standard deviation) of 0.35 (0.21) and median of 0.28 (Figure 2). The mean (standard deviation) racial isolation was 0.54 (0.23) among non-Hispanic black patients and 0.24 (0.17) among non-Hispanic white patients.
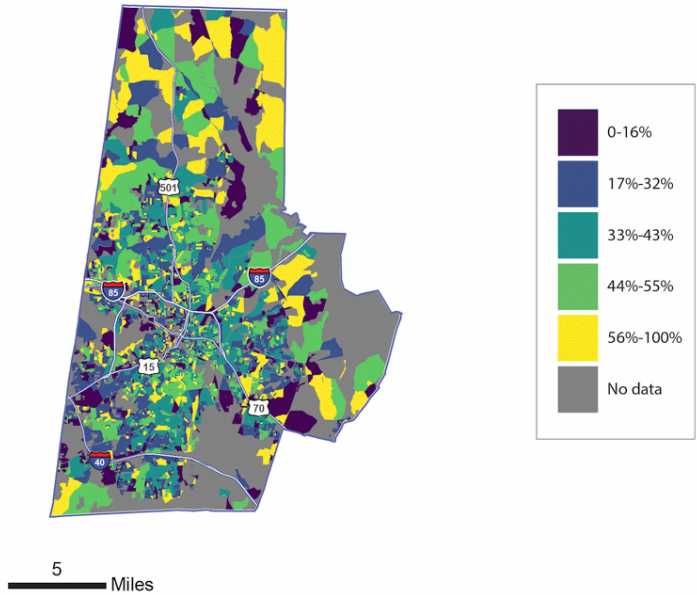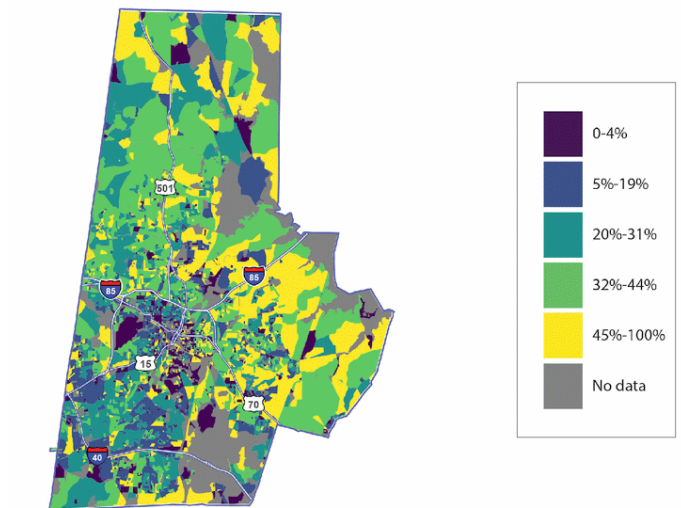
Figure 1. Proportion of patients with hypertension in 2010 Census blocks, by quintile, Duke Medicine Enterprise Data Warehouse patient data, Durham, North Carolina. A, Non-Hispanic black patients. B, Non-Hispanic white patients.
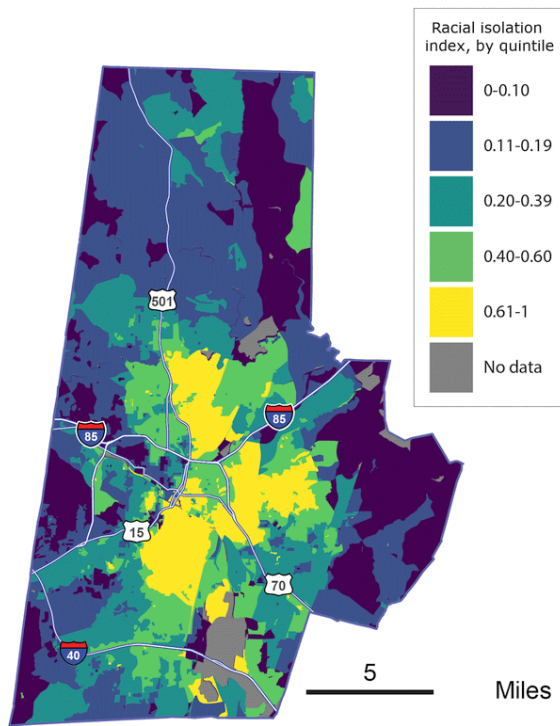
Figure 2. Index value, by quintile, for census-block–level racial isolation of non-Hispanic black residents, Durham, North Carolina. Index of racial isolation is scaled from 0 to 1, with 1 indicating complete isolation.

Model choice

For both non-Hispanic black and non-Hispanic white patients, DIC analysis indicated that the spatially structured and convolu- tion models were indistinguishable from one another (difference in DIC ≤5) but preferred over the standard and random-intercept models (Table 2). We chose the spatially structured model over the convolution model for ease of interpretation.

Non-Hispanic black patients

Among non-Hispanic black patients, a 0.20 increase in racial isol-ation was associated with 1.06 (95% credible interval, 1.03–1.10) higher odds of hypertension in the spatially structured model after adjusting for patient age and sex. In the null model, the residual geographic heterogeneity (residual variation on the binomial scale associated with the spatially structured random effect) was approx-imately 0.36. With the addition of patient age and sex, heterogen-eity decreased by 83%, to 0.06. Inclusion of racial isolation fur-ther decreased heterogeneity by 33%, to 0.04.

Overall, block-level odds of hypertension among non-Hispanic black patients ranged from 0.62 to 1.88 (Appendix). The blocks

with the greatest contributions to hypertension from racial isola-tion corresponded to areas with higher racial isolation values in central and south central Durham. The magnitude of the associ-ation with racial isolation and the width of 95% credible intervals was similar across models (Table 3), although credible intervals in spatial models were wider than those in the aspatial (standard and random intercept) models.

Non-Hispanic white patients

Among non-Hispanic white patients, a 0.20 increase in racial isol-ation of non-Hispanic black patients was associated with 1.11 (95% credible interval, 1.07–1.16) higher odds of hypertension, after adjusting for patient age and sex. Residual geographic hetero-geneity in the null model for non-Hispanic white patients was ap-proximately 0.59. The addition of patient age and sex to the mod-el decreased residual heterogeneity by 66%, to 0.20; the sub-sequent addition of racial isolation decreased residual heterogen-eity by 20%, to 0.16.

Overall, block-level odds of hypertension among non-Hispanic white patients ranged from 0.32 to 2.41 (Appendix). The mag-nitude of the association with racial isolation was larger in the standard and random-intercept models than in the spatially struc-tured and convolution models (Table 3). The 95% credible inter-vals were also wider in the spatially structured and convolution models than in the aspatial models.

Sensitivity analysis

Using the Watanabe–Akaike information criterion instead of DIC would not have resulted in selection of different models (Ap-pendix). The cross-sectional associations between hypertension and block-level proportion of non-Hispanic black residents was smaller than, but not significantly different from, the cross-sec-tional associations between hypertension and racial isolation. In race-stratified models with and without data on health insurance status, 95% credible intervals for cross-sectional associations for non-Hispanic black and non-Hispanic white patients in the spa-tially structured model overlapped with those reported in the main analysis (Appendix).

Discussion

An underlying spatially patterned phenomenon fully characterized residual geographic heterogeneity in hypertension among both non-Hispanic black and non-Hispanic white patients in Durham, North Carolina. Block-level odds of hypertension were more var-ied among non-Hispanic white patients than non-Hispanic black patients. Patient age and sex accounted for a larger proportion of residual heterogeneity among non-Hispanic black patients than non-Hispanic white patients, whereas the inclusion of racial isola-

tion more similarly proportionately reduced residual geographic heterogeneity among both non-Hispanic black and non-Hispanic white patients. The cross sectional association estimated between racial isolation and hypertension for non-Hispanic white patients was larger than that estimated for non-Hispanic black patients. Furthermore, for non-Hispanic white patients, the cross-sectional associations from aspatial models were larger than those from spatial models. Aspatial models also produced narrower credible intervals than did spatial models.

To date, spatial methods have not been used to study associations between racial isolation and hypertension. We found that non-Hispanic black and non-Hispanic white patients in Durham have, on average, distinct residential contexts, which may lead to separate neighborhood risk factors for hypertension. The exclusive role of the spatially structured random effect in unobserved geographic heterogeneity suggests the presence of local environmental risk factors whose effects on hypertension spill over census-block boundaries.

The larger range in overall block-level odds of hypertension among non-Hispanic white patients than non-Hispanic black patients may indicate underlying differences in race-specific study samples. Non-Hispanic white residents are more spread out than non-Hispanic black residents across Durham, creating more widely varying neighborhood environments for non-Hispanic white residents.

The racial isolation index used in our study measured the geographic separation of black people from other racial/ethnic groups. Non-Hispanic white patients residing in blocks with high values of racial isolation lived in predominantly black neighborhoods and may have greater exposure to neighborhood conditions associated with higher rates of hypertension (eg, unhealthy food environments, poor access to health care). In contrast, when non-Hispanic white patients lived in blocks with low values for racial isolation (which means predominantly white neighborhoods given our definition of racial isolation), they may benefit from health-promoting neighborhood conditions. Non-Hispanic white people who are subject to the same census-block conditions (ie, blocks with high levels of racial isolation) as non-Hispanic black people may be worse off than other non-Hispanic white people because they do not reap neighborhood benefits that provide a health advantage to most other non-Hispanic white people, a premise supported by the findings of others (14). However, the contribution of the spatially structured random effect to overall odds of hypertension suggests that we may not have accounted for other spatially patterned characteristics (eg, healthy food availability) (26).

Our study has several limitations. One is the cross-sectional study design, which precludes causal inference. Second, although we used ICD-9 codes to identify patients with hypertension, we may not have captured data on all patients with hypertension in our study sample. Third, the association observed between racial isolation and hypertension may have been subject to confounding from factors for which we did not control. In the sensitivity analysis, we used insurance status as a proxy for individual-level SES. We observed that inclusion of insurance status, which was missing for approximately 31% of the sample, did not result in significantly different estimated associations between racial isolation and hypertension. Another limitation relates to the study sample's representativeness of Durham County's population and the generalizability of results. During the study period, approximately 84% of Durham County residents received care from a Duke Medicine provider at least once, but the study sample excluded patients with residential addresses that could not be found or matched in a reference address data set. The nongeocodable patients, who were removed from analysis, may systematically differ from geocodable patients, who were included in the analysis, in characteristics affecting exposure or health or both.

Despite these limitations, our study enriches the existing body of research on links between racial residential segregation and health, specifically hypertension. Researchers have observed associations between racial residential segregation and health (18,27), but only a few studies have examined segregation and hypertension. Of those that have, most were cross-sectional studies that relied on metropolitan-level measures of segregation or used exclusively aspatial models. For spatially dependent health outcomes, a spatial modeling approach yields more conservative inference; significance in the spatial model, with potentially inflated variances, should also imply significance in the nonspatial model (16,28,29). Furthermore, the local spatial measure of block-level racial isolation may be more closely linked than segregation measures estimated at the metropolitan or city level to individual health outcomes because it is a proxy for the concentration of multiple disadvantages into a single, local ecologic space (6).

Spatial analysis provides an innovative mechanism for evaluating the extent to which residual geographic patterning persists after adjusting for variables that may relate to hypertension, some of which may cluster spatially. Here, we identified blocks and areas of Durham County in which spatially correlated latent risk factors other than racial isolation may be associated with hypertension. In blocks with other neighborhood-based spatially patterned risk factors that contribute to hypertension, additional research is needed to identify what these additional neighborhood characteristics are and how they might be addressed to reduce hypertension. We also identified blocks in Durham with the greatest overall odds of hypertension, which can be used to inform targeted interventions to reduce hypertension risk or manage chronic hypertension.

Author Information

Corresponding Author: Mercedes A. Bravo, PhD, Children's Environmental Health Initiative, Biosciences Research Collaborative, 6500 S Main St, Houston, TX 77030. Telephone: 919-368-0434. Email: mercedes.a.bravo@rice.edu.

Author Affiliations: [1]Department of Statistics, Rice University, Houston, Texas. [2]Children's Environmental Health Initiative, Rice University, Houston, Texas. [3]Department of Medicine, Endocrinology, Metabolism, and Nutrition, Duke University School of Medicine, Durham, North Carolina.

References

1. American Heart Association. Health threats from high blood pressure. 2018. http://www.heart.org/HEARTORG/Conditions/HighBloodPressure/LearnHowHBPHarmsYourHealth/Health-Threats-From-High-Blood-Pressure_UCM_002051_Article.jsp. Accessed July 12, 2018.
2. Centers for Disease Control and Prevention. High blood pressure frequently asked questions (FAQs). 2016. https://www.cdc.gov/bloodpressure/faqs.htm. Accessed July 12, 2018.
3. Hajjar I, Kotchen TA. Trends in prevalence, awareness, treatment, and control of hypertension in the United States, 1988–2000. JAMA 2003;290(2):199–206.
4. Mensah GA, Mokdad AH, Ford ES, Greenlund KJ, Croft JB. State of disparities in cardiovascular health in the United States. Circulation 2005;111(10):1233–41.
5. Massey DS, Denton NA. The dimensions of residential segregation. Soc Forces 1988;67(2):281–315.
6. Williams DR, Collins C. Racial residential segregation: a fundamental cause of racial disparities in health. Public Health Rep 2001;116(5):404–16.
7. Bravo MA, Anthopolos R, Kimbro RT, Miranda ML. Residential racial isolation and spatial patterning of type 2 diabetes mellitus in Durham, North Carolina. Am J Epidemiol 2018;187(7):1467–76.
8. Anthopolos R, James SA, Gelfand AE, Miranda ML. A spatial measure of neighborhood level racial isolation applied to low birthweight, preterm birth, and birthweight in North Carolina. Spat Spatio-Temporal Epidemiol 2011;2(4):235–46.
9. Hearst MO, Oakes JM, Johnson PJ. The effect of racial residential segregation on black infant mortality. Am J Epidemiol 2008;168(11):1247–54.
10. Jackson SA, Anderson RT, Johnson NJ, Sorlie PD. The relation of residential segregation to all-cause mortality: a study in black and white. Am J Public Health 2000;90(4):615–7.
11. Jones A. Segregation and cardiovascular illness: the role of individual and metropolitan socioeconomic status. Health Place 2013;22:56–67.
12. Kershaw KN, Diez Roux AV, Burgard SA, Lisabeth LD, Mujahid MS, Schulz AJ. Metropolitan-level racial residential segregation and black-white disparities in hypertension. Am J Epidemiol 2011;174(5):537–45.
13. White K, Borrell LN, Wong DW, Galea S, Ogedegbe G, Glymour MM. Racial/ethnic residential segregation and self-reported hypertension among US- and foreign-born blacks in New York City. Am J Hypertens 2011;24(8):904–10.
14. Thorpe RJJ Jr, Brandon DT, LaVeist TA. Social context as an explanation for race disparities in hypertension: findings from the Exploring Health Disparities in Integrated Communities (EHDIC) Study. Soc Sci Med 2008;67(10):1604–11.
15. Kershaw KN, Robinson WR, Gordon-Larsen P, Hicken MT, Goff DC Jr, Carnethon MR, et al. Association of changes in neighborhood-level racial residential segregation with changes in blood pressure among black adults: The CARDIA study. JAMA Intern Med 2017;177(7):996–1002.
16. Waller LA, Gotway CA. Applied spatial statistics for public health data. Hoboken (NJ): John Wiley and Sons; 2004.
17. Kershaw KN, Albrecht SS. Racial/ethnic residential segregation and cardiovascular disease risk. Curr Cardiovasc Risk Rep 2015;9(3):10.
18. Acevedo-Garcia D, Lochner KA, Osypuk TL, Subramanian SV. Future directions in residential segregation and health research: a multilevel approach. Am J Public Health 2003;93(2):215–21.
19. US Census Bureau. American factfinder factsheet: summary file 1 100% data, Durham County, North Carolina, and United States. 2010. http://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=DEC_10_SF1_P9&prodType=table. Accessed June 14, 2016.
20. Besag J, York J, Mollie A. Bayesian image restoration, with two applications in spatial statistics (with discussion). Ann Inst Stat Math 1991;43(1):1–59.

21. Banerjee S, Carlin B, Gelfand A. Hierarchical modeling and analysis for spatial data. New York (NY): Chapman and Hall; 2004.

22. Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit. J R Stat Soc Series B Stat Methodol 2002;64(4):583–639.

23. Law J, Haining R. A Bayesian approach to modeling binary data: the case of high-intensity crime areas. Geogr Anal 2004; 36(3):197–216.

24. Watanabe S. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. J Mach Learn Res 2010;11(2010):3571–94.

25. Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J R Stat Soc Series B Stat Methodol 2009; 71(2):319–92.

26. Kaiser P, Diez Roux AV, Mujahid M, Carnethon M, Bertoni A, Adar SD, et al. Neighborhood environments and incident hypertension in the Multi-Ethnic Study of Atherosclerosis. Am J Epidemiol 2016;183(11):988–97.

27. Kramer MR, Hogue CR. Is segregation bad for your health? Epidemiol Rev 2009;31(1):178–94.

28. Reich BJ, Hodges JS, Zadnik V. Effects of residual smoothing on the posterior of the fixed effects in disease-mapping models. Biometrics 2006;62(4):1197–206.

29. Hanks E, Schliep E, Hooten M, Hoeting J. Restricted spatial regression in practice: geostatistical models, confounding, and robustness under model misspecication. Environmetrics 2015; 26(4):243–54.

Tables

Table 1. Summary Statistics of Patient Characteristics in the Duke Medicine Enterprise Data Warehouse Electronic Health Records (n = 147,351), Durham, North Carolina, 2007–2011[a]

Characteristic	Non-Hispanic Black, No. (%)	Non-Hispanic White, No. (%)
Total	65,026 (44.1)	82,325 (55.9)
Hypertension	24,517 (37.7)	21,836 (26.5)
Male	26,157 (40.2)	35,183 (42.7)
Age, y		
18–21	6,473 (10.0)	4,205 (5.1)
22–29	10,962 (16.9)	14,680 (18.1)
30–39	12,360 (19.0)	15,392 (18.7)
40–49	12,590 (19.4)	12,436 (15.1)
50–64	14,557 (22.4)	19,626 (23.8)
≥65	8,084 (12.4)	15,986 (19.4)
Racial isolation, percentile[a]		
<20th	2,424 (3.7)	27,001 (32.8)
20th–39th	5,952 (9.2)	23,566 (28.6)
40th–59th	11,613 (17.9)	17,638 (21.4)
60th–79th	18,871 (29.0)	10,893 (13.2)
≥80th	26,166 (31.8)	3,227 (5.0)

[a] The racial isolation index ranges from 0 to 1. In the 3,439 blocks with ≥1 patient in the analysis data set, the 20th, 40th, 60th, and 80th percentiles of racial isolation correspond to racial isolation values of 0.11, 0.21, 0.37, and 0.63, respectively. Data on racial isolation determined by 2010 census block of residence.

Table 2. Comparison of Standard, Random Intercept, Spatially Structured, and Convolution Race-Stratified Logistic Regression Models, Study of Racial Isolation and Spatial Patterning of Hypertension in Durham, North Carolina, 2007–2011[a]

Race	Deviance Information Criterion[b]			
	Standard Model	Random Intercept	Spatially Structured Model	Convolution Model
Non-Hispanic black	63,419	63,327	63,279[c]	63,276[d]
Non-Hispanic white	69,255	68,714	68,417[c,d]	68,417[d]

[a] All models were adjusted for individual-level patient age and sex and block-level racial isolation of non-Hispanic black patients. Patient data obtained from electronic health records in the Duke Medicine Enterprise Data Warehouse for 2007–2011. Data on racial isolation determined by 2010 census block of residence.

[b] The deviance information criterion is a generalization of the Akaike information criterion. Taking into account both model fit and model complexity, smaller values indicate a preferred model (22). Using the Watanabe–Akaike information criterion produced the same preferred models.

[c] The selected model.

[d] Model with the lowest deviance information criterion value across row.

Table 3. Odds Ratios (95% Credible Interval) for Hypertension per 0.20-Unit Increase in Racial Isolation, in Race-Stratified Logistic Regression Models, Study of Racial Isolation and Spatial Patterning of Hypertension in Durham, North Carolina, 2007–2011[a]

Race	Standard Model	Random Intercept Model	Spatially Structured Model	Convolution Model
Non-Hispanic black	1.09 (1.07–1.11)	1.08 (1.06–1.10)	1.06 (1.03–1.10)	1.07 (1.03–1.10)
Non-Hispanic white	1.19 (1.17–1.22)	1.19 (1.17–1.23)	1.11 (1.07–1.16)	1.11 (1.07–1.16)

[a] The standard deviation of racial isolation was 0.17 for non-Hispanic white patients and 0.23 for non-Hispanic black patients. For purposes of comparison, odds ratios are presented per 0.20 racial isolation units. Patient data obtained from electronic health records in the Duke Medicine Enterprise Data Warehouse for 2007–2011. Data on racial isolation determined by 2010 US census block of residence.

Appendix. Supplemental Information

Details on the calculation of the racial isolation index are provided in the Appendix, which is available at https://rice.box.com/v/BravoetalSupplementalMaterial. Also included in the Appendix are sensitivity analysis results, including model selection results using the Watanabe–Akaike information criterion, cross-sectional associations estimated between racial isolation and hypertension for individuals after controlling for patient-level insurance status in the selected model. The map decompositions show the overall odds of hypertension in addition to the contribution of racial isolation and the spatially structured block-level random effect to overall odds of hypertension for the average non-Hispanic black patient and the average non-Hispanic white patient in each block.

ORIGINAL RESEARCH

Modeling the Importance of Within- and Between-County Effects in an Ecological Study of the Association Between Social Capital and Mental Distress

Tse-Chuan Yang, PhD[1]; Stephen A. Matthews, PhD[2]; Feinuo Sun[1]; Marina Armendariz[3]

PEER REVIEWED

Summary

What is already known on this topic?

At the individual level, high levels of social capital are associated with low levels of mental distress.

What is added by this report?

We used ecological data to demonstrate that social capital and prevalence of mental distress are spatially clustered in US counties. We showed that social capital decreases the prevalence of mental distress within a county, but this within-county association is weaker than the between-county association.

What are the implications for public health practice?

Our results suggest that policy interventions to promote population-level mental health should consider broader multi-county contexts and the coordination of actions within the consortia of neighboring counties.

Abstract

Introduction

Levels of mental distress in the United States are a health policy concern. The association between social capital and mental distress is well documented, but evidence comes primarily from individual-level studies. Our objective was to examine this association at the county level with advanced spatial econometric methods and to explore the importance of between-county effects.

Methods

We used County Health Rankings and Roadmaps data for 3,106 counties of the contiguous United States. We used spatial Durbin modeling to assess the direct (within a county) and indirect (between neighboring counties) effects of social capital on mental distress. We also examined the spatial spillover effects from neighboring counties based on higher-order spatial weights matrices.

Results

Counties with the highest prevalence of mental distress were found in regional clusters where levels of social capital were low, including the Black Belt, central/southern Appalachia, on the Mississippi River, and around some Indian Reservations. Most of the association between social capital and mental distress was indirect, from the neighboring counties, although significant direct effects showed the within-county association. Models also confirmed the importance of county-level socioeconomic status.

Conclusion

We found that county social capital is negatively related to mental distress. Counties are not isolated places and are often part of wider labor and housing markets, so understanding spatial dependencies is important in addressing population-level mental distress.

Introduction

Mental distress refers to a range of symptoms and experiences that cause problems in the way individuals think, feel, or behave (1). Levels of mental stress in the United States have increased since the 1990s (2). Social capital reflects the potential resources, both tangible and invisible, embedded in social relations or networks. Social capital can be measured at both individual and ecological levels (3), and its association with mental distress has been well documented (4–6). However, most evidence comes from individu-

al-level studies, and the ecological findings are inconclusive due to diverse analytic units and methods used (4,6).

We identified 2 major gaps in our understanding of how social capital is associated with mental distress at the ecological level. First, research overlooks the spatial dependencies between these 2 measures by simplifying the contribution of neighboring areas into a single parameter estimate, which is also sometimes misinterpreted (7). Second, whether mental distress of a focal area is affected by the social capital of neighboring areas remains underexplored, and little is known about the importance of distance — as measured by spatial order (ie, n^{th} spatial lags) — in explaining spillover effects of social capital on mental distress.

To address these gaps, we applied spatial Durbin modeling approaches to the County Health Rankings and Roadmaps (CHRR) data set (8) for the United States (9). We examined 2 hypotheses: 1) within a county, higher levels of social capital are associated with lower prevalence of mental distress, even after controlling for other confounders; and 2) the levels of mental distress of a county are negatively influenced by the low social capital of neighboring counties, and the effect is strongest from the neighboring adjacent counties than from counties farther away.

Methods

Data and measures

The 2018 CHRR synthesizes both health and socioeconomic information from national data sets, such as the Behavioral Risk Factor Surveillance System (BRFSS), the Dartmouth Atlas of Health Care, and the American Community Survey. Although the 2018 CHRR covers all US counties, we focused on the counties in the contiguous United States (N = 3,106). All data were publicly available, so no institutional review board approval was needed.

The dependent variable was county-level mental distress. Mental health considers stress, depression, and problems with emotions, and this measure emphasizes those residents with more chronic and severe mental health issues (8). In the BRFSS it was measured with "frequent mental distress," which is the percentage of adults who reported more than 14 days in response to the question, "How many days during the past 30 days was your mental health not good?" For the counties with limited data, the entire BRFSS sample and census population estimates were used to estimate this variable (8).

The key independent variable was the social capital index developed by Rupasingha and colleagues (10), which was created by applying principal component analysis (PCA) to 4 variables: number of establishments per 10,000 population, voter turnout, census response rate, and number of nonprofit organizations. Higher so-

cial capital index values refer to stronger social connections among residents. This social capital index has been used in county level analysis (11,12), but its application to mental health is limited.

We also considered other covariates. The socioeconomic status (SES) index is a PCA-derived score using percentage of population older than 25 who have at least some college education (factor loading = 0.805), unemployment rate (factor loading = −0.752), child poverty rate (factor loading = −0.935), and logged median household income (factor loading = 0.900). Approximately 72% of variation among these 4 variables can be captured with a single factor; a higher SES index score indicates higher socioeconomic status.

Several variables reflect the demographic composition of a county. Age was measured with percentage of population younger than 18 and percentage of population older than 65. Racial/ethnic composition was based on percentage of non-Hispanic blacks, non-Hispanic Asians, and Hispanics. We included percentage population that was female, not proficient in English, and living in a rural area, and we included the ratio of household income at the 80th and 20th percentiles. We checked the variance inflation factors among the independent variables and found that all were smaller than 4, indicating that multicollinearity was not a concern.

Analytic approach

To test our hypotheses, we used the spatial Durbin model, developed in spatial econometrics but rarely used in health research. A spatial Durbin model can be expressed as follows (7,13):

$$(I_n - \rho W)y = \alpha l_n + X\beta + WX\theta + \varepsilon$$

where both the spatially lagged dependent (ρWy) and independent variables ($WX\theta$) are included (9). The endogeneity in the model makes the interpretations of the estimates richer (7). Explicitly, the spatial Durbin model allows researchers to separate the direct (within a county) and indirect (to/from neighboring counties) effects of an independent variable on the dependent variable. The equation above can be rewritten:

$$y = (I_n - \rho W)^{-1}\alpha l_n + (I_n - \rho W)^{-1}X\beta + (I_n - \rho W)^{-1}WX\theta + (I_n - \rho W)^{-1}\varepsilon$$

The partial derivatives of y with respect to the rth independent variable (X_r) across the n observations in the study region can be expressed as follows:

$$\partial y/\partial X_r = (I_n - \rho W)^{-1}(I_n\beta_r + W\theta_r)$$

where $\partial y/\partial X_r$ indicates an n x n matrix, and β_r and θ_r represent the parameter estimates associated with the independent variable in a

county and in neighboring counties. Several implications from this equation highlight the benefit of a spatial Durbin approach (7,13,14). Specific to this study, the third equation indicates that the change in a county's social capital index will not only lead to the change in mental distress in the same county, but also influence the frequent mental distress in other counties. The former refers to the direct effects [average of the main diagonal elements of $(I_n - \rho W)^{-1}(I_n \beta_r + W \theta_r)$ matrix], whereas the latter indicates the indirect effects (average of the off-diagonal elements). Furthermore, the partial derivatives of y are a function of $(I_n - \rho W)^{-1}$ and can be expanded as a linear combination of powers of the spatial weights matrix (W): $I_n + \rho W + \rho^2 W^2 + \rho^3 W^3 + \dots$. The powers of W correspond to the counties themselves (zero-order), adjacent neighbors (first-order), neighbors of adjacent neighbors (second-order), and so on. It is possible to partition both the direct and indirect impacts of social capital on mental distress by using the powers of spatial weight matrix. Consequently, researchers can generate a "spatial profile" of the importance of neighboring areas with the partitioning results. In this way we can test the second hypothesis.

Although other forms of spatial econometrics models handle spatial association (eg, spatial lag and spatial error), the spatial Durbin model is the most appropriate spatial regression form, particularly when the generating process underlying the observed data is unknown (7). Both spatial and aspatial exploratory data analysis were done before the spatial Durbin and partitioning analysis. The Markov chain Monte Carlo method was used to calculate the direct and indirect effects and the partitioning results. All analyses were conducted with the *spdep* package (15) in R (16). Comparisons between the spatial Durbin model and other conventional spatial models are available on request.

Results

On average, 12% of the adult population aged 18 to 85 in a county reported more than 14 days of mental distress in the past 30 days (Table 1). The relatively small standard deviation of mental distress (1.88) suggests that in most counties at least 8.5% of adults aged 18–85 reported mental distress. The social capital index had a mean value of 0 and a standard deviation of 1.26.

Counties with high prevalence of mental distress (4th and 5th quintiles) were concentrated in the South (particularly the Black Belt), central and southern Appalachia, and the Mississippi River Valley through Oklahoma (Figure). Clusters of high levels of mental distress were also found in Indian Reservations (eg, the Four Corners and several counties in the Dakotas).
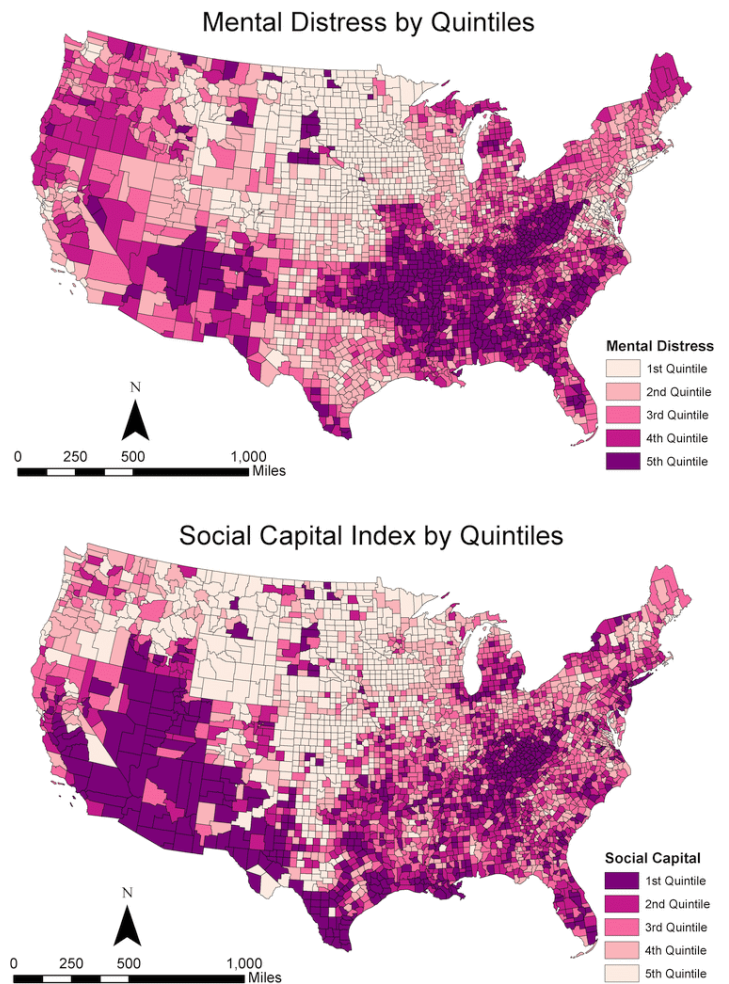
Figure. Spatial distribution of mental distress and social capital in the United States, County Health Rankings and Roadmaps, United States, 2018. Map A depicts county-level prevalence of mental distress, and Map B depicts the social capital index of counties.

The spatial distribution of social capital was the opposite of the spatial distribution of mental distress. Many counties with high prevalence of mental distress had low social capital. Counties with high social capital and low mental distress were clustered in the Great Plains and the Midwest. Overall, spatial analyses indicated that mental distress and social capital are negatively associated.

The spatial Durbin modeling results are shown in Table 2. The direct effect of social capital (−0.087) was smaller than the indirect effect. A one-unit increase in the average social capital index in neighboring counties was associated with a 0.234 percentage point decrease in mental distress of a focal county. The indirect effect of

social capital on mental distress was roughly 2.7 times (−0.234 divided by −0.087 = 2.69) stronger than the direct effect.

SES index had the strongest total effect on mental distress (Table 2). A one-unit increase in SES index was associated with a 1.263 percentage point decrease in the prevalence of mental distress within a county. High SES in neighboring counties was associated with decreased mental distress (percentage point decrease, 0.532). The percentage of female population was associated with spatial variation in mental distress. For every one percentage point increase in female population, mental distress increased by 0.114. Furthermore, a higher income ratio (ie, higher income inequality) was associated with a higher mental distress level within a county; however, the indirect effect of income ratio was negatively associated with mental distress (−0.472).

Results of how the effect of social capital and other covariates on mental distress are transited through neighboring counties are presented in Table 3. The direct effect of social capital at the zero-order (W_0) was −0.072, indicating that almost 83% (−0.072 divided by −0.087 = 0.828) of the direct effect came from a county itself and the other 17% could be attributed to the inter-county dependencies. The immediate neighbors (W_1) appeared not to matter, but the second-order neighboring counties contributed to the direct effect of social capital. The contribution of the third-order (W_3) neighbors became much smaller, yet remained significant.

Estimates of indirect effect of social capital decreased from the first-order to the higher orders. More than 25% (−0.062 divided by −0.234 = 0.265) of the indirect effect came from the first-order neighbors, but higher-order neighbors still contributed to the overall indirect effect.

Discussion

We found strong evidence for our first hypothesis, that county social capital is negatively related to mental distress and that this relationship holds even after considering other confounders. The direct effect of social capital on the prevalence of mental distress was negative and significant in the spatial Durbin model. The partitioning results further indicated that more than 80% of the direct effect was within-county and that neighboring counties strengthened the association of social capital with mental distress.

Our study is explicitly ecological, and the findings contribute to the wider literature of social capital and mental distress. How do we understand this relationship? On one hand, our social capital index reflects the potential connections and social ties among residents (3,10). These social relations create trust and reciprocity that can be used to cope with negative emotions, stress, anxiety, and depression (4,6,17). Individuals living in counties with strong

social capital receive better social support than those living in areas with weak social capital. As a result, the prevalence of mental distress decreases with the increase in social capital. Additionally, strong social capital facilitates a community's capacity for action and cooperative social activities (18,19), which produces an environment conducive to economic development and community well-being (20). Although the findings suggesting that social capital buffers against the potential negative impacts of economic and social adversities are inconclusive, our results support the ecological finding that social capital may lower population prevalence of mental distress.

We also hypothesized that the prevalence of mental distress of a county is negatively associated with social capital of neighboring counties and this relationship decreases with distance (ie, increasing spatial lag order). Our results support this hypothesis but also indicate that the spatial spillover effect is complex. The indirect effect of social capital on mental distress was roughly 2.7 times stronger than the direct effect, suggesting a strong spatial spillover effect from neighboring counties. This indicates that counties with low prevalence of mental distress benefit indirectly from the strong social capital of neighboring counties. Our results suggest that the first-order neighbors are the most important contributors. Although other neighbors remain connected, their contributions decline as spatial order increases. The spatial clustering patterns in the Figure highlighted the strong spatial dependence embedded in social capital and mental distress, and the partitioning results showed how the indirect effects work through spatial adjacency.

As the indirect effect is reciprocal, it indicates that a one-unit increase in social capital of a focal county was related to a 0.234 percentage point decrease in mental distress of neighboring counties. The significant indirect effect of social capital confirms that the spatial association between social capital and mental distress was not only a within-county phenomenon, but influenced by inter-county spatial dependencies, which is captured by the exogenous interactions ($(I_n - \rho W)^{-1} WX$) in the model. The nonsignificant zero-order indirect effect of social capital (−0.015) suggests that the indirect effect of social capital may be largely due to spatial spillover associated with the spatial structure (ie, form of W) and the spillover effect from the first-order neighbors was more crucial than from neighboring counties at the higher orders.

Our study has limitations. The results and conclusions may change if the underlying data are aggregated to different geographies, which is a modifiable area unit problem (21,22). As a sensitivity analysis, we included county area in the model and found that county size is not statistically related to mental distress. That is, while large counties may have neighbors that are geographically farther away than smaller counties, our findings are unaltered. Furthermore, our conclusions cannot be generalized to the individual

level (23). There is no consensus on how to measure social capital at the aggregate level, and the social capital index may not fully reflect the complexities of this construct (3,24). Other mental health measures (eg, mental disorders) should be considered to bolster the beneficial effect of social capital on mental health at the county level. Importantly, both mental distress and social capital are not race/ethnicity specific, which limits our understanding of the dynamics between these 2 variables within each racial/ethnic group. This warrants future endeavors to develop race/ethnicity-specific measurements. Moreover, given the programming limitation, our analysis does not weigh the influence of a county on another by total population. Future research should incorporate population size into the spatial weight matrix. Finally, the analysis is cross-sectional, and it is possible that high prevalence of mental distress leads to weak social capital. The causal relationship between these 2 variables needs to be clarified.

The spatial dependencies between social capital and mental distress provide insights. First, the spatial spillover process generates an indirect effect (from neighboring counties) that is stronger than the direct effect (within a county). The spatial dependencies cannot be identified with conventional spatial regression models, which in part may explain why ecological-level evidence is mixed (4). Future ecological work should consider spatial dependencies to address some of the inconsistencies in the social capital literature. Second, related to policy implications, the indirect effect of social capital on mental distress suggests that improving social capital in a certain county will have spillover effects and reduce mental distress in adjacent or nearby counties. In addition, cross-county collaboration to improve social capital and connections (ie, regional interventions) should be considered to maximize the effect of social capital on mental health.

Beyond social capital, we found that SES played a critical role in explaining the prevalence of mental distress. SES has the strongest impact on mental distress within a county, and this indicates that the spatial variation in mental distress may be a consequence of spatial economic inequality (25). The importance of SES has been discussed (2,26,27), and our results confirm this relationship. Income ratio is also important and our finding echoes the literature, suggesting income inequality follows the social relativity theory as neighbors with high income inequality reduce the sense of relative deprivation, which in turn improves population health (28). Moreover, prevalence of mental distress increases with the percentage of female population within (and beyond) a county. This positive association also corresponds to the extant literature (2,29). It is plausible that females encounter unique social and psychological stressors (eg, pregnancy and social role expectations) that affect their mental distress (2,6,29), although men's reluctance to disclose mental distress may also be a factor (30). The importance

of these variables is confirmed in sensitivity analysis where all the independent variables are standardized (results available on request), indicating that social capital and SES are not only statistically significant but also play a substantive role in explaining the spatial variation in mental distress.

In sum, we contribute to the mental health literature in 2 ways. First, we provide robust evidence for the beneficial association between aggregate levels of social capital and mental distress in US counties. Second, our adoption of spatial Durbin models showed the complicated inter-county dependencies and the relationship between these 2 measures. To our knowledge, no prior research has used the spatial Durbin modeling to clarify how levels of mental distress in a county are affected by neighboring counties.

Acknowledgments

Author Information

Corresponding Author: Tse-Chuan Yang, PhD, Department of Sociology, University at Albany, SUNY, 315 AS, 1400 Washington Ave, Albany, NY 12222. Telephone: 518-442-4647. Email: tyang3@albany.edu.

Author Affiliations: [1]Department of Sociology, University at Albany, SUNY, Albany, New York. [2]Department of Sociology and Criminology, Pennsylvania State University, University Park, Pennsylvania. [3]Department of Biobehavioral Health, Pennsylvania State University, University Park, Pennsylvania.

References

1. EthnoMed. What is mental distress? Washington (DC): National Institutes of Health; 2015.
2. Centers for Disease Control and Prevention (CDC). Self-reported frequent mental distress among adults — United States, 1993–2001. MMWR Morb Mortal Wkly Rep 2004; 53(41):963–6.
3. Putnam RD. Bowling alone: America's declining social capital. In: Crothers L, Lockhart C, editors. Culture and politics. New York (NY): Palgrave Macmillan; 2000. p. 223-34.
4. De Silva MJ, McKenzie K, Harpham T, Huttly SR. Social capital and mental illness: a systematic review. J Epidemiol Community Health 2005;59(8):619–27.

5. Song L. Social capital and psychological distress. J Health Soc Behav 2011;52(4):478–92.

6. Almedom AM. Social capital and mental health: an interdisciplinary review of primary evidence. Soc Sci Med 2005;61(5):943–64.

7. LeSage J, Pace RK. Introduction to spatial econometrics. New York (NY): Chapman and Hall/CRC; 2009.

8. County health rankings and roadmaps 2018. Madison (WI): University of Wisconsin Population Health Institute; 2018.

9. Tabb LP, McClure LA, Quick H, Purtle J, Diez Roux AV. Assessing the spatial heterogeneity in overall health across the United States using spatial regression methods: The contribution of health factors and county-level demographics. Health Place 2018;51:68–77.

10. Rupasingha A, Goetz SJ, Freshwater D. The production of social capital in US counties. J Socioeconomics 2006; 35(1):83–101.

11. Yang TC, Jensen L, Haran M. Social capital and human mortality: explaining the rural paradox with county-level mortality data. Rural Sociol 2011;76(3):347–74.

12. Goetz SJ, Davlasheridze M, Han Y. County-level determinants of mental health, 2002–2008. Soc Indic Res 2015; 124(2):657–70.

13. Anselin L. Spatial econometrics: methods and models. Berlin (Germany): Springer Science and Business Media; 2013.

14. Elhorst JP. Applied spatial econometrics: raising the bar. Spat Econ Anal 2010;5(1):9–28.

15. Bivand R, Piras G. Comparing implementations of estimation methods for spatial econometrics. J Stat Softw 2015; 63(18):1–36.

16. R Core Team. R: a language and environment for statistical computing; 2018. https://www.r-project.org/. Accessed April 14, 2019.

17. McKenzie K, Whitley R, Weich S. Social capital and mental health. Br J Psychiatry 2002;181(4):280–3.

18. Glaeser EL, Laibson D, Sacerdote B. An economic approach to social capital. Econ J (Lond) 2002;112(483):F437–58.

19. Paxton P. Is social capital declining in the United States? A multiple indicator assessment. Am J Sociol 1999; 105(1):88–127.

20. Wilson PA. Building social capital: a learning agenda for the twenty-first century. Urban Stud 1997;34(5-6):745–60.

21. Openshaw S. The modifiable areal unit problem. Concepts and techniques in modern geography. Washington (DC): Geo Books; 1984.

22. Fotheringham AS, Wong DW. The modifiable areal unit problem in multivariate statistical analysis. Environ Plan 1991; 23(7):1025–44.

23. Piantadosi S, Byar DP, Green SB. The ecological fallacy. Am J Epidemiol 1988;127(5):893–904.

24. Mitchell CU, LaGory M. Social capital and mental distress in an impoverished community. City Community 2002; 1(2):199–222.

25. Wilkinson RG, Pickett KE. The problems of relative deprivation: why some societies do better than others. Soc Sci Med 2007;65(9):1965–78.

26. Link BG, Phelan J. Social conditions as fundamental causes of disease. J Health Soc Behav 1995;35(Spec No):80–94.

27. Strine TW, Balluz L, Chapman DP, Moriarty DG, Owens M, Mokdad AH. Risk behaviors and healthcare coverage among adults by frequent mental distress status, 2001. Am J Prev Med 2004;26(3):213–6.

28. Yang TC, Noah A, Shoff C. Exploring geographic variation in US mortality rates using a spatial Durbin approach. Popul Space Place 2015;21(1):18–37.

29. Rosenfield S, Mouzon D. Gender and mental health. In: Aneshensel CS, Phelan JC, Bierman A, editors. Handbook of the sociology of mental health. New York (NY): Springer; 2013. p. 277-96.

30. Substance Abuse and Mental Health Services Administration, Center for Mental Health Services. Mental health: culture, race and ethnicity — a supplement to mental health. A report of the Surgeon General. Rockville. Maryland (MD): US Department of Health and Human Services; 2001.

Tables

Table 1. Descriptive Statistics of Variables (N = 3,106), Ecological Study of the Association Between Social Capital and Mental Distress, County Health Rankings and Roadmaps, United States, 2018

Variable	Minimum	Maximum	Mean (Standard Deviation)
Adults reporting mental distress, %	8.03	21.34	12.21 (1.88)
Social capital index	−3.18	21.81	0.00 (1.26)
Socioeconomic status index	−3.80	2.99	0.00 (1.00)
Younger than 18 y, %	5.15	40.79	22.34 (3.40)
Older than 65 y, %	4.63	56.31	18.45 (4.51)
Non-Hispanic black, %	0.00	85.15	9.03 (14.37)
Non-Hispanic Asian, %	0.00	36.50	1.40 (2.41)
Hispanic, %	0.50	96.25	9.33 (13.73)
Not English proficient, %	0.00	32.69	1.75 (2.93)
Female, %	27.80	56.55	49.93 (2.22)
Rural resident, %	0.00	100.00	58.52 (31.44)
80th/20th income ratio	0.00	8.93	4.52 (0.74)

Table 2. Decomposition Estimates of the Direct and Indirect Effects on Percentage of Adults Reporting Mental Distress, Ecological Study of the Association Between Social Capital and Mental Distress, County Health Rankings and Roadmaps, United States, 2018

Variable	Direct Effect	Indirect Effect	Total
Social capital index	−0.087[a]	−0.234[a]	−0.321[a]
Socioeconomic status index	−1.263[a]	−0.532[a]	−1.795[a]
Younger than 18 y, %	0.018[a]	−0.102[a]	−0.084[a]
Older than 65 y, %	−0.075[a]	0.005	−0.070[a]
Non-Hispanic black, %	−0.010[a]	−0.009	−0.020[a]
Non-Hispanic Asian, %	−0.014	0.009	−0.004
Hispanic, %	−0.028[a]	0.005	−0.023[a]
Not English proficient, %	0.006	−0.010	−0.004
Female, %	0.114[a]	0.132[a]	0.247[a]
Rural resident, %	0.001	−0.002	−0.001
80th/20th income ratio	0.315[a]	−0.472[a]	−0.158

[a] Significant at $P < .05$.

Table 3. Spatial Partitioning Results of Direct and Indirect Effects on Percentage of Adults Reporting Mental Distress, Ecological Study of the Association Between Social Capital and Mental Distress, County Health Rankings and Roadmaps, United States, 2018

Variable	Direct					Indirect				
	W_0	W_1	W_2	W_3	W_4	W_0	W_1	W_2	W_3	W_4
Social capital index	−0.072[a]	−0.002	−0.007[a]	−0.002[a]	−0.002[a]	−0.015	−0.062[a]	−0.039[a]	−0.032[a]	−0.023[a]
Socioeconomic status index	−1.229[a]	0.093[a]	−0.090[a]	−0.005[a]	−0.015[a]	0.742[a]	−0.448[a]	−0.168[a]	−0.183[a]	−0.122[a]
Younger than 18 y, %	0.024[a]	−0.006[a]	0.001	−0.001[a]	0.000	−0.047[a]	−0.011	−0.013[a]	−0.008[a]	−0.006[a]
Older than 65 y, %	−0.076[a]	0.007[a]	−0.005[a]	0.000	−0.001[a]	0.057[a]	−0.021[a]	−0.005	−0.007	−0.005[a]
Non-Hispanic black, %	−0.010[a]	0.001	−0.001[a]	0.000[a]	0.000[a]	0.004	−0.004[a]	−0.002[a]	−0.002	−0.001[a]
Non-Hispanic Asian, %	−0.014	0.002	−0.001	0.000	0.000	0.013	−0.002	0.000	−0.001	0.000
Hispanic, %	−0.028[a]	0.003[a]	−0.002[a]	0.000	0.000[a]	0.022[a]	−0.007[a]	−0.001	−0.002	−0.001[a]
Not English proficient, %	0.006	−0.001	0.000	0.000	0.000	−0.007	0.000	−0.001	0.000	0.000
Female, %	0.106[a]	−0.005[a]	0.008[a]	0.001[a]	0.002[a]	−0.039[a]	0.054[a]	0.027[a]	0.025	0.017[a]
Rural resident, %	0.001	0.000	0.000	0.000	0.000	−0.001	0.000	0.000	0.000	0.000
80th/20th income ratio	0.345[a]	−0.048[a]	0.020[a]	−0.004[a]	0.002[a]	−0.388[a]	0.017	−0.043[a]	−0.013	−0.014

[a] Significant at $P < .05$.

ORIGINAL RESEARCH

Identification of Resilient and At-Risk Neighborhoods for Cardiovascular Disease Among Black Residents: the Morehouse-Emory Cardiovascular (MECA) Center for Health Equity Study

Jeong Hwan Kim, MD[1]; Tené T. Lewis, PhD[2]; Matthew L. Topel, MD, MSc[1]; Mohamed Mubasher, PhD[3];

Chaohua Li, MPH[4]; Viola Vaccarino, MD, PhD[1,2]; Mahasin S. Mujahid, PhD[5]; Mario Sims, PhD, MS[6];

Arshed A. Quyyumi, MD[1]; Herman A. Taylor Jr, MD[7]; Peter T. Baltrus, PhD[3,4]

PEER REVIEWED

Summary

What is already known about this topic?

Residential neighborhood and neighborhood socioeconomic status (SES) are important determinants of cardiovascular disease (CVD) outcomes. It remains understudied what types of neighborhoods promote resilience or increase risk of CVD beyond the effect of neighborhood SES, especially among black Americans, who have a disparately higher prevalence of CVD than white Americans.

What is added by this report?

In the Atlanta, Georgia, metropolitan area, using the census tract-level rates of cardiovascular mortality and morbidity for black residents during 2010–2014, we identified 106 resilient neighborhoods and 121 at-risk neighborhoods where black residents had substantially lower-than-expected and higher-than-expected rates of CVD events, respectively, despite similarities in their neighborhood income levels. Yet, certain socioeconomic indicators of inequalities remained important determinants of neighborhood-level CVD risk.

What are the implications for public health practice?

Better characterization of resilient and at-risk neighborhood for black Americans helps identify neighborhood-level factors that promote resilience to CVD and helps guide community-level interventions to improve CVD outcomes for black residents in high-risk areas.

Abstract

Introduction

Despite the growing interest in place as a determinant of health, areas that promote rather than reduce cardiovascular disease (CVD) in blacks are understudied. We performed an ecologic analysis to identify areas with high levels of CVD resilience and risk among blacks from a large southern, US metropolitan area.

Methods

We obtained census tract–level rates of cardiovascular deaths, emergency department (ED) visits, and hospitalizations for black adults aged 35 to 64 from 2010 through 2014 for the Atlanta, Georgia, metropolitan area. Census tracts with substantially lower rates of cardiovascular events on the basis of neighborhood socioeconomic status were identified as resilient and those with higher rates were identified as at risk. Logistic regression was used to estimate the odds ratios (OR) and 95% confidence intervals (CIs) of being classified as an at-risk versus resilient tract for differences in census-derived measures.

Results

We identified 106 resilient and 121 at-risk census tracts, which differed in the rates per 5,000 person years of cardiovascular outcomes (mortality, 8.13 vs 13.81; ED visits, 32.25 vs 146.3; hospitalizations, 26.69 vs 130.0), despite similarities in their median black income ($46,123 vs $45,306). Tracts with a higher percentage of residents aged 65 or older (odds ratio [OR], 2.29; 95% CI, 1.41–3.85 per 5% increment) and those with incomes less than 200% of the federal poverty level (OR, 1.19; 95% CI, 1.02–1.39

per 5% increment) and greater Gini index (OR, 1.56; 95% CI, 1.19– 2.07 per 0.05 increment) were more likely to be classified as at risk than resilient neighborhoods.

Discussion

Despite matching on median income level, at-risk neighborhoods for CVD among black populations were associated with a higher prevalence of socioeconomic indicators of inequality than resilient neighborhoods.

Introduction

Despite the recent, overall reduction in cardiovascular events in the United States, cardiovascular disease (CVD) rates are still higher among black Americans than among white Americans (1,2). Although this interracial disparity in CVD is a public health concern, a substantial degree of intraracial heterogeneity exists within the black population that is often overlooked. More than 50% of black Americans have no form of CVD or cardiovascular risk factors (3). Nevertheless, the factors that promote resilience to CVD among blacks are understudied.

Factors that confer cardiovascular resilience are likely multifactorial, consisting of individual and environmental elements (3). Recent studies have demonstrated residential "place" as a determinant of cardiovascular outcomes (4–7). For example, neighborhood characteristics such as food access, aspects of the built environment, safety, and social cohesion have been individually linked with the cardiovascular health of the residents (7). Furthermore, across racial groups, there is significant variability in CVD by national (6,8) and regional geographic locations (5,9). This geographic variability suggests that certain residential contexts promote cardiovascular health while others increase cardiovascular risk and disease. A better characterization of the spatial contexts that positively promote cardiovascular health (ie, areas with cardiovascular resilience, particularly for black residents), is important in understanding the CVD burden for black Americans and guiding interventions to improve outcomes among them.

We investigated the resilience of neighborhoods against expected CVD rates among black adults in Atlanta, Georgia. By using census tract–level cardiovascular mortality and morbidity rates, we identified neighborhoods that were resilient or at risk for CVD among black residents. Specifically, we identified resilient and at-risk neighborhoods that were not predominantly confounded by differences in neighborhood socioeconomic status (SES), an established determinant of cardiovascular outcomes (7,10–12). Lastly, we conducted an ecologic-level analysis of the census-derived measures to identify the characteristics that distinguish resilient and at-risk areas.

Methods

Geographic region of the study. This study was completed as part of the Morehouse–Emory Cardiovascular (MECA) Center for Health Equity project. Census tract was used as the unit of analysis. Data were obtained and analyzed for the 992 census tracts in the 36-county Atlanta–Athens–Clarke–Sandy Springs combined statistical area that makes up the Atlanta metropolitan area (Figure 1).
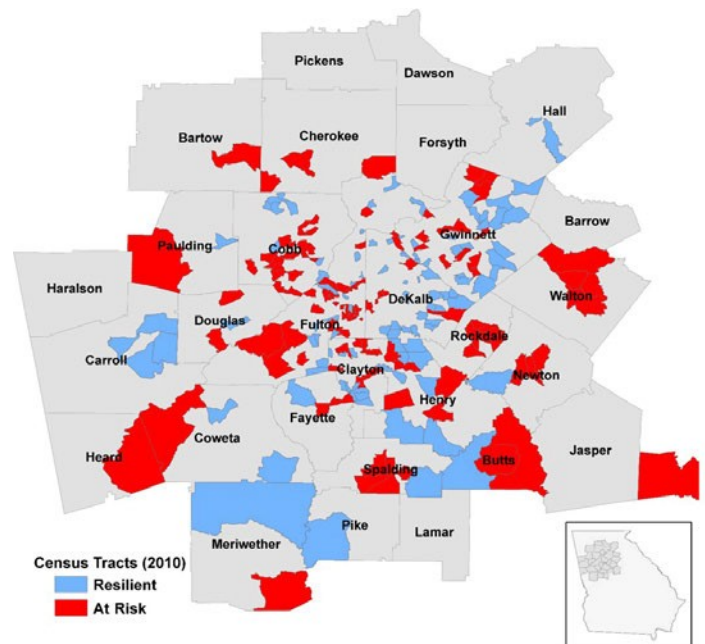
Figure 1. Study region of the Morehouse–Emory Cardiovascular Center for Health Equity project conducted in the Atlanta, Georgia, metropolitan area with 2010 census tract boundaries. Resilient and at-risk census tracts identified by the residual percentile method are indicated.

Mortality data. Cardiovascular mortality data for the 5-year period from 2010 through 2014 were obtained from the Georgia Department of Public Health. We received the counts of all deaths attributable to cardiovascular causes (identified as ICD 10 codes I00–I78, from the *International Classification of Diseases, Tenth Revision* [13] or ICD 9 codes 390-434 and 436–448 from the *International Classification of Diseases, Ninth Revision* [14]) for blacks aged 35 to 64, the age group that captured most of the population with CVD risk while excluding those aged 65 or older to minimize the confounding by noncardiac comorbidities. Counts for census tracts with fewer than 5 deaths were censored for confidentiality reasons, which resulted in a total of 347 census tracts with uncensored data. Additionally, to minimize the number of census tracts censored because of few events and to ensure stable

events rates over the 5-year period, only the tracts with at least 200 black adults aged 35 to 64 were included (N = 346). Counts of deaths were then divided by the black population aged 35 to 64 living in the respective census tracts (2010 US Census data) (15) to generate the mortality rate for each census tract. The rates were reported as the number of events per 5,000 person-year (per 1,000 people over the 5-year period).

Morbidity data. Cardiovascular morbidity data from 2010 through 2014 were obtained from the Georgia Hospital Association. We obtained aggregated counts of emergency department (ED) visits and hospitalizations for cardiovascular reasons, identified with ICD 10 codes I00–I78 (13) or ICD 9 codes 390–434 and 436–448 (14) for blacks aged 35 to 64 from 2010 through 2014. Census tracts with fewer than 6 events were censored for confidentiality reasons, resulting in 802 tracts with uncensored data for ED visit and 763 tracts for hospitalization data. As with mortality, only tracts with at least 200 black adults aged 35 to 64 were included (N = 693 for ED visits; N = 675 for hospitalizations). Counts of ED visits and hospitalizations were divided by the population of blacks aged 35 to 64 living in the respective census tract (2010 US Census data) (15) to calculate the rates of hospitalization and ED visits for each census tract. The rates were reported as the number of events per 5,000 person-year.

Census-derived measures. We obtained census tract data from the 2010 US Decennial Census (15) to characterize the demographic and socioeconomic composition of the identified at-risk and resilient census tracts. The variables selected included factors that have been previously linked with CVD, such as SES and housing-related indicators (5,10,16), and measures of demographic composition. Demographic data obtained were percentage female, black median age, percentage aged 65 or older, percentage aged 17 or younger, percentage minority population, percentage black population, percentage speaking English less than well, percentage of single-parent households, and percentage civilians with a disability. For the measures of SES, we obtained median black household income, percentage education certifications (high school, college), percentage unemployed, percentage with incomes below the federal poverty level, percentage with incomes below 200% of the federal poverty level (ie, percentage of the population with income below twice the federal poverty level, as an index of the proportion in or near poverty), and Gini index (17) (a measure of income inequality from perfect equality [0], where everyone receives the same income, to perfect inequality [1], where a single person receives the total income of the community). For housing-related measures, median home value, percentage living in multiunit structures, percentage living in mobile homes, percentage living in crowded units (defined as housing units occupied by more than 1 person per room), and percentage living in group quarters. Finally, the percentage of households without a vehicle was assessed as a measure of transportation accessibility.

Identification of resilient and at-risk census tracts. We identified census tracts that were resilient and at risk based on the aforementioned measures of cardiovascular outcomes: deaths, ED visits, and hospitalizations. First, we identified low-rate and high-rate census tracts solely on the basis of the distribution of the outcome measures. A census tract was considered low-rate on one of the 3 measures if its rate was in the bottom quartile of the measure and high-rate if its rate was in the highest quartile of the measure. Then, if a census tract was considered low-rate on at least 2 of the 3 measures and not high-rate for any measure, the tract was classified as a low-rate census tract. Similarly, being labeled as a high-rate tract on at least 2 of the 3 measures and not low-rate on any measure classified the tract as high-rate.

Because it is well documented that neighborhood SES is a strong determinant of cardiovascular outcomes (5,10,11), we identified areas that were not predominantly confounded by differences in neighborhood SES. We used the residual percentile method, which is similar to a method used to by Fry-Johnson et al (18) to identify counties with low infant mortality rates independent of county-level SES. By using this method (Figure 2), we identified census tracts that had substantially lower or higher rates of CVD outcomes than the rates that would be expected on the basis of their neighborhood SES. Census tracts with lower than expected CVD outcome rates were defined as resilient, and those with higher than expected CVD rates were defined as at-risk. To do so, a negative binomial model was built for each of the 3 measures. Each model was adjusted for census tract-level socioeconomic variables for blacks, including age distribution (in 5-year age groups), percentage male, and median black household income. Census tracts without any missing covariate were included in the model (N = 346 for mortality; N = 689 for ED visits; N = 671 for hospitalizations). Census tracts with model residuals in the highest 25% (substantially higher rates than predicted) were considered at risk for the measure. Similarly, tracts with model residuals in the lowest 25% (substantially lower rates than predicted) were considered resilient for the measure. Census tracts at risk or resilient on at least 2 of 3 measures were finally labeled as at-risk or resilient census tracts, respectively, and included in our analysis. Any census tract designated at risk for one measure but resilient for any other measures, or vice versa, was excluded.

Figure 2. The steps in the identification of at-risk and resilient census tracts by the residual percentile method. Census tract-level CV outcome data for blacks aged 35 to 64 from 992 census tracts in 36 counties in the Atlanta–Athens–Clarke–San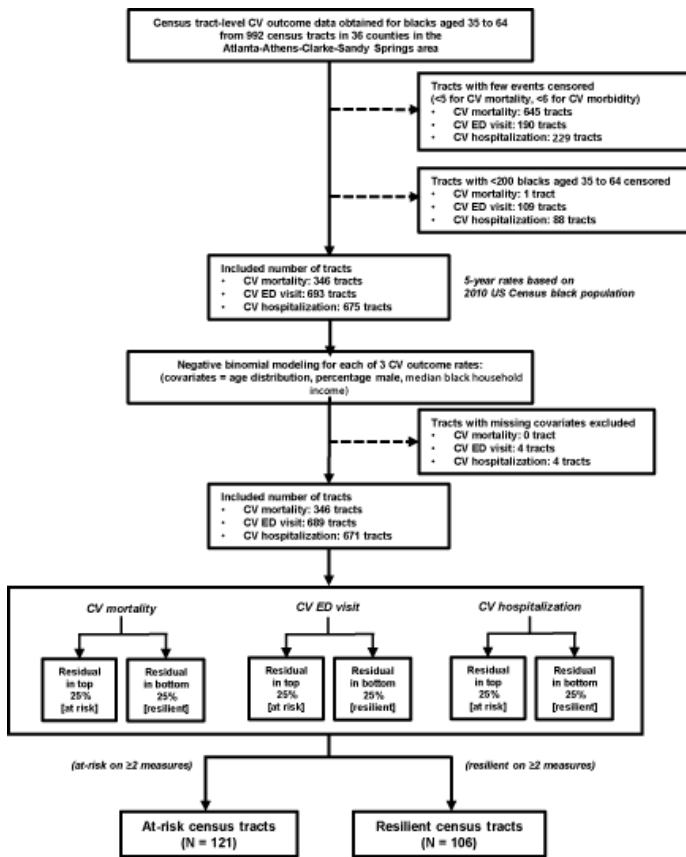dy Springs combined statistical area were used to identify 121 at-risk and 106 resilient census tracts. Abbreviations: CV, cardiovascular; ED, emergency department.

Statistical analysis. We used *t* tests to compare demographic and socioeconomic measures of at-risk and resilient census tracts, which we identified by the residual percentile method. The measures that were significantly different were subsequently analyzed by using logistic regression models. The OR and 95% CI for being labeled at-risk census tracts compared with resilient tracts were estimated in bivariate and multivariable models, for 5% increment in the included census tract measures. We verified absence of any major collinearity among the explanatory variables by computing the condition index (19) in the fully adjusted model (27.49). *P* < .05 was considered significant. Statistical analyses were performed by using SAS version 9.3 (SAS Institute Inc).

Results

In our initial analyses, unadjusted for neighborhood SES, we identified 130 low-rate and 137 high-rate census tracts. Tracts selected using this approach differed in their CVD outcome measures as expected (mortality: 6.27 for low-rate tracts vs 15.75 for high-rate tracts; ED visits: 27.67 for low-rate tracts vs 159.70 for high-rate tracts; hospitalizations: 21.60 for low-rate tracts vs 165.10 for high-rate tracts; per 5,000 person-year), but they also had substantial difference in the median black household income levels ($60,980 for low-rate tracts vs $29,015 for high-rate tracts). By using the residual percentile method, we identified 106 resilient and 121 at-risk census tracts. The resilient census tracts had lower rates of cardiovascular mortality, hospitalization, and ED visits than the at-risk census tracts, but the median black household income levels of the resilient and the at-risk census tracts did not differ from each other substantially (Table 1). Furthermore, resilient and at-risk census tracts were located throughout the metropolitan Atlanta area without clustering of either resilient or at-risk tracts, and resilient and at-risk census tracts were also often adjacent to one another (Figure 1).

The median age of black residents was similar in resilient and at-risk census tracts, but the proportion of residents aged 65 or older was significantly lower in resilient census tracts than in at-risk census tracts (*P* < .001) (Table 2). The proportion of women and black residents were also similar in both neighborhood types. However, fewer civilians with a disability resided in resilient census tracts than in at-risk tracts (*P* < .001).

For socioeconomic measures, resilient census tracts had a higher percentage of college graduates and those with some college education than at-risk census tracts (*P* = .01 and .007, respectively). Similarly, there were more people with high school diploma or less in at-risk census tracts than in resilient tracts (*P* < .001). Though the median black household income was similar and the percentage of people with incomes below the federal poverty level were similar in the 2 groups, resilient census tracts had fewer residents with incomes below 200% of the federal poverty level than at-risk census tracts and had significantly lower Gini index than at-risk census tracts (0.38 vs 0.42, *P* < .001). Other housing measures did not differ significantly between resilient and at-risk tracts, but more households in at-risk census tracts had no vehicle than in resilient tracts (*P* = .02).

Six measures that differed significantly (*P* < .05) between resilient and at-risk census tracts were included in regression analyses: percentage aged 65 or older, percentage of civilians with a disability, percentage with no high school diploma, percentage with incomes below 200% of the federal poverty level, Gini index, and percent-

age with no vehicle in household (Table 3). After simultaneous adjustment in the model, census tracts with a 5% increment in the proportion aged 65 or older were 2.29 times (95% CI, 1.41–3.85) more likely to be categorized as at-risk tracts. Similarly, tracts with 5% increment in the percentage below 200% poverty were 1.19 times (95% CI, 1.02–1.39) more likely to be designated as at-risk tracts. Finally, tracts with a 0.05 higher Gini index were 1.56 times (95% CI, 1.19–2.07) more likely to be classified as at-risk tracts.

Discussion

We identified several demographic and socioeconomic indicators of income and education inequality at the ecologic level that distinguished at-risk neighborhoods from resilient neighborhoods; having a higher proportion of residents aged 65 or older and residents with income below 200% of the federal poverty level and greater income inequality were independent factors that separated at-risk neighborhoods from resilient neighborhoods. To our knowledge, this study is the first to use census tract–level data to identify areas resilient to and at risk for CVD for black residents in a large US metropolitan area.

Our approach to identify resilient and at-risk neighborhoods was unique in that we quantified the deviation of cardiovascular mortality and morbidity for neighborhoods from what would be predicted on the basis of their neighborhood SES. Over the past 2 decades, studies have demonstrated that living in socioeconomically disadvantaged neighborhoods is associated with a greater burden of cardiovascular risk and disease (7,12). This association has been demonstrated not only with cardiovascular risk factors (11,20,21), but also with incidence of CVD (5,22) and cardiovascular mortality (10,23). However, despite the growing interest in neighborhoods as a determinant of health, less is known about outlier communities that have an unusually lower or higher burden of CVD than what would be expected given their socioeconomic composition. Understanding of those outlier communities will elucidate neighborhoods' health-promoting factors better than using SES.

Reports of such outlier communities date back as early as the 1960s (24), but contemporary data from the United States is still largely lacking. The bulk of available evidence on resilient neighborhood comes from research in Europe (25–28) and New Zealand (29), in which neighborhoods with higher or lower rates of all-cause mortality and morbidity than predicted from neighborhood SES were identified, similar to the approach we used in this analysis. However, our analysis differed from these reports in 2 major aspects. First, we examined cardiovascular-specific mortality and morbidity whereas the other studies examined all-cause mortality or morbidity. As previously reported (27), the resilience of neighborhoods may differ depending on the etiologies of mortality, and examination of cause-specific mortality and morbidity as in our analysis helps identify potential mechanistic pathways between neighborhood characteristics and CVD more directly. Second, previous studies extracted mortality and morbidity data from the entire population of the examined communities, potentially masking the racial/ethnic differences in the association between neighborhoods and individuals. On the other hand, we focused on a specific racial group, blacks, to explore the intraracial differences between types of neighborhood on CVD and eventually to help design effective interventions to improve neighborhoods for better cardiovascular outcomes of among black residents.

We also identified several independent features that distinguished resilient and at-risk neighborhoods for CVD in black residents. Not only do these factors illustrate the primary ecologic-level determinants of neighborhood resilience or risk for CVD for black residents, but they also could provide insights into policy design or community-level interventions to improve cardiovascular outcomes among blacks. First, despite similarities in the median age and the proportion of population aged 17 or younger, at-risk census tracts had a higher proportion of residents aged 65 or older than resilient census tracts. A similar finding was also previously reported in relation to all-cause mortality (26). Interestingly, the cardiovascular outcome data used in our analysis did not include people aged 65 or older. Thus, although an older age is a known risk factor for cardiovascular mortality and morbidity (30), the proportion of those aged 65 or older likely represents a proxy for contextual factors of the at-risk neighborhood environment. For example, a higher proportion of elderly residents may correlate with a stagnant or declining overall population with fewer middle-aged working residents, whereas a greater influx of residents, likely with more economic opportunities, may be associated with resilient neighborhoods (29,31). Further characterization of the population composition with trajectory may help further elucidate the significance of the percentage of the elderly in the CVD resilience and risk of the overall neighborhood.

Secondly, both a higher proportion of those with incomes under 200% of the federal poverty level and greater income inequality were also independently associated with at-risk neighborhoods compared with resilient neighborhoods. Although the median black income and percentage of those under the poverty level were similar in resilient and at-risk neighborhoods, our results suggest that even moderate deprivation of income (ie, those in the near-poverty and the resultant income equality despite similarities in the median income) could adversely affect CVD outcomes among black residents. In addition to the level of neighborhood income it-

self (7,12), income inequality has been previously associated with CVD burden (32,33). Thus, our findings reconfirm that socioeconomic deprivation, even at a moderate degree, may affect cardiovascular resilience and risk at the ecologic level. Whether income deprivation and inequality represent proxies for other contextual factors of neighborhoods remains to be investigated. Although limited in our analysis, further characterization of people with incomes at the poverty or near-poverty level would be important, because they may be the vulnerable population that would most benefit from the appropriate aid to improve their cardiovascular outcomes or prevention measures.

Our study has limitations. Because of its cross-sectional design, any inference of causation from the observed findings is limited. Longitudinal analyses of the neighborhood resilience and the neighborhood-level cardiovascular outcomes would be needed. Furthermore, the definition of neighborhood in a fixed unit of census tracts may have masked variability of smaller communities and residential contexts. Similar analysis in smaller units, such as census block, may be informative to validate or augment our analysis. Third, because the data examined were limited at the ecologic level, the subjective, contextual factors of living in a given neighborhood are not accounted for in our analysis. However, our work was undertaken as the first cornerstone of the larger MECA project, which aims for a multilevel exploration of cardiovascular resilience of US black adults and lays a foundation for continued investigation. In the subsequent stages of the MECA project, we plan to examine the characteristics of the identified at-risk and resilient neighborhoods at the individual level, which would enable us to better understand the contextual versus compositional factors contributing risk or resilience to the residents of the selected tracts.

In conclusion, by using neighborhood-level data on cardiovascular mortality and morbidity for black residents, we identified resilient and at-risk neighborhoods for CVD among black adults in a large southern US city. These resilient and at-risk neighborhoods substantially differed in the rates of cardiovascular mortality and morbidity despite their similar income levels, suggesting that they represent a distinct residential context, or place, that promotes or jeopardizes the cardiovascular health of its black residents beyond the effect of neighborhood SES. However, even with our definitions of resilient and at-risk neighborhoods, certain socioeconomic indicators of inequality remained important predictors of CVD risk at the neighborhood level. Further exploration of contextual factors other than neighborhood SES are needed to fully characterize the factors that constitute a residential place that either promotes or threatens the cardiovascular health of its black residents.

Author Information

Corresponding Author: Peter Baltrus, PhD, Morehouse School of Medicine, National Center for Primary Care, Room 310, 720 Westview Dr, Atlanta, GA 30310. Telephone: 404-520-4417. Email: pbaltrus@msm.edu.

Author Affiliations: [1]Department of Medicine, Division of Cardiology, Emory University School of Medicine, Atlanta, Georgia. [2]Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, Georgia. [3]Department of Community Health and Preventive Medicine, Morehouse School of Medicine, Atlanta, Georgia. [4]National Center for Primary Care, Morehouse School of Medicine, Atlanta, Georgia. [5]Division of Epidemiology, School of Public Health, University of California, Berkeley, Berkeley, California. [6]Department of Medicine, University of Mississippi Medical Center, Jackson, Mississippi. [7]Department of Medicine, Morehouse School of Medicine, Atlanta, Georgia.

References

1. Carnethon MR, Pu J, Howard G, Albert MA, Anderson CAM, Bertoni AG, et al.; American Heart Association Council on Epidemiology and Prevention; Council on Cardiovascular Disease in the Young; Council on Cardiovascular and Stroke Nursing; Council on Clinical Cardiology; Council on Functional Genomics and Translational Biology; and Stroke Council. Cardiovascular health in African Americans: a scientific statement from the American Heart Association. Circulation 2017;136(21):e393–423.
2. Van Dyke M, Greer S, Odom E, Schieb L, Vaughan A, Kramer M, et al. Heart Disease death rates among blacks and whites aged ≥35 years — United States, 1968–2015. MMWR Surveill Summ 2018;67(5):1–11.
3. Zautra AJ, Hall JS, Murray KE; the Resilience Solutions Group. Resilience: a new integrative approach to health and mental health research. Health Psychol Rev 2008;2(1):41–64.

4. Kiefe CI, Williams OD, Bild DE, Lewis CE, Hilner JE, Oberman A. Regional disparities in the incidence of elevated blood pressure among young adults: the CARDIA study. Circulation 1997;96(4):1082–8.

5. Diez Roux AV, Merkin SS, Arnett D, Chambless L, Massing M, Nieto FJ, et al. Neighborhood of residence and incidence of coronary heart disease. N Engl J Med 2001;345(2):99–106.

6. Kershaw KN, Diez Roux AV, Carnethon M, Darwin C, Goff DC Jr, Post W, et al. Geographic variation in hypertension prevalence among blacks and whites: the multi-ethnic study of atherosclerosis. Am J Hypertens 2010;23(1):46–53.

7. Diez Roux AV, Mujahid MS, Hirsch JA, Moore K, Moore LV. The impact of neighborhoods on CV risk. Glob Heart 2016; 11(3):353–63.

8. Casper M, Kramer MR, Quick H, Schieb LJ, Vaughan AS, Greer S. Changes in the geographic patterns of heart disease mortality in the United States: 1973 to 2010. Circulation 2016; 133(12):1171–80.

9. Vaughan AS, Kramer MR, Casper M. Geographic disparities in declining rates of heart disease mortality in the southern United States, 1973-2010. Prev Chronic Dis 2014;11:E185.

10. Borrell LN, Diez Roux AV, Rose K, Catellier D, Clark BL; Atherosclerosis Risk in Communities Study. Neighbourhood characteristics and mortality in the Atherosclerosis Risk in Communities Study. Int J Epidemiol 2004;33(2):398–407.

11. Clark CR, Ommerborn MJ, Hickson DA, Grooms KN, Sims M, Taylor HA, et al. Neighborhood disadvantage, neighborhood safety and cardiometabolic risk factors in African Americans: biosocial associations in the Jackson Heart study. PLoS One 2013;8(5):e63254.

12. Diez Roux AV, Mair C. Neighborhoods and health. Ann N Y Acad Sci 2010;1186(1):125–45.

13. National Center for Health Statistics. Centers for Disease Control and Prevention. International Classification of Diseases, Tenth Revision. https://www.cdc.gov/nchs/icd/icd10.htm. Accessed December 19, 2015.

14. National Center for Health Statistics. Centers for Disease Control and Prevention. International Classification of Diseases, Ninth Revision. https://www.cdc.gov/nchs/icd/icd9.htm. Accessed: December 19, 2015.

15. US Decennial Census, 2010. https://www.census.gov/programs-surveys/decennial-census/decade.2010.html. Accessed December 19, 2015

16. Bang DW, Manemann SM, Gerber Y, Roger VL, Lohse CM, Rand-Weaver J, et al. A novel socioeconomic measure using individual housing data in cardiovascular outcome research. Int J Environ Res Public Health 2014;11(11):11597–615.

17. Cowell FA. Measurement of inequality. In: Atkinson A, Bourguignon F, editors. Handbook of income distribution, volume 1. Amsterdam (NL): Elsevier; 2000. p. 87–166.

18. Fry-Johnson YW, Levine R, Rowley D, Agboto V, Rust G. United States black:white infant mortality disparities are not inevitable: identification of community resilience independent of socioeconomic status. Ethn Dis 2010;20(1 Suppl):S1–131–5.

19. Habshah M, Sarkar SK, Rana S. Collinearity diagnostics of binary logistic regression model. Journal of Interdisciplinary Mathematics 2010;13(3):253–67.

20. Cubbin C, Sundquist K, Ahlén H, Johansson SE, Winkleby MA, Sundquist J. Neighborhood deprivation and cardiovascular disease risk factors: protective and harmful effects. Scand J Public Health 2006;34(3):228–37.

21. Diez-Roux AV, Nieto FJ, Muntaner C, Tyroler HA, Comstock GW, Shahar E, et al. Neighborhood environments and coronary heart disease: a multilevel analysis. Am J Epidemiol 1997;146(1):48–63.

22. Sundquist K, Winkleby M, Ahlén H, Johansson S-E. Neighborhood socioeconomic environment and incidence of coronary heart disease: a follow-up study of 25,319 women and men in Sweden. Am J Epidemiol 2004;159(7):655–62.

23. Diez Roux AV, Borrell LN, Haan M, Jackson SA, Schultz R. Neighbourhood environments and mortality in an elderly cohort: results from the cardiovascular health study. J Epidemiol Community Health 2004;58(11):917–23.

24. Stout C, Marrow J, Brandt EN Jr, Wolf S. Unusually low incidence of death from myocardial infarction: Study of an italian american community in pennsylvania. JAMA 1964; 188(10):845–9.

25. Doran T, Drever F, Whitehead M. Health underachievement and overachievement in English local authorities. J Epidemiol Community Health 2006;60(8):686–93.

26. van Hooijdonk C, Droomers M, van Loon JAM, van der Lucht F, Kunst AE. Exceptions to the rule: healthy deprived areas and unhealthy wealthy areas. Soc Sci Med 2007; 64(6):1326–42.

27. Tunstall H, Mitchell R, Gibbs J, Platt S, Dorling D. Is economic adversity always a killer? Disadvantaged areas with relatively low mortality rates. J Epidemiol Community Health 2007;61(4):337–43.

28. Cairns JM, Curtis SE, Bambra C. Defying deprivation: a cross-sectional analysis of area level health resilience in England. Health Place 2012;18(4):928–33.

29. Pearson AL, Pearce J, Kingham S. Deprived yet healthy: neighbourhood-level resilience in New Zealand. Soc Sci Med 2013;91:238–45.

30. Yazdanyar A, Newman AB. The burden of cardiovascular disease in the elderly: morbidity, mortality, and costs. Clin Geriatr Med 2009;25(4):563–77, vii.

31. Mitchell R, Gibbs J, Tunstall H, Platt S, Dorling D. Factors which nurture geographical resilience in Britain: a mixed methods study. J Epidemiol Community Health 2009; 63(1):18–23.

32. Diez-Roux AV, Link BG, Northridge ME. A multilevel analysis of income inequality and cardiovascular disease risk factors. Soc Sci Med 2000;50(5):673–87.

33. Kim D, Kawachi I, Hoorn SV, Ezzati M. Is inequality at the heart of it? Cross-country associations of income inequality with cardiovascular diseases and risk factors. Soc Sci Med 2008;66(8):1719–32.

Tables

Table 1. Mean Rates of Cardiovascular Outcomes and Median Household Income for Black Residents in Resilient and At-Risk Census Tracts[a], Atlanta, Georgia, 2010–2014

Variable	Resilient Tract (n = 106)	At Risk Tract(n = 121)	P Value
Mortality rate[b]	8.1	13.8	<.001
Emergency department visits[b]	32.3	146.3	<.001
Hospitalization rate[b]	26.7	130.0	<.001
Median household income, $	46,123	45,306	.79

[a] Selected by the residual percentile method.
[b] Number of events per 5,000 person-year.

Table 2. Comparison of Demographic, Socioeconomic, Housing and Transportation Characteristics of Resilient and At-Risk Census Tracts, Atlanta, Georgia[a]

Variable	Resilient Tract (n = 106)	At-Risk Tract (n = 121)	P Value
Demographic characteristic			
% Female	54.8	55.6	.29
Median black age, y	32.3	32.1	.77
% Aged ≥65 y	7.8	10.4	<.001
% Aged ≤17 y	26.4	25.3	.19
% Racial/ethnic minority population	67.7	62.5	.14
% Black population	48.8	45.3	.38
% Speaking English less than well	4.8	4.0	.34
% Single-parent households	13.9	14.0	.88
% Civilians with a disability	9.7	12.0	<.001
Socioeconomic status of residents			
Median black income, $	46,123	45,306	.79
% With no high school diploma	13.3	16.3	.02
% With high school diploma or less	34.8	43.3	<.001
% With some college	35.8	32.4	.007
% College graduate	29.4	24.4	.01
% Unemployed	13.2	13.4	.85
% With income below federal poverty level	20.2	22.8	.14
% With income below 200% of federal poverty level	33.7	40.7	.003
Gini index[b]	0.38	0.42	<.001
Housing			
Median home value, $	181,761.00	176,008.00	.62
% Multi-unit structure	18.3	13.8	.10
% Mobile home	2.5	2.5	.97
% Crowded unit	3.2	3.1	.96
% Living in group quarter	0.9	1.7	.27
Transportation: % with no vehicle in household	7.6	10.8	.02

[a] Values are mean values of percentage values unless noted otherwise.

[b] A measure of income inequality from perfect equality (0), where everyone receives the same income, to perfect inequality (1), where a single person receives the total income of the community.

Table 3. Predictors of Census Tracts Being At Risk Versus Resilient (N = 227), Atlanta Metropolitan Area[a]

Variable	Crude	Adjusted
	Odds Ratio (95% Confidence Interval)	
% Aged ≥65 y	2.11 (1.51–3.03)[b]	2.29 (1.41–3.85)[b]
% With disability	1.77 (1.31–2.43)[b]	1.12 (0.70–1.81)
% With no high school diploma	1.19 (1.03–1.38)[b]	0.98 (0.79–1.22)
% With annual income below 200% of federal poverty level	1.12 (1.04–1.22)[b]	1.19 (1.02–1.39)[b]
Gini index[c], per 0.05 increment	1.59 (1.28–2.02)[b]	1.56 (1.19 -2.07)[b]
% With no vehicle in household	1.17 (1.02–1.35)[b]	0.82 (0.66–1.02)

[a] Crude and adjusted odds ratios of being classified as an at-risk census tract versus a resilient census tracts are shown for 5% increments in each of the examined factors except for Gini index (per 0.05 unit increment).
[b] Significant ($P < .05$) results.
[c] A measure of income inequality from perfect equality [0], where everyone receives the same income, to perfect inequality [1], where a single person receives the total income of the community.

GIS SNAPSHOTS

Sidewalk Conditions in Northern New Jersey: Using Google Street View Imagery and Ordinary Kriging to Assess Infrastructure for Walking

Jesse J. Plascak, PhD[1,2]; Adana A. M. Llanos, PhD, MPH[1,2]; Laxmi B. Chavali, MPH[1];

Cathleen Y. Xing, MPH[1]; Nimit N. Shah, BS[1]; Antoinette M. Stroup, PhD[2,3]; Jessica Plaha[4];

Emily M. McCue[5]; Andrew G. Rundle, PhD[6]; Stephen J. Mooney, PhD[7]

PEER REVIEWED
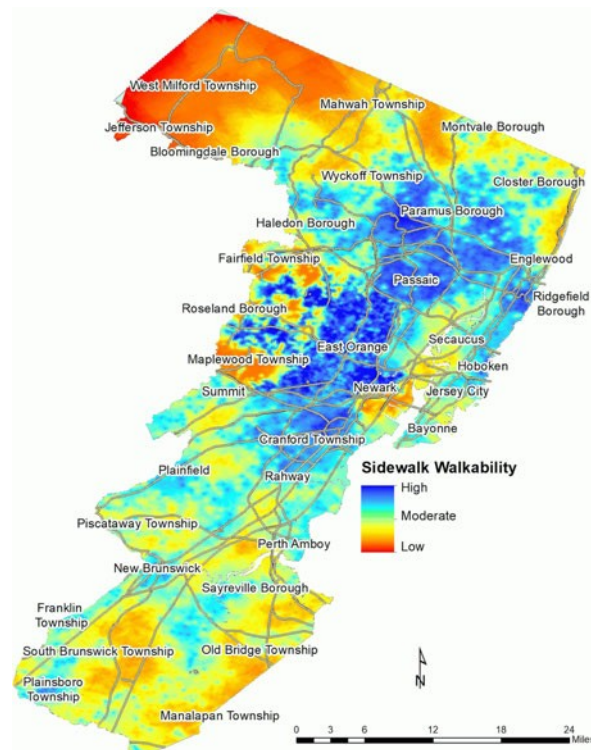
Estimated presence or absence of sidewalks and conditions of sidewalks in northeastern New Jersey. The map depicts an index of sidewalk walkability estimated from virtual street audits at 11,282 locations using Google Street View and spatial interpolation techniques. Levels of walkability ranged from low (no sidewalk or poor condition) to moderate (fair condition) to high (good condition). Precise measures of sidewalk conditions can help identify barriers to walking-based physical activity and key areas for intervention to maintain and modify sidewalk conditions.

Background

Morbidity and mortality from chronic conditions such as obesity, diabetes, heart disease, and depression among the US population are a critical public health issue (1,2). Substantial evidence indicates that aerobic physical activity, including walking, can reduce the risk of numerous physical and mental health conditions (3,4). Walking is an excellent way to achieve the recommended amount of aerobic physical activity (≥150 minutes per week) (5). According to the Behavioral Risk Factor Surveillance System, 48.7% of US adults in 2015 did not attain the recommended amount of weekly aerobic physical activity (6).

Various built environment characteristics, including sidewalk characteristics (eg, connectivity, continuity, width, barriers, condition), could influence walkability and physical activity (7). Although street audits that observe built environment characteristics in communities are common, few studies have assessed differences in observed characteristics at the sidewalk level or address level across large, generalizable geographic areas (8). The objective of this study was to describe, in map format, sidewalk characteristics at the address level in densely populated urban and suburban areas of northeastern New Jersey, where 51.7% of adults do not participate in at least 150 minutes of weekly aerobic physical activity (6).

Methods

We characterized sidewalks during virtual street audits via the Google Street View application, Computer Assisted Neighborhood Visual Assessment System (CANVAS) (9). Virtual street audits have been validated and demonstrated to be more cost-effective than in-person audits because of lower travel time and costs (9). We used CANVAS to assess several sidewalk characteristics, including 2 items within the 360° view at each audited location: sidewalk presence (yes or no) and sidewalk condition (poor [numerous breaks, uneven sidewalk], fair [some unevenness], or good [even, no breaks]).

We selected audited locations from non-highway roads in 6 counties in New Jersey. We selected locations approximately 150 m apart in densely populated Essex County (which encompasses Newark) and locations elsewhere approximately 600 m apart. The higher-density sampling allowed for investigation of the spatial autocorrelation structure of sidewalk characteristics and motivated the less dense sampling scheme of the 5 counties other than Essex. CANVAS auditors completed a 4-hour training session to collect data consistently for the presence or absence of sidewalks and the condition of sidewalks. Auditors were trained to report the worst sidewalk condition if sidewalks of different conditions were present at the same location. Of the 8,100 audited sidewalks observed in Essex County, 405 (5%) were rerated by each of all the auditors to provide estimates of test–retest and inter-rater agreement reliability (9). Auditors performed ratings on computers that had 2 monitors: one monitor displayed data input forms and the second monitor displayed the Google Street View scene. We downloaded and analyzed data on completed ratings; 11,282 locations were available for analysis.

Data on sidewalk presence and condition were combined into a sidewalk walkability variable with the following possible ordinal values: 0 (no sidewalk), 1 (poor sidewalk condition), 2 (fair sidewalk condition), and 3 (good sidewalk condition). Test–retest and inter-rater reliability were high in the reliability subsample (all intraclass correlation coefficients ≥0.89). Spatial analyses indicated that measured sidewalk walkability values correlated with other values at locations separated up to 4,200 m (2.6 miles). We used ordinary kriging to estimate sidewalk walkability values across the study area (9). Kriging models are spatial interpolation methods that predict sidewalk walkability at nonaudited locations based, in part, on the observed similarity between walkability values assessed at audited locations. Ordinary kriging results in continuous predictions, and we plotted these continuous predictions on a map as a range of walkability, from low (no sidewalk or poor condition) to moderate (fair condition) to high (good sidewalk condition). We analyzed concurrent validity of the sidewalk walkability construct through a census tract–level Spearman correlation coefficient ($\rho = 0.22$, $P < .001$) of the relationship between average sidewalk walkability in each tract and proportion of commuters in that tract who reported walking to work in the 2012–2016 American Community Survey (10). We used SAS version 9.4 (SAS Institute Inc) and ArcGIS version 10.5 (Esri) in all analyses.

Main Findings

We found several geographic patterns in sidewalk walkability in northern New Jersey. The presence of any sidewalk and the presence of sidewalks in fair or good condition were more common in urban cores (Newark, East Orange, Passaic, and Hoboken) than outside these cores and occurred less frequently as distance from these cores increased (for example, in northern West Milford Township, northern Mahwah Township, and southern Manalapan Township). However, we found heterogeneity in sidewalk walkability at a smaller geographic scale, which was subtle in the urban cores but more apparent in the western suburbs of Newark and East Orange (for example, in Roseland Borough). Generally, sidewalks were absent or in poor condition along major roads in otherwise walkable urban cores.

Action

We used virtual street audits and spatial interpolation techniques to construct a detailed high-resolution map of sidewalk conditions in northeastern New Jersey. Such high-resolution maps can be informative and powerful tools, offering finer-grain detail on sidewalk conditions than would be available in tabular format or a choropleth map. We demonstrated that the use of innovative, spatially based sampling and estimation methods, publicly available Google Street View scenery, and the CANVAS application can allow for large-scale, routine, and standardized collection of variables related to sidewalk characteristics. Such information can be useful both for research and practice. For researchers, precise measures of sidewalk conditions can help identify barriers to walking-based physical activity. For practitioners, this map may help identify key areas for intervention to maintain and modify sidewalk conditions (11). Improvements made to walkability may be one of the most cost-effective strategies for increasing physical activity and reducing disparities in chronic disease, particularly among populations that do not achieve recommended amounts of physical activity (4). A map indicating regions for improvement in walkability may facilitate identification of regions in need of sidewalk improvements to support walking-based physical activity. Future research should extend this measure across all of New Jersey and further explore potential correlates of sidewalk conditions, such as race/ethnicity and socioeconomic status.

Acknowledgments

Author Information

Corresponding Author: Jesse J. Plascak, PhD, Department of Biostatistics and Epidemiology, School of Public Health, Rutgers, The State University of New Jersey, 683 Hoes Lane West, Room 209, Piscataway, NJ 08854. Telephone: 732-235-4664. Email: jesse.plascak@rutgers.edu.

Author Affiliations: [1]Department of Biostatistics and Epidemiology, School of Public Health, Rutgers, The State University of New Jersey, Piscataway, New Jersey. [2]Rutgers Cancer Institute of New Jersey, New Brunswick, New Jersey. [3]New Jersey State Cancer Registry, Trenton, New Jersey. [4]School of Arts and Sciences, Rutgers, The State University of New Jersey, New Brunswick, New Jersey. [5]School of Environmental and Biological Sciences, Rutgers, The State University of New Jersey, New Brunswick, New Jersey. [6]Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, New York. [7]Department of Epidemiology, University of Washington, Seattle, Washington.

References

1. Ward BW, Schiller JS, Goodman RA. Multiple chronic conditions among US adults: a 2012 update. Prev Chronic Dis 2014;11:E62.
2. Pratt LA, Brody DJ. Depression and obesity in the U.S. adult household population, 2005–2010. NCHS Data Brief 2014; (167):1–8.
3. US Department of Health and Human Services. 2018 Physical Activity Guidelines Advisory Committee scientific report. Washington (DC): US Department of Health and Human Services; 2018.
4. US Department of Health and Human Services. Step it up! The Surgeon General's call to action to promote walking and walkable communities. Washington (DC): Office of the Surgeon General; 2015.
5. Garber CE, Blissmer B, Deschenes MR, Franklin BA, Lamonte MJ, Lee IM, et al. American College of Sports Medicine position stand. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise. Med Sci Sports Exerc 2011;43(7):1334–59.
6. Centers for Disease Control and Prevention. Behavioral Risk Factor Surveillance System prevalence and trends data: 2015. https://www.cdc.gov/brfss/brfssprevalence. Accessed August 7, 2018.
7. Cerin E, Saelens BE, Sallis JF, Frank LD. Neighborhood Environment Walkability Scale: validity and development of a short form. Med Sci Sports Exerc 2006;38(9):1682–91.
8. Lopez RP, Hynes HP. Obesity, physical activity, and the urban environment: public health research needs. Environ Health 2006;5(1):25.
9. Mooney SJ, Bader MDM, Lovasi GS, Teitler JO, Koenen KC, Aiello AE, et al. Street audits to measure neighborhood disorder: virtual or in-person? Am J Epidemiol 2017; 186(3):265–73.

10. US Census Bureau. American Community Survey, 2012–2016. American Community Survey 5-year estimates; using American FactFinder. 2018. http://factfinder2.census.gov. Accessed August 30, 2018.

11. Zandieh R, Flacke J, Martinez J, Jones P, van Maarseveen M. Do inequalities in neighborhood walkability drive disparities in older adults' outdoor walking? Int J Environ Res Public Health 2017;14(7):E740.

GIS SNAPSHOTS

An Objective Walkability Index for Public Health and Planning in Peel Region, Ontario, Canada

Maria Mukhtar, MA[1]; David Guillette[2]; Natalie Lapos, RN, MN[1]; Sandra Fitzpatrick, RD, MHSc[1]; Ron Jaros, MES[2]

PEER REVIEWED
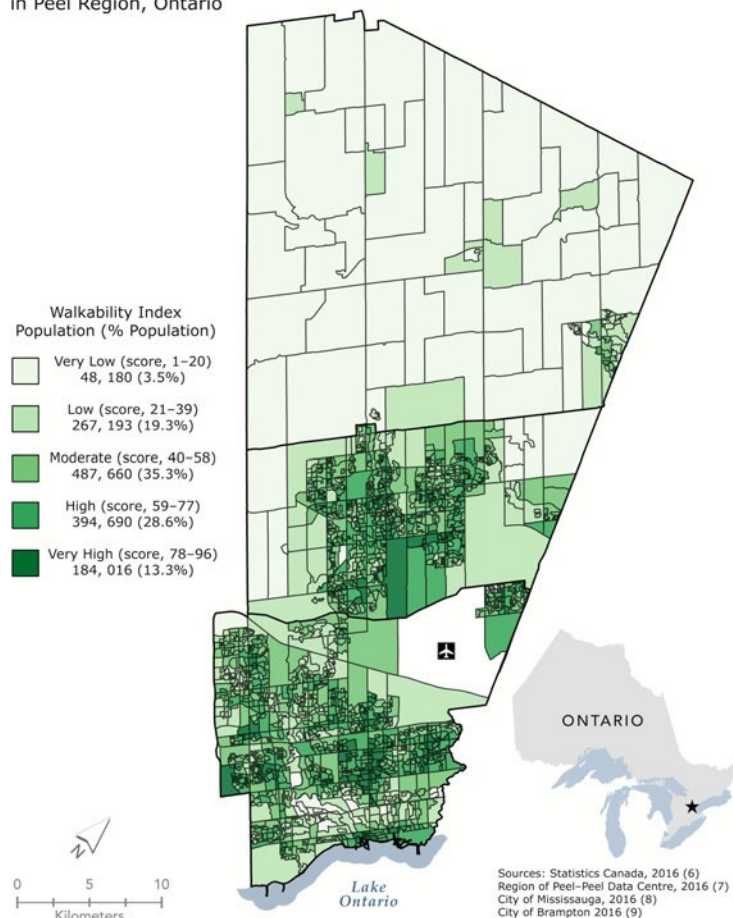
The Peel Walkability Composite Index is used as an evaluation tool and consists of 3 equally weighted components: access to retail and service outlets, access to schools and green spaces, and residential density and diversity. Understanding the capacity of the built environment to facilitate walking for utilitarian purposes (1) allows public health departments to advocate for strategic land use and infrastructure developments that promote an increase in population levels of physical activity.

Background

During the past decade, autocentric suburban regions in Canada experienced tremendous growth. Autocentric built environments discourage active transportation and are linked to chronic disease risk factors (eg, low physical activity levels) (2). Peel Region is a large suburban municipality in Canada with a population of 1.38 million people and an average annual growth rate of 1.3%. To promote healthier communities, the Region of Peel–Public Health partnered with land-use planners on a public health intervention that incorporated policies in Peel's Regional Official Plan (www.peelregion.ca/planning/officialplan), which requires a health assessment on development applications. Evaluation of this intervention relies on the Peel Walkability Composite Index (PWCI).

The PWCI is part of a larger initiative to produce indicators that measure and monitor built environment infrastructure throughout Peel. The PWCI includes indicators that measure built environment features influenced by the policy intervention. Collectively, the indicators operationalize the larger construct of neighborhood walkability and are thus composed into a single evaluation metric (ie, the PWCI).

Well-established walkability indices, including the Physical Activity in Localities and Community Environments (3) and the Neighborhood Quality of Life Study (4), empirically demonstrate the relationship between environmental attributes (ie, residential and retail density, street connectivity, and land-use mix) and physical activity outcomes. A lack of diversity in the attributes used to construct these indices is an acknowledged limitation (3,5). The PWCI was constructed by using a diverse range of objective indicators and was designed to ensure measurement repeatability.

Data Sources and Map Logistics

The PWCI must be repeatable to capture differences in the index score over time. We created the PWCI in 2 stages: 1) we determined the measures to include in the index by using principal component analysis (PCA), and 2) we determined an appropriate weighting scheme to ensure measurement repeatability.

Using PCA on measures of density, diversity, and connectivity is a common approach to creating a walkability index. For the PWCI, we used PCA only to screen and select variables to construct the index. We completed PCA by using the following 14 indicators in SPSS software version 21.0.0.2 (IBM Corporation):

- residential density (Census 2016 [6])
- population density (Census 2016 [6])
- population-plus-employment density (Census 2016 [6] and Municipal Employment Surveys 2015–2016 [7–9])
- proximity of residents to frequent transit (Census 2016 [6] and General Transit Feed Specification 2016 [8,9])
- proximity of residents to green spaces (Census 2016 [6], Active Recreation, Parks, Trails, Peel Data Centre 2016 [7], Parks [8,9] and Conservation Areas [10,11])
- proximity of residents to food stores (Census 2016 [6] and Food Check Peel 2016 [7])
- proximity of residents to schools (Census 2016 [6] and Schools, Peel Data Centre 2016 [7])
- proximity of residents to community and retail services (Census 2016 [6], Municipal Employment Surveys 2015–2016 [7–9], Food Check Peel 2016 [7] and Child Care, Land Marks, Peel Data Centre 2016 [7])
- diversity of land use (Parcel Based Land Use 2016 [7])
- diversity of housing stock (Census 2016 [6])
- intersection density (Single-Line Street Network, Peel Data Centre, 2016 [7])
- percentage of sidewalks with tree canopy (Peel Land Cover, Peel Data Centre 2016 [7] and Sidewalks 2016 [7–9])
- proximity of residents to bicycle networks (Census 2016 [6] and Trails, Peel Data Centre 2016 [7])
- percentage of local roads with speeds below 40 km/hour (Single-Line Street Network, Peel Data Centre 2016 [7])

Indicators had high face validity and were constructed at the level of the Canadian Census dissemination area. We calculated proximity indicators by using 400-m, 800-m, or 1,600-m network distances from points of interest to residential parcels to account for population weighting within dissemination areas. We standardized indicator values by z scores before inclusion in the PCA.

Because of multicollinearity (bivariate correlation scores >0.8), inadequate measures of sampling adequacy (values <0.5 from anti-image correlation matrix), and high levels of nonredundant residuals (>0.05), we removed 6 of the 14 indicators from the PCA: population density, population-plus-employment density, intersection density, percentage of sidewalk with tree canopy, proximity of residents to bicycle networks, and percentage of local roads with speeds below 40 km/hour. We extracted 3 components with eigenvalues greater than 0.95; these components accounted for 62.4% of the total variance. The retained 8 indicators loaded on 3 components: access to retail and service outlets (proximity of residents to grocery stores, +0.85; proximity of residents to community and retail services, +0.85; diversity of land use, +0.57); access to schools and green spaces (proximity of residents to green spaces, +0.80; proximity of residents to schools, +0.74); and residential density and diversity (residential density, +0.86; diversity of housing stock, +0.73, proximity of residents to frequent transit, +0.43).

We constructed the PWCI by averaging the sum of the normalized scores for the standardized indicators that loaded on the extracted principal components for each dissemination area. We normalized the retained 8 indicators (on a scale of 0 to 100) and averaged them by using equal weighting to create the composite index. This process resulted in dissemination area PWCI scores ranging from 1 to 96. These scores provide a 2016 benchmark walkability score. Using equal weighting ensures that component loading values will not influence the capacity of the PWCI to monitor changes in scores over time. We divided the composite index into 5 classes in equal intervals of walkability, from very low (score of 1–20) to very high (score of 78–96). We mapped these classes to illustrate the spatial distribution of walkability in Peel.

Highlights

Many residents of Peel (41.9%) live in areas classified as highly or very highly walkable. These areas are in the downtown cores of cities that have zoning bylaws that encourage higher density and greater mix of land use. Approximately one-third of residents (35.3%) live in a moderately walkable area. These areas are in the inner suburbs, close to city cores, and benefit from proximity to schools and green space. Almost a quarter of residents (22.8%) live in areas with very low or low walkability, along suburban edges. A cluster of areas with very low walkability in the southwest is due to pedestrian barriers, including a highway and the Credit River. Planning policies in these areas encourage very low-density development, contributing to minimal walkable destinations.

Action

The indicator data for the PWCI will be rerun every 5 years, in sequence with the Canadian Census, to monitor changes in the spatial distribution of walkability in Peel. The PWCI is an evidence-informed tool that local elected officials, planners, and public health departments can use to evaluate health-promoting built environment policies and inform future land-use policies. Understanding the spatial distribution of walkable built environments promotes strategic investments in infrastructure that are aimed at increasing levels of physical activity among adults (1).

Acknowledgments

Author Information

Corresponding Author: Maria Mukhtar, MA, Region of Peel, Public Health, Division of Chronic Disease and Injury Prevention, 7120 Hurontario St, Mississauga, Ontario, L5M 2C2. Telephone: 905-791-7800, ext. 2010. Email: maria.mukhtar@peelregion.ca.

Author Affiliations: [1]Division of Chronic Disease and Injury Prevention, Region of Peel–Public Health, Ontario, Canada. [2]Division of Information Management, Region of Peel–Peel Data Centre, Ontario, Canada.

References

1. Mayne DJ, Morgan GG, Jalaludin BB, Bauman AE. The contribution of area-level walkability to geographic variation in physical activity: a spatial analysis of 95,837 participants from the 45 and Up Study living in Sydney, Australia. Popul Health Metr 2017;15(1):38.
2. Tam T. The chief public health officer's report on the state of public health in Canada 2017— designing healthy living. Ottawa (ON): Public Health Agency of Canada; 2017. Report no. 170223. ISSN: 1924-7087.
3. Leslie E, Coffee N, Frank L, Owen N, Bauman A, Hugo G. Walkability of local communities: using geographic information systems to objectively assess relevant environmental attributes. Health Place 2007;13(1):111–22.
4. Frank LD, Sallis JF, Saelens BE, Leary L, Cain K, Conway TL, et al. The development of a walkability index: application to the Neighborhood Quality of Life Study. Br J Sports Med 2010;44(13):924–33.
5. Giles-Corti B, Macaulay G, Middleton N, Boruff B, Bull F, Butterworth I, et al. Developing a research and practice tool to measure walkability: a demonstration project. Health Promot J Austr 2014;25(3):160–6.
6. Statistics Canada. Census profile: age, sex, type of dwelling, families, households, marital status, language, income, immigration and ethnocultural diversity, housing, aboriginal peoples, education, labour, journey to work, mobility and migration, and language of work for Canada, provinces and territories, census divisions, census subdivisions and dissemination areas, 2016census. Statistics Canada Catalogue no. 98-401-X2016044. http://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/download-telecharger/comp/page_dl-tc.cfm?Lang=E. Updated November 29, 2017. Accessed July 10, 2018.

7. Region of Peel. Open data: active recreation, childcare centres, fresh food locations (Food Check Peel), landmarks (points of interest), parcel based land use, parks (Town Of Caledon), sidewalks, single line street network, employment survey (Town Of Caledon), trails. http://opendata.peelregion.ca/a-to-z.aspx. Updated 2016. Accessed January 8, 2018.

8. City of Mississauga. Open data portal: business directory, general transit feed specification (MiWay), parks, sidewalks. http://data.mississauga.ca. Updated 2016. Accessed January 18, 2018.

9. City of Brampton. Open data portal: business listings, general transit feed specification (Brampton Transit), parks, sidewalks. http://geohub.brampton.ca. Updated 2016. Accessed January 8, 2018.

10. Toronto and Region Conservation Authority. Open data portal: TRCA properties. https://data.trca.ca. Updated 2016. Accessed January 8, 2018.

11. Credit Valley Conservation Authority. Conservation areas. Updated 2016. Accessed January 8, 2018.

ORIGINAL RESEARCH

Assessment of Town and Park Characteristics Related to Physical Activity in the Lower Mississippi Delta

Jessica L. Thomson, PhD[1]; Melissa H. Goodman, PhD[1]; Alicia S. Landry, PhD, RD, LDN, SNS[2]

PEER REVIEWED

Summary

What is already known on this topic?

Exercise during and after pregnancy can convey important health benefits for both mother and infant. However, the built environment of rural communities can present challenges for engaging in physical activity.

What is added by this report?

The built environments of the Lower Mississippi Delta towns included in this study lacked key programs, policies, and amenities associated with physical activity among residents.

What are the implications for public health practice?

Before conducting physical activity interventions in rural communities, it may be necessary to first assess the built environments of target populations.

Abstract

Introduction

Our objective was to determine aspects of the built environment that may have contributed to the low levels of physical activity reported in both the gestational and postnatal periods by women participating in a diet and physical activity intervention in the rural Lower Mississippi Delta.

Methods

The built environments of 12 towns were measured by using the Rural Active Living Assessment tools and the Community Park Audit Tool. Correlations between town assessment scores and town size variables were computed by using Kendall τ coefficient.

The street distance from a participant's home address to the nearest park was computed by using network analysis in ArcGIS.

Results

Rural Active Living Assessment scores were low with mean values between 0% (town policy) and 68% (parks and playgrounds) of the highest possible scores. The mean (standard deviation) number of parks per town was 2.6 (3.2), and 55% of the 31 parks were in the 2 largest towns. Most parks (87%) had a single amenity while 1 park had more than 4 amenities. Distance from a participant's home to the nearest park ranged from less than 0.1 to 8.8 miles (mean [standard deviation], 1.2 [1.8]).

Conclusion

These 12 Lower Mississippi Delta towns scored low on assessments of physical environment features and amenities, town characteristics, and programs and policies associated with physical activity in rural communities. To increase the physical activity levels of rural residents, it may be necessary to first improve the built environment in which they live.

Introduction

Lifestyle choices throughout pregnancy can play crucial roles in both the mother's and her unborn child's health. Exercise during pregnancy can ease discomforts such as back pain, boost mood and energy levels, improve sleep, prevent excess weight gain, and increase stamina and muscle strength (1–3). Continuing to exercise after giving birth is essential for strengthening and toning abdominal muscles, boosting energy, promoting better sleep, relieving stress, and losing pregnancy weight gain (4,5). Yet less than one-fourth of pregnant women in the United States meet recommendations for physical activity (PA) (6), and women's participation in exercise programs diminishes after giving birth (1).

From 2013 through 2016, a diet and PA intervention was conducted with pregnant women and their infants residing in the rural Lower Mississippi Delta region of the United States. The Delta

Healthy Sprouts (DHS) Project compared the effect of 2 maternal, infant, and early childhood home visiting curricula on health behaviors of women and their infants (7). Analysis of the project's PA data indicated that baseline PA was low among DHS participants, and positive PA changes were not observed in the gestational or postnatal periods for this cohort of women (8,9). We conducted an observational ancillary investigation, the Delta Neighborhood Physical Activity Study, to determine aspects of these women's built environment that may have contributed to their low levels of PA.

Methods

Study setting

The Delta Neighborhood PA Study included the 12 towns in which DHS participants resided. Parks within these towns were identified 1) by contacting local governing bodies, including city or town hall, mayor's office, town or county office, parks and recreation department, and park commission office; 2) by conducting internet searches; and 3) by study staff members' knowledge of the towns. The study was approved and classified as exempt by the Institutional Review Board of Delta State University (IRB protocol number 16–028). Data collection occurred from August 2016 to September 2017.

Measures

The built environments of the 12 towns were measured by using the Rural Active Living Assessment (RALA) tools: the Program and Policy Assessment (PPA) tool, the Town-Wide Assessment (TWA) tool, and the Street Segment Assessment tool. These 3 observational tools are designed to assess via surveys physical environment features and amenities, town characteristics, and community programs and policies that can affect PA among residents in rural communities (10). Information necessary to complete these surveys was obtained via contact with local governing bodies and school officials, internet searches, staff members' knowledge of the towns, and direct observation. The PPA tool consisted of 20 questions that provided an inventory of each town's programs and policies related to PA. Items included policies for bikeways or walkways, presence of a public recreation department or a private organization offering PA programs, local public transportation, school walking programs, sponsored PA for schoolchildren, a late bus option for children participating in after school activities, and the percentage of children living within 1 mile of their school. The TWA tool consisted of questions about 18 town characteristics and an inventory of 14 recreational amenities that measured each town's physical characteristics on a broad level. Town characteristics included county and town size measures, topography, presence of a town center, street patterns, and

location of schools (elementary, middle, high, and magnet). Magnet schools were added to capture the presence of this type of school in some of the towns. The recreational amenity inventory looked for hiking or walking trails, biking paths, public parks, swimming beaches, public use swimming pools, rivers with watersport access, lakes with watersport access, skate parks, ice skating rinks, roller skating rinks, recreational centers, private fitness facilities, playgrounds, and playing fields or courts. We added "lakes with watersport access" to capture the presence of this amenity in one of the towns. The Street Segment Assessment tool consisted of 28 questions that measured each town's physical characteristics on a detailed (micro) level. Data from the Street Segment Assessment tool is reported elsewhere (11).

Although 2 of the towns exceeded the recommended population size (<10,000) for use of the RALA tools, the surveys were used to assess all of the towns for measurement consistency and comparison among towns. Component and total scores were computed by using scoring algorithms provided in the RALA code and scoring book (10). The higher the assessment scores, the more conducive the town's built environment was to engagement in PA by its residents.

Because the TWA did not provide a detailed assessment of public parks, the Community Park Audit Tool (CPAT) was used to collect specific information regarding features of the towns' public parks (12). A public outdoor space with at least 1 identifiable activity area (eg, green space, playground, field, court, walking trail) was used as the operational definition for a public park. School playgrounds were not included in this assessment. The CPAT survey consists of 4 sections: park information (6 items), access and surrounding neighborhood (11 items), park activity areas (15 items), and park quality and safety (16 items). To avoid redundancy, scoring for the TWA parks component was based on data captured with the CPAT because it contained the same information (and more) as that captured with the TWA tool. We used summary measures to present the data captured with the CPAT because no scoring algorithm is available for this instrument.

For RALA training, senior researchers and research associates (data collectors) watched a recorded web-based seminar that discussed the 3 tools. The webinar is available from the Active Living Research team (10). Senior research members reviewed and discussed the RALA codebook with research associates before data collection and verified sources used for obtaining town information after data collection. Training for use of the CPAT consisted of review and discussion of the user guide by senior research members with research associates and field testing of the

instrument on 3 parks in a nearby town that was not included in the study catchment. Additionally, we randomly selected 10% of the parks for duplicate measurement by senior research members for quality assurance purposes. Discrepancies between duplicate measurements were discussed and resolved.

We re-created the RALA and CPAT instruments as electronic surveys by using Snap Surveys software (version 11.20, Snap Surveys Ltd). All data were collected via tablets loaded with Snap Surveys software and stored on the Snap WebHost, an online, mobile, and secure survey management system.

Statistical analyses

Statistical analyses were performed by using SAS (version 9.4, SAS Institute Inc). We considered results significant at the nominal level of $P < .05$. Kendall τ coefficient, a nonparametric measure of an association's strength and direction, was used to compute correlations between town assessment scores and town size variables because most of the distributions were highly skewed. Analyses of the RALA data sets were conducted both with and without the 2 towns that exceeded the recommended population size (<10,000) for use of the RALA tools. We used the longitude and latitude coordinates of each park's center to mark its location. The street path distance from a participant's home address to the nearest park was computed by using network analysis in ArcGIS (version 10.4, Esri). Three of the 12 towns did not contain any parks. For 2 of the 4 participants living in these 3 towns, their nearest park was in a measured town. For the other 2 participants, their nearest park was in neighboring towns that were not measured because no participants lived in these towns. Although the 2 parks in the nonmeasured towns were used for computing distance to the nearest park, these 2 parks were not measured because we focused on the towns in which participants lived.

Results

Rural Active Living Assessment

Most (63%) DHS participants lived within the boundaries of the 2 largest towns in the intervention. At baseline, none of the 82 DHS participants met the recommended 150 minutes per week of moderate intensity PA (13). However, 5 participants were classified as engaging in moderate amounts of PA, while the other 77 were classified as engaging in low amounts (13). Four of the 5 participants who engaged in moderate amounts of PA lived in the largest town while the fifth participant lived in the third largest town. Mean town population size was 5,319, and mean density was 1,280 residents per square mile (Table 1). Town PPA component and total scores were low, with mean values between 0% (town policy) and 50% (school policy) of the highest possible

scores on the assessment. Town TWA component and total scores also were low, with mean values between 19% (amenity) and 68% (parks and playgrounds) of the highest possible scores on the assessment. Mean scores were lower when the 2 largest towns were excluded from the analyses, with the exception of the town policy score (zero for all towns).

With all towns included, town population was significantly associated with the PPA total score and its school policy component score (Table 2). Town population also was significantly associated with the TWA total score and its school and parks and playgrounds components scores. Town population density was significantly associated with the PPA total score and its school program and school policy component scores as well as the TWA school component score. All correlations were in the positive direction indicating that as town size increased, assessment scores also increased. Town area (square miles) was not significantly associated with any of assessment scores when all towns were included in the analyses. Associations generally increased in magnitude when the 2 largest towns were excluded from the analyses.

Community parks audit

All 31 parks were measured on a weekday in the fall of 2016. Three of the 12 towns did not have any parks, 4 towns had a single park, 2 towns had 3 parks, and the remaining 3 towns had 4, 6, and 11 parks. The mean number (standard deviation [SD]) of parks per town was 2.6 (3.2) and over half of the parks (n = 17) were in the 2 largest towns. Most parks were easy to find (25 [81%]), accessible (29 [94%]), and had at least 6 entry points or an open boundary (18 [58%]) (Table 3). Of the 7 neighborhood concerns observed, the most frequent was no or low street lighting (26%), followed by graffiti (13%) and poorly maintained property (13%). For 14 (45%) of the parks, no neighborhood concerns were observed. A smaller number of concerns were observed for the parks themselves including graffiti (26%), excessive litter (13%), and poor maintenance (7%). In 19 (61%) of the parks, no park concerns were observed. In terms of aesthetics, almost all (94%) of the parks had scattered trees present, although only 8 (26%) parks featured landscaping (eg, flower beds, pruned bushes).

Most parks (87%) had a single activity amenity while 1 park had more than 4 amenities. The most common amenities were open or green spaces (87%) followed by playgrounds (77%) and basketball courts (52%). Amenities were in good condition, ranging from 67% for volleyball courts and swimming pools to 100% for baseball fields, trails, sports fields, tennis courts, and fitness equipment or stations. In terms of features, most parks had lights (87%), trash cans (84%), and benches for sitting (68%). Less than half the parks had restrooms (45%), picnic tables (45%), picnic shelters (42%), grills or fire pits (36%), or drinking fountains (19%). In all

31 parks, the following features were absent: map of the park, public transit stop near park, bike racks, bike routes bordering park, splash pads, off-leash dog parks, family restrooms, and baby changing stations in the restrooms.

Street path distance from a participant's home to the nearest park ranged from less than 0.1 to 8.8 miles (mean [SD], 1.2 [1.8]). For the 5 participants who engaged in moderate amounts of PA at baseline, mean (SD) distance from home to the nearest park was 1.0 (1.3) miles, compared with 1.2 (1.8) miles for the participants who engaged in low amounts of PA at baseline. Additionally, 3 of the 5 participants who engaged in moderate amounts of PA lived close (one-half mile or less) to a park. In comparison, 47% of participants who engaged in low amounts of PA lived within one-half mile of a park.

Discussion

We presented physical activity–related characteristics of the towns in which DHS participants lived as well as features and amenities of the parks in these towns. Results indicate that the built environment may have played a role in the low levels of PA observed in this cohort of rural, Southern, primarily African American women. On average, assessment scores for programs, policies, features, and amenities related to PA were low for the towns in which these women lived. Mean PPA and TWA scores for towns in our study (26 and 32) were lower than those reported in studies assessing rural towns in the Deep South (55 and 59), Appalachia region of North Carolina (not assessed and 50), Washington Latino communities (69 and 63), and Hawaii (39 and 67) (14–17). Similar to findings in our study, all but 1 Deep South community scored zero on the PPA town policy (14). However, all towns in the previous 4 studies had recreational amenities (14–17), while 2 towns in our study had no recreational amenities. Furthermore, no town in our study had public transportation systems. Results from a systematic review of correlates of PA suggest that recreational facilities must be present and either close to an person's residence with safe walking routes or accessible by public transportation to promote participation in PA at such facilities (18).

In our study, towns with higher populations had higher TWA total and parks and playgrounds component scores, indicating that larger towns were more conducive to engagement in PA by their residents. A similar relationship between town size and TWA scores was present in the Washington towns, but not in the Deep South, North Carolina Appalachian, or Hawaiian towns (14–17). In our study, 4 of the 5 participants who engaged in moderate amounts of PA at baseline lived in the most populated town with the highest parks and playgrounds component score (25) and the second highest TWA score (60). The other participant lived in the third

most populated town also with the highest parks and playgrounds component score (25) and the third highest TWA score (52). Results should be interpreted cautiously because of the small number of participants who engaged in moderate amounts of PA.

Almost half of DHS participants lived within one-half mile walking distance of a park and approximately three-fourths of the parks contained playgrounds, an amenity associated with park-based PA in women (19). However, only one-fourth of the parks had bordering sidewalks, which suggests that walking routes to parks lacked this safety feature. In a telephone survey conducted with 1,176 South Carolina residents, more African American women reported greater maintenance of sidewalks as a correlate of PA than white women did (20). Likewise, the presence of sidewalks and feeling safe and secure from crime and traffic were closely linked to the decision to be physically active in minority women (21). Hence, walking routes to and around public parks may have been a contributing factor to the low levels of PA observed among DHS participants. Potentially compounding the issue of safe walking routes is aesthetics, which also was identified as an important environmental design aspect in the systematic review of correlates and determinants of PA in adults and children (18). In our study, all but 2 of the 31 parks had scattered trees present, but only 8 featured landscaping. Thus, the lack of aesthetic features in most of the parks may have at least partly discouraged engaging in PA in the parks in this cohort of women.

The built environment likely played a role in the low levels of PA observed in these women; however, the influence of their personal health characteristics bears mentioning. During pregnancy, PA levels are known to decrease (22), probably because of anatomic and physiologic changes that occur. Additionally, two-thirds of the women in our study were overweight or obese before becoming pregnant and scored relatively low for PA self-efficacy at baseline (13). Overweight classification and lower self-efficacy for participating in PA have been negatively related to PA levels (18).

Strengths of this study are the use of multiple validated and objective tools to assess town and park characteristics and exploration of potential associations between study participants' PA levels with town and park measures and features. The population studied also is a strength because rural, Southern, African American adults are at increased risk for inadequate amounts of PA (23–25). A limitation is the small sample sizes for both study participants and towns, which may have limited the ability to find significant associations in the data. Additionally, the nonrandom selection of towns and parks limits the generalizability of the study's results.

The Lower Mississippi Delta towns included in this study generally scored low on assessments of physical environment features and amenities, town characteristics, and community programs and

policies that can affect PA among residents in rural communities. Furthermore, although most DHS participants lived close to a park, the parks lacked features known to be associated with PA, such as safe walking routes and aesthetics. To increase PA levels of rural residents, it may be necessary to first improve the built environment in which they live.

Acknowledgments

Author Information

Corresponding Author: Jessica L. Thomson, PhD, USDA Agricultural Research Service, 141 Experiment Station Road, Stoneville, MS 38776. Telephone: 225-892-3662. Fax: 662-686-5309. Email: jessica.thomson@ars.usda.gov.

Author Affiliations: [1]US Department of Agriculture, Agricultural Research Service, Stoneville, Mississippi. [2]Department of Family and Consumer Sciences, University of Central Arkansas, Conway, Arkansas.

References

1. American College of Obstetricians and Gynecologists. ACOG committee opinion no. 650: physical activity and exercise during pregnancy and the postpartum period. Obstet Gynecol 2015;126(6):e135–42.
2. Hollenbach D, Broker R, Herlehy S, Stuber K. Non-pharmacological interventions for sleep quality and insomnia during pregnancy: a systematic review. J Can Chiropr Assoc 2013;57(3):260–70.
3. Nascimento SL, Surita FG, Cecatti JG. Physical exercise during pregnancy: a systematic review. Curr Opin Obstet Gynecol 2012;24(6):387–94.
4. Liu J, Smith MG, Dobre MA, Ferguson JE. Maternal obesity and breast-feeding practices among white and black women. Obesity (Silver Spring) 2010;18(1):175–82.
5. Roy BA. Postpartum exercise. ACSMs Health Fit J 2014; 18(6):3–4.
6. Evenson KR, Wen F. National trends in self-reported physical activity and sedentary behaviors among pregnant women: NHANES 1999–2006. Prev Med 2010;50(3):123–8.
7. Thomson JL, Tussing-Humphreys LM, Goodman MH. Delta Healthy Sprouts: a randomized comparative effectiveness trial to promote maternal weight control and reduce childhood obesity in the Mississippi Delta. Contemp Clin Trials 2014; 38(1):82–91.
8. Thomson JL, Tussing-Humphreys LM, Goodman MH, Olender SE. Physical activity changes during pregnancy in a comparative impact trial. Am J Health Behav 2016; 40(6):685–96.
9. Thomson JL, Tussing-Humphreys LM, Goodman MH, Landry AS. Enhanced curriculum intervention did not result in increased postnatal physical activity in rural, Southern, primarily African American women. Am J Health Promot 2018;32(2):464–72.
10. Hartley D, Yousefian A, Umstattd R, Hallam J, Economos C, Hyatt R, et al.The Rural Active Living Assessment (RALA) Tools. 2009. https://activelivingresearch.org/rural-active-living-assessment-rala-tools. Accessed July 23, 2018.
11. Thomson JL, Goodman MH, Landry AS. Assessment of neighborhood street characteristics related to physical activity in the Lower Mississippi Delta. Health Promot Perspect 2019; 9(1):24–30.
12. Kaczynski A, Stanis S, Besenyi G. Community Park Audit Tool (CPAT). 2012. https://activelivingresearch.org/community-park-audit-tool-cpat. Accessed July 23, 2018.
13. Thomson JL, Tussing-Humphreys LM, Goodman MH, Olender S. Baseline demographic, anthropometric, psychosocial, and behavioral characteristics of rural, Southern women in early pregnancy. Matern Child Health J 2016; 20(9):1980–8.
14. Robinson JC, Carson TL, Johnson ER, Hardy CM, Shikany JM, Green E, et al. Assessing environmental support for better health: active living opportunity audits in rural communities in the southern United States. Prev Med 2014;66:28–33.
15. Hege A, Christiana RW, Battista R, Parkhurst H. Active living in rural Appalachia: using the Rural Active Living Assessment (RALA) tools to explore environmental barriers. Prev Med Rep 2017;8:261–6.
16. Perry CK, Nagel C, Ko LK, Duggan C, Linde S, Rodriguez EA, et al. Active living environment assessments in four rural Latino communities. Prev Med Rep 2015;2:818–23.
17. Hafoka SF. Assessing the active living environment in three rural towns with a high proportion of Native Hawaiians and other Pacific Islanders. Health Promot Perspect 2017; 7(3):134–9.
18. Bauman AE, Reis RS, Sallis JF, Wells JC, Loos RJF, Martin BW, et al. Correlates of physical activity: why are some people physically active and others not? Lancet 2012; 380(9838):258–71.

19. Kaczynski AT, Besenyi GM, Stanis SAW, Koohsari MJ, Oestman KB, Bergstrom R, et al. Are park proximity and park features related to park use and park-based physical activity among adults? Variations by multiple socio-demographic characteristics. Int J Behav Nutr Phys Act 2014;11(1):146.

20. Sharpe PA, Granner ML, Hutto BE, Wilcox S, Peck L, Addy CL. Correlates of physical activity among African American and white women. Am J Health Behav 2008;32(6):701–13.

21. Committee on Physical Activity, Health, Transportation, and Land Use. Does the built environment influence physical activity? Washington (DC): Institute of Medicine; 2005. TRB special report 282.

22. Borodulin K, Evenson KR, Herring AH. Physical activity patterns during pregnancy through postpartum. BMC Womens Health 2009;9(1):32.

23. An R, Xiang X, Yang Y, Yan H. Mapping the prevalence of physical inactivity in U.S. states, 1984–2015. PLoS One 2016;11(12):e0168175.

24. Pleis JR, Lethbridge-Cejku M. Summary health statistics for U.S. adults: National Health Interview Survey, 2005. Vital Health Stat 10 2006;(232):1–153.

25. Patterson PD, Moore CG, Probst JC, Shinogle JA. Obesity and physical inactivity in rural America. J Rural Health 2004;20(2):151–9.

Tables

Table 1. Characteristics of Towns in the Delta Neighborhood Physical Activity Study, 2016–2017

Characteristic	Range of Possible Scores	All Towns (n = 12)				Largest Towns[a] Excluded (n = 10)		
		Mean (SD)	Median	Min[b]	Max	Mean (SD)	Median	Max
Population[c]		5,319 (9,739)	1,743	337	34,400	1,709 (1,419)	1,321	4,481
Area (square miles)[c]	NA	4 (7)	1	1	27	2 (1)	1	4
Density (per square mile)[c]		1,280 (701)	1,262	461	2,519	1,188 (706)	1,178	2,519
Program and Policy Assessment								
Town program score[d,e]	0–30	6 (9)	0	0	26	3 (6)	0	14
Town policy score[d,f]	0–10	0 (0)	0	0	0	0 (0)	0	0
School program score[d,g]	0–30	5 (9)	0	0	30	3 (5)	0	10
School policy score[d,h]	0–30	15 (13)	15	0	30	14 (13)	15	30
Total score[d,i]	0–100	26 (25)	29	0	86	19 (18)	22	40
Town-Wide Assessment								
School count	NA	5 (7)	3	0	23	3 (3)	3	9
School score[d,j]	0–21	6 (6)	5	0	15	5 (6)	2	15
Amenity type count[d,k]	NA	4 (2)	4	0	8	3 (2)	3	5
Amenity total count[d,l]	NA	10 (11)	6	0	37	5 (5)	4	12
Amenity score[d,m]	0–53	10 (9)	9	0	29	7 (6)	8	18
Parks and playgrounds score[d,n]	0–25	17 (11)	23	0	25	15 (11)	19	25
Total score[d,o]	0–99	32 (21)	36	0	62	27 (18)	29	52

Abbreviations: Min, minimum; Max, maximum; NA, not applicable; SD, standard deviation.
[a] Excluded 2 towns with populations exceeding recommended size (<10,000) for Rural Active Living Assessment tools.
[b] Minimum values are the same for both sets of towns.
[c] Source: www.factfinder.census.gov.
[d] Higher scores indicate the town's built environment was more conducive to physical activity.
[e] Composed of 6 items concerning public and private recreation.
[f] Composed of 1 item concerning bikeways/walkways required for new infrastructure.
[g] Composed of 2 items concerning public access to recreation facilities and late bus options.
[h] Composed of 3 items concerning walking and safe routes to school and sponsored physical activity programs.
[i] Sum of scores for town program, town policy, school program, and school policy.
[j] Composed of 4 items concerning walkability to schools (elementary, middle, high, and magnet).
[k] Count of different types of amenities (each of 17 types counted only once).
[l] Count of total number of amenities (may include multiples of same type).
[m] Composed of 13 items concerning location of amenities from town centers.
[n] Composed of 4 items concerning location of parks, playgrounds, and sports fields and courts from town centers.
[o] Sum of scores for school, amenity, and parks and playgrounds.

Table 2. Associations Among Town Size Measures and Rural Active Living Assessment Scores, Delta Neighborhood Physical Activity Study, 2016–2017

Town Size	Statistic	Program and Policy Assessment Scores[a]				Town-Wide Assessment Scores			
		Town Program	School Program	School Policy	Total	School	Amenity	Parks and Playgrounds	Total
All towns included (n = 12)									
Population	KTC	0.30	0.49	0.54	0.58	0.77	0.32	0.76	0.72
	P value	.30	.09	.04	.03	.004	.22	.004	.004
Area (square miles)	KTC	0.31	0.00	0.07	0.11	0.36	0.07	0.51	0.35
	P value	.33	>.99	.81	.72	.23	.82	.08	.22
Population density (per square mile)	KTC	0.00	0.62	0.60	0.63	0.56	0.22	0.37	0.45
	P value	>.99	.03	.03	.02	.03	.40	.16	.07
Largest towns[b] excluded (n = 10)									
Population	KTC	0.56	0.48	0.53	0.60	0.68	0.55	0.79	0.78
	P value	.02	.05	.03	.01	.004	.02	.001	<.001
Area (square miles)	KTC	0.64	0.20	0.24	0.33	0.42	0.46	0.64	0.56
	P value	.02	.46	.37	.20	.10	.07	.01	.02
Population density (per square mile)	KTC	0.15	0.40	0.53	0.50	0.61	0.29	0.39	0.47
	P value	.56	.11	.03	.03	.009	.21	.10	.03

Abbreviation: KTC, Kendall τ correlation.
[a] Town policy was not included since all towns scored 0 points on this component.
[b] Excluded 2 towns with populations exceeding recommended size (<10,000) for Rural Active Living Assessment.

Table 3. Characteristics of Parks (N = 31) Included in the Delta Neighborhood Physical Activity Study, 2016–2017

Characteristic	n (%)
Park Characteristics	
Easy to find	25 (80.6)
Accessible	29 (93.5)
Points of entry	
1	3 (9.7)
2–5	10 (32.3)
≥6 or open boundary	18 (58.1)
Parking type	
Lot	18 (58.1)
On street	14 (45.2)
None	1 (3.2)
Bordering sidewalks[a]	8 (25.8)
Bordering traffic signs[b]	25 (80.6)
Main land use	
Residential	22 (71.0)
Institutional (school)	1 (3.2)
Commercial	4 (12.9)
Natural	4 (12.9)
Neighborhood concerns[c]	
No or low street lighting	8 (25.8)
Graffiti	4 (12.9)
Poorly maintained property	4 (12.9)
Excessive litter	3 (9.7)
Heavy traffic	3 (9.7)
Vacant/abandoned buildings	2 (6.5)
Unfavorable buildings	2 (6.5)
None	14 (45.2)
Park concerns[d]	
Graffiti	8 (25.8)
Excessive litter	4 (12.9)
Poor maintenance	2 (6.5)
None	19 (61.3)

[a] All sidewalks were useable, but only 5 of the 8 parks had curb cuts or ramps.
[b] 24 parks had stop signs, 1 park had a stop light, and no parks had crosswalks.
[c] None of the surrounding neighborhoods had vandalism, excessive noise, lack of eyes on the street, or threatening persons or behavior.
[d] None of the parks had vandalism, excessive noise or animal waste, threatening persons or behavior, or dangerous spots.
[e] 24 parks had playgrounds, but 1 park had 2 playground areas so denominator is 25 for playground features.
[f] 12 parks had baseball fields, but 2 parks had 2 baseball fields so denominator is 14 for field condition.
[g] 8 parks had trails, but 1 park had 2 trails so denominator is 9 for trail features.
[h] Football or soccer fields; 4 parks had sports fields, but 1 park had 2 sport fields so denominator is 5 for field condition.
[i] Included portable toilets.

(continued)

Table 3. Characteristics of Parks (N = 31) Included in the Delta Neighborhood Physical Activity Study, 2016–2017

Characteristic	n (%)
Aesthetic features	
Scattered trees	29 (93.5)
Dense trees	10 (32.3)
Landscaping	8 (25.8)
Water	6 (19.4)
Historical/educational	4 (12.9)
Activity Area Characteristics	
Total number	
1	27 (87.1)
2	1 (3.2)
3	2 (6.5)
≥4	1 (3.2)
Open or green space	27 (87.1)
Playground[e]	24 (77.4)
Good condition	22 (88.0)
Colorful equipment	22 (88.0)
Shade cover ≥25%	11 (44.0)
Bench	18 (72.0)
Separation from road	11 (44.0)
Basketball court	16 (51.6)
Good condition	13 (81.3)
Baseball field[f]	12 (38.7)
Good condition	14 (100.0)
Trail[g]	8 (25.8)
Good condition	9 (100.0)
Connected to activity areas	8 (88.9)
Bench for sitting	4 (44.4)
Sport field[h]	4 (12.9)
Good condition	5 (100.0)
Tennis court	4 (12.9)
Good condition	4 (100.0)
Swimming pool	3 (9.7)

[a] All sidewalks were useable, but only 5 of the 8 parks had curb cuts or ramps.

[b] 24 parks had stop signs, 1 park had a stop light, and no parks had crosswalks.

[c] None of the surrounding neighborhoods had vandalism, excessive noise, lack of eyes on the street, or threatening persons or behavior.

[d] None of the parks had vandalism, excessive noise or animal waste, threatening persons or behavior, or dangerous spots.

[e] 24 parks had playgrounds, but 1 park had 2 playground areas so denominator is 25 for playground features.

[f] 12 parks had baseball fields, but 2 parks had 2 baseball fields so denominator is 14 for field condition.

[g] 8 parks had trails, but 1 park had 2 trails so denominator is 9 for trail features.

[h] Football or soccer fields; 4 parks had sports fields, but 1 park had 2 sport fields so denominator is 5 for field condition.

[i] Included portable toilets.

(continued)

Table 3. Characteristics of Parks (N = 31) Included in the Delta Neighborhood Physical Activity Study, 2016–2017

Characteristic	n (%)
Good condition	2 (66.7)
Volleyball court	3 (9.7)
Good condition	2 (66.7)
Fitness equipment or station	2 (6.5)
Good condition	2 (100.0)
Feature Characteristics	
Lights	27 (87.1)
Trash can	26 (83.9)
Overflowing	0 (0.0)
Restroom[i]	14 (45.2)
Good condition	12 (85.7)
Bench for sitting	21 (67.7)
Good condition	19 (90.5)
Drinking fountain	6 (19.4)
Good condition	2 (33.3)
Picnic table	14 (45.2)
Good condition	14 (100.0)
Picnic shelter	13 (41.9)
Grill/fire pit	11 (35.5)
Animal rules posted	11 (35.5)
Interior road	4 (12.9)
Recycling container	1 (3.2)
Vending machine	1 (3.2)

[a] All sidewalks were useable, but only 5 of the 8 parks had curb cuts or ramps.

[b] 24 parks had stop signs, 1 park had a stop light, and no parks had crosswalks.

[c] None of the surrounding neighborhoods had vandalism, excessive noise, lack of eyes on the street, or threatening persons or behavior.

[d] None of the parks had vandalism, excessive noise or animal waste, threatening persons or behavior, or dangerous spots.

[e] 24 parks had playgrounds, but 1 park had 2 playground areas so denominator is 25 for playground features.

[f] 12 parks had baseball fields, but 2 parks had 2 baseball fields so denominator is 14 for field condition.

[g] 8 parks had trails, but 1 park had 2 trails so denominator is 9 for trail features.

[h] Football or soccer fields; 4 parks had sports fields, but 1 park had 2 sport fields so denominator is 5 for field condition.

[i] Included portable toilets.

ORIGINAL RESEARCH

Understanding the Density and Distribution of Restaurants in Los Angeles County to Inform Local Public Health Practice

Lauren N. Gase, PhD, MPH[1,2]; Gabrielle Green, MPP[1]; Christine Montes, MPH[1]; Tony Kuo, MD, MSHS[1]

PEER REVIEWED

Summary

What is already known on this topic?

Studies have found associations between neighborhood sociodemographics and restaurant density and restaurant type, often categorizing restaurants as "fast food" or "full service."

What is added by this report?

This study provides insight into the potential reach of program or policy strategies that target chain restaurants. To inform local public health planning, we examined where restaurants, including chain restaurants, were located in Los Angeles County, California.

What are the implications for public health practice?

Results highlight the limited reach of strategies targeting chain restaurants. Other jurisdictions can build on the methods used in our study to enhance understanding of their own local landscape.

Abstract

Introduction

To describe the potential reach of restaurant-based strategies that seek to improve the healthfulness of menu options, it is important to understand the local restaurant environment, including the extent to which restaurants subject to policy mandates are located in communities disproportionately affected by diet-related diseases.

Methods

This cross-sectional study examined the restaurant environment in Los Angeles County, a large jurisdiction with diverse geographic and socioeconomic characteristics, specifically 1) the number and characteristics of restaurants; 2) the association between neighborhood sociodemographics and restaurant density; and 3) the association between neighborhood sociodemographics and restaurant characteristics, including chain status (large chain, small chain, independent restaurant). Data sources were 1) industry data on restaurant location and characteristics (N = 24,292 restaurants) and 2) US Census data on neighborhood sociodemographics (N = 247 neighborhoods). We conducted descriptive and bivariate analyses at the restaurant and neighborhood level.

Results

Countywide, only 26.5% of all restaurants were part of a large chain (a chain with ≥20 locations). We found positive associations between restaurant density and neighborhood proportions of non-Hispanic white residents and residents with more than a high school education. We found limited support to suggest a greater density of large chains in neighborhoods with lower socioeconomic status.

Conclusion

Results highlight the potentially limited reach of strategies targeting chain restaurants and point to the importance of including small chain restaurants and independent restaurants in public health efforts to improve the healthfulness of restaurants. Understanding where restaurants are in relation to priority populations is a critical step to planning strategies that address diet-related disparities.

Introduction

As consumers purchase more meals away from home than previously, strategies to increase the healthfulness of food and beverages offered at restaurants have garnered increased attention (1,2). Examples of restaurant-focused policies include menu labeling (3), ordinances banning restaurants from giving away free toys with children's meals unless the meal meets nutritional guidelines (4), and ordinances mandating that restaurants serve healthy

beverages as the default option (5). Examples of voluntary initiatives include increasing the number of healthy options, offering smaller portion sizes, or offering attractive pricing for healthy options (6,7).

Both mandatory policies and voluntary initiatives have typically targeted chain restaurants: restaurants that do business under the same name and offer substantially the same menu items. National menu labeling policy included in the Patient Protection and Affordable Care Act applies only to restaurants with 20 or more locations (3). Voluntary initiatives also frequently aim to engage, and often have high rates of participation from, chain restaurants (8,9). Chains have generally been the focus of restaurant-based strategies for 3 reasons. First, chains are believed to pose a higher risk to consumers than other types of restaurants: chains usually offer foods that have minimal nutritional value, are widely accessible, and/or have many repeat customers (10). Second, the cost of implementing initiatives is lower for chains than for other types of restaurants, because chains can more easily absorb the costs associated with healthy eating strategies, such as the costs of nutritional analyses (11,12). Third, chains are better equipped than other types of restaurants to adhere to the requirements of initiatives, such as following standardized recipes (13).

To advance restaurant-based strategies at the local level, it is important to understand where restaurants are located, including the extent to which restaurants subject to policy mandates are in communities disproportionately affected by diet-related diseases. Although previous studies examined neighborhood sociodemographics associated with restaurant density, most categorized restaurants as "fast food" or "full service" as opposed to "chain" or "nonchain" (14–16). Although useful, such studies provide little insight into the potential reach of program or policy strategies that target chain restaurants. To inform local public health planning, we aimed to describe where restaurants, especially chain restaurants, are located in Los Angeles County, California.

Methods

This cross-sectional study, conducted in late 2016 and early 2017, sought to answer the following questions: 1) What is the current number and what are the characteristics of restaurants in Los Angeles County?; 2) What neighborhood sociodemographic characteristics are associated with a greater presence of restaurants (in general)?; and 3) What neighborhood sociodemographic characteristics are associated with restaurant characteristics, in particular the density of chain restaurants? Although our primary goal was to inform local decision making, we hope that other jurisdictions can build on the methods used in our study to enhance understanding of their own local landscape.

We used data from 2 sources: industry data on restaurant characteristics and US Census data on neighborhood characteristics. Information on restaurant characteristics was provided by a market research firm that tracks restaurant industry trends nationally. The firm defines a restaurant as a location whose primary purpose is to serve food away from home on an open, commercial basis. Twenty market research staff members at this market research firm work daily on data validation through a multitiered strategy that includes monthly searches for social media reviews, auto- mated telephone calls to check a restaurant's telephone connectiv- ity, and direct outreach though email and telephone surveys.

The primary restaurant characteristic of interest was chain status. We determined chain status on the basis of the number of restaurant locations that conducted business using the same name, that offered similar menu items, and whose link could be verified via internet or telephone. To understand the potential reach of strategies at the national and local level, we created 2 variables for chain status, one based on the number of locations nationally and one based on the number of locations in Los Angeles County. We classified restaurants as independent (single location), small chain (2–19 locations), or large chain (≥20 locations), in accordance with previous research and policy scope (eg, the menu labeling policy included in the Patient Protection and Affordable Care Act) (17,18).

We also categorized restaurants according to industry market segment and cuisine type. Industry market segment was coded by the market research firm as 1) quick service (patrons order at counter; meals typically under $10), 2) fast casual (patrons order at counter; slightly higher price point than quick service), 3) midscale dining (offers sit-down/full table service; typically does not serve alcohol; entrée prices generally ≤$20), 4) casual dining (offers sit-down/full table service; typically serves alcohol; entrée prices generally $15–$25), or 5) fine dining (entrée prices are generally >$25). These market segments are recognized industry standards, allowing for comparison across geographic regions. Restaurant cuisine type, coded by the market research firm on the basis of the primary type of food served by the restaurant, was categorized as American/Southern (bar and grill, diner, sports bar, brew pub), Asian (Chinese, Japanese, Korean, Thai, other Asian), Latino (Mexican, South American), coffee/bakery/dessert (bagel, coffee shop, ice cream, smoothie, donut), burger, pizza, sandwich/deli, European (Italian, Mediterranean, French, other European), or other (African, Caribbean, Indian, seafood, mixed ethnicity, steakhouse, barbecue).

We selected neighborhood characteristics on the basis of population groups that tend to be disproportionately affected by diet-related diseases. We collected the following census tract–level data from the 2010–2014 American Community Survey 5-Year Estim-

ates (19–23): 1) percentage of non-Hispanic white residents; 2) percentage of the population aged 25 or older with more than a high school education; 3) the percentage of the population with income below the poverty level in the last 12 months; 4) median household income in the last 12 months, in 2014 inflation-adjusted dollars; and 5) total population.

Data cleaning, geocoding, and aggregation

The original data set contained 24,884 restaurants, as of September 19, 2016. Staff members of the Los Angeles County Department of Public Health (DPH) conducted a 3-stage cleaning and geocoding process. First, they flagged possible duplicate records on the basis of similarities in restaurant name, street address, and/or telephone number. Second, they flagged possible unidentified chain locations on the basis of similarities in restaurant name and telephone number. All flags were investigated via internet search. Third, they geocoded restaurant addresses in ArcMap version 10.3.1 (ESRI) by using Los Angeles County's Countywide Address Management System address locator (24). After cleaning, 24,292 restaurants remained in the final restaurant data set.

We defined Los Angeles County neighborhoods according to the *Los Angeles Times'* Mapping L.A. project, which defines neighborhoods (a city, a community within a city, or an unincorporated area of the county) that are meaningful to residents and align with census-tract boundaries (25). We used the ArcMap Dissolve tool to aggregate census data to the neighborhood level. We constructed the following neighborhood-level measures of restaurant density: 1) total number of restaurants, 2) total number of restaurants per 1,000 residents (number of restaurants divided by the neighborhood population, multiplied by 1,000), 3) percentage of restaurants that were large chains (number of restaurants categorized as large chains [≥20 locations, based on the number of locations nationally or in Los Angeles County] divided by the number of total restaurants), 4) percentage of restaurants in each industry market segment (eg, quick service, fast casual), and 5) percentage of restaurants serving each type of cuisine (eg, American/Southern, Asian).

We excluded 8 neighborhoods that had fewer than 1,000 residents because the neighborhoods entirely or primarily were non-neighborhood–type complexes (eg, theme park, recreation area, health care campus). We excluded as outliers 3 neighborhoods that had more than 15 restaurants per 1,000 residents because they were large business or entertainment districts where daytime populations greatly exceed residential populations. Thus, we included 247 neighborhoods in the final neighborhood data set. In a sensitivity analysis of the relationship between restaurant chain density and neighborhood sociodemographic characteristics, we excluded 32 neighborhoods that had fewer than 10 restaurants.

Data analysis

We conducted descriptive and bivariate analyses at the restaurant level to examine the number and characteristics of restaurants countywide. We conducted descriptive and linear regression analyses at the neighborhood level to examine the association between neighborhood sociodemographic characteristics and 1) the number of restaurants per 1,000 residents and 2) the percent- age of restaurants categorized as large chains. We conducted all analyses by using Stata version 14.1 (StataCorp LP). All materials were reviewed and approved by the Los Angeles County Depart- ment of Public Health Institutional Review Board.

Results

The final sample consisted of 24,292 restaurants in Los Angeles County. On the basis of the number of locations nationally, we classified 26.5% of restaurants as large chain, 11.3% as small chain, and 62.2% as independent (Table 1). The 6,430 restaurant locations categorized as large chains represented 278 restaurant brands. On the basis of the number of locations in Los Angeles County, we classified 21.2% restaurants as large chain. The 5,145 locations categorized as large chains represented only 59 brands. These 59 brands represented 80% of all large chain restaurants in the county.

Large chain restaurants were more likely than other types of restaurants to be quick service or fast casual. On the basis of the number of locations nationally, 66.2% of large chains were quick service and 21.5% were fast casual, compared with 29.9% and 5.7% of independent restaurants that were quick service or fast casual, respectively. Large chains were most likely to serve coffee/bakery/dessert (22.4%), burger (19.7%), and sandwich (14.2%) cuisines. Independent restaurants were most likely to serve Asian (23.4%), American/Southern (21.5%), or Latino (17.0%) cuisines.

Although most large chains were classified as quick service or fast casual restaurants, not all quick service and fast casual restaurants were classified as large chains. Among quick service restaurants (n = 9,571), less than half (44.5%) were classified as large chains on the basis of the number of locations nationally.

Neighborhood sociodemographic characteristics and restaurant density

The average number of restaurants, by neighborhood, was 94.4 (standard deviation [SD], 117.5) (Table 2). The number of restaurants ranged from 0 (4 neighborhoods) to more than 500 restaurants (4 neighborhoods); the median was 58 (interquartile range, 101). The average number of restaurants was 2.3 (SD, 1.8) per 1,000 residents but ranged from 0 to 11.6. The median was 1.9 (interquartile range, 1.6).

Neighborhood education level, racial/ethnic composition, and poverty were significantly associated with the number of restaurants in the neighborhood. On average, for every 1-point increase in the percentage of residents with more than a high school education, the number of restaurants per 1,000 residents would be expected to increase by 2.5 (95% confidence interval [CI], 1.5–3.5). On average, for every 1-point increase in the percentage of non-Hispanic white residents in the neighborhood, the number of restaurants per 1,000 residents would be expected to increase by 1.9 (95% CI, 1.1–2.7). On average, for every 1-point increase in the percentage of residents below the poverty level, the number of restaurants per 1,000 residents would be expected to decrease by 3.3 (95% CI, −5.5 to −1.1). Median household income was not significantly associated with restaurant density.

Neighborhood sociodemographic characteristics and restaurant characteristics

Chain status. On the basis of the number of locations nationally, the average density of large chain restaurants by neighborhood was 26.5% (SD, 15.0%), although the density of large chain restaurants ranged from 0% to 100% across neighborhoods. On the basis of the number of locations in Los Angeles County, the average density of large chain restaurants was 21.7% (SD, 12.4%); density ranged from 0% to 66.7%.

When we examined the number of locations of restaurants nationally, we found no significant associations between neighborhood sociodemographic characteristics and chain density. Neighborhoods with a greater density of large chains tended to have a lower percentage of non-Hispanic white residents and a lower percentage of residents with more than a high school education. When we examined the number of restaurant locations in Los Angeles County, we found significant associations between a greater density of large chain restaurants and a lower percentage of non-Hispanic white residents and a lower percentage of residents with more than a high school education (Table 3). Results did not substantively change in magnitude or significance when we considered only neighborhoods with 10 or more restaurants.

Restaurants by industry market segment and cuisine. We found high correlations between the proportion of residents with more than a high school education and the proportion of 1) quick-service restaurants ($r = -0.42$) and 2) fine-dining restaurants ($r = 0.46$). In neighborhoods with the lowest quartile of residents with more than a high school education, roughly half of the restaurants were quick service (mean = 0.49; SD, 0.13), and in neighborhoods with the highest quartile of residents with more than a high school education, approximately one-third (mean = 0.32; SD =

0.14) of restaurants were quick service. We found correlations and proportions of similar magnitude between these 2 restaurant segments and the percentage of non-Hispanic white residents.

For cuisine type, we found high correlations between the proportion of non-Hispanic white residents and the proportion of restaurants that served European cuisine ($r = 0.69$) and Latino cuisine ($r = -0.51$). The proportion of residents with more than a high school education was strongly correlated with the proportion of restaurants that served Latino cuisine ($r = -0.69$), European cuisine ($r = 0.58$), coffee/bakery/dessert cuisine ($r = 0.44$), and burger cuisine ($r = -0.43$).

Discussion

Our study suggests that a limited proportion of restaurants in Los Angeles County are part of a large chain. A policy that targets large chains (≥20 locations nationally) would affect only about one-quarter of all restaurants in Los Angeles County. Estimates of chain density in our study are lower than national estimates, which suggest that 40% of all restaurants are part of a large chain (3). Our study highlights the potentially limited reach of policy strategies (such as menu labeling) that would target chain restaurants in Los Angeles County and point to the importance of reaching out to and collaborating with small chain restaurants and independent restaurants as part of a comprehensive local strategy to improve the healthfulness of restaurants. The importance of targeting such restaurants is underscored by recent work demonstrating that non-chain restaurants offer high-calorie food, on par with their chain counterparts (26). Independent and small chain restaurants may face challenges to participating in restaurant-based strategies, especially when recipe analysis or sales tracking is required. To address these barriers, lessons may be drawn from work with small food store owners, who often struggle to stock healthy items because of limitations related to infrastructure, staff expertise, and access to appropriate suppliers (27).

In general, we found more restaurants in neighborhoods that had a greater percentage of non-Hispanic white residents and residents with more than a high school education; these populations tend to be less affected by diet-related diseases. Where restaurants choose to locate is driven by various market forces, including the proportion of targeted households, traffic generators, and sales generators (28). Given that spending on food purchased away from home increases with income (29), we were not surprised that the density of restaurants was greater in Los Angeles County's higher-income neighborhoods, where income level might allow for a greater amount of discretionary spending than in lower-income neighborhoods. Few studies have examined restaurant density (as a whole) in relationship to neighborhood demographic characteristics. One

national study showed that higher-income neighborhoods and neighborhoods with predominantly black or African American and racially mixed residents had lower levels of access to both fast-food and full-service restaurants (16). Previous work in Los Angeles County found that lower-middle and upper-middle socioeconomic census tracts had the highest total number of restaurants, compared with very low- and very high-income tracts (14). Previous literature does not clarify whether a greater density of restaurants is protective or detrimental. Greater restaurant density could mean more consumer choice to seek out healthy options. Alternatively, given that consuming food at restaurants is associated with greater intake of energy, fat, and sodium, relative to consuming foods prepared at home (30), greater restaurant density could lead to less healthy behaviors and outcomes.

Our study provides limited support to the idea that large chains are more heavily concentrated in certain neighborhoods (neighborhoods with a lower percentage of non-Hispanic white residents and a lower percentage of residents with more than a high school education). We did not observe significant relationships between chain density and neighborhood sociodemographic characteristics when chains were defined according to the number of locations nationally; rather, we observed relationships between chain density and race/ethnicity and education level only when we examined restaurants with 20 or more locations in Los Angeles County. Previous research on the relationship between density of chain restaurants and neighborhood sociodemographic characteristics is limited, and these studies tended to conflate chain status with fast-food service style and cuisine. In our study, although many chain restaurants were classified as quick service (fast food), not all quick service restaurants were part of a chain.

Although the completeness and large sample sizes of the data sources are strengths, our study has several limitations. First, our analysis does not provide any information on the reasons or causal mechanisms underlying the observed associations between community characteristics and restaurant density and characteristics. Restaurant location is driven by various market forces; restaurants often choose to cluster in commercial areas. Second, we used sociodemographic indicators as rough markers of disproportionate disease burden. Mapping the restaurant landscape in relationship to the prevalence of diet-related disease is an important area for future work. Third, although people are likely to visit restaurants outside their neighborhood, our study treated neighborhood boundaries as rigid boundaries and did not account for travel distance to restaurants or the profile of restaurants near a person's school or work. For our analysis to have useful and relevant neighborhood boundaries, we used neighborhood definitions that were meaningful to residents, while respecting census boundaries to accurately integrate demographic data. However, calculating restaurant density according to census units did not account for edge effects, particularly along major arterials. Finally, the market research data did not provide any information on the relative healthfulness, size, or sales of restaurants. We used market research data rather than administrative data because they provided a more accurate picture of the number and types of restaurants, particularly information on chain status (the major study question). Future studies would benefit from data sets that include information on the relative healthfulness of food options.

Our study provides insight on the potential importance of including small chain and independent restaurants in efforts to advance healthier food access in communities. The extent to which restaurant-based strategies are an effective means to target diet-related disparities remains unclear. Additional work is needed to better understand the extent to which restaurant-based initiatives can effectively reach people most in need. It is important to consider where restaurants — especially those affected or targeted by program or policy work (such as chains) — are located in relation to priority populations. Our study answers a question that is infrequently examined yet of critical importance to advance local public health practice. Applied researchers and evaluators in other jurisdictions can build on the methods used in our study to gain a deeper understanding of their local landscape.

Acknowledgments

Author Information

Corresponding Author: Lauren N. Gase, PhD, MPH, Senior Researcher, Spark Policy Institute, 2717 Welton St, Denver, CO 80205. Telephone: 303-455-1740. Email: laurengase@gmail.com.

Author Affiliations: [1]Division of Chronic Disease and Injury Prevention, Los Angeles County Department of Public Health, Los Angeles, California. [2]Spark Policy Institute, Denver, Colorado.

References

1. Story M, Kaphingst KM, Robinson-O'Brien R, Glanz K. Creating healthy food and eating environments: policy and environmental approaches. Annu Rev Public Health 2008; 29(1):253–72.

2. US Food and Drug Administration. Calories count and Keystone report; 2006. https://www.fda.gov/Food/FoodScienceResearch/ConsumerBehaviorResearch/default.htm. Accessed December 4, 2018.

3. US Food and Drug Administration. Food labeling: nutrition labeling of standard menu items in restaurants and similar retail food establishments; 2014. https://www.fda.gov/downloads/Food/IngredientsPackagingLabeling/LabelingNutrition/UCM423985.pdf. Accessed January 16, 2018.

4. Otten JJ, Saelens BE, Kapphahn KI, Hekler EB, Buman MP, Goldstein BA, et al. Impact of San Francisco's toy ordinance on restaurants and children's food purchases, 2011–2012. Prev Chronic Dis 2014;11:E122.

5. Davis City Council. Ordinance establishing the default beverages offered with children's meals; 2015. https://cityofdavis.org/home/showdocument?id=3656. Accessed December 4, 2018.

6. Glanz K, Hoelscher D. Increasing fruit and vegetable intake by changing environments, policy and pricing: restaurant-based research, strategies, and recommendations. Prev Med 2004; 39(Suppl 2):S88–93.

7. Gittelsohn J, Lee-Kwan SH, Batorsky B. Community-based interventions in prepared-food sources: a systematic review. Prev Chronic Dis 2013;10:E180.

8. National Restaurant Association. Kids LiveWell Program. 2015. http://www.restaurant.org/Industry-Impact/Food-Healthy-Living/Kids-LiveWell-Program. Accessed January 16, 2018.

9. UNC Center for Health Promotion and Disease Prevention. Smart Meal Seal program; 2012. http://www.centertrt.org/content/docs/Intervention_Documents/Intervention_Templates/Smart_Meal_Seal_Template.pdf. Accessed December 4, 2018].

10. Pomeranz JL, Brownell KD. Legal and public health considerations affecting the success, reach, and impact of menu-labeling laws. Am J Public Health 2008;98(9):1578–83.

11. Goldman TR. Health policy brief: the FDA's menu-labeling rule. Bethesda (MD): Health Affairs; June 25, 2015. https://www.healthaffairs.org/do/10.1377/hpb20150713.56602/full/. Accessed December 4, 2018.

12. Roberto CA, Schwartz MB, Brownell KD. Rationale and evidence for menu-labeling legislation. Am J Prev Med 2009; 37(6):546–51.

13. Britt JW, Frandsen K, Leng K, Evans D, Pulos E. Feasibility of voluntary menu labeling among locally owned restaurants. Health Promot Pract 2011;12(1):18–24.

14. Inagami S, Cohen DA, Brown AF, Asch SM. Body mass index, neighborhood fast food and restaurant concentration, and car ownership. J Urban Health 2009;86(5):683–95.

15. Kwate NOA, Yau C-Y, Loh J-M, Williams D. Inequality in obesigenic environments: fast food density in New York City. Health Place 2009;15(1):364–73.

16. Powell LM, Chaloupka FJ, Bao Y. The availability of fast-food and full-service restaurants in the United States: associations with neighborhood characteristics. Am J Prev Med 2007; 33(4,Suppl):S240–5.

17. Bleich SN, Wolfson JA, Jarlenski MP. Calorie changes in chain restaurant menu items: implications for obesity and evaluations of menu labeling. Am J Prev Med 2015; 48(1):70–5.

18. Krieger JW, Chan NL, Saelens BE, Ta ML, Solet D, Fleming DW. Menu labeling regulations and calories purchased at chain restaurants. Am J Prev Med 2013;44(6):595–604.

19. US Census Bureau. 2010–2014 American Community Survey 5-year estimate, B01003: total population; 2015. https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_14_5YR_B01003&prodType=table. Accessed December 13, 2016.

20. US Census Bureau. 2010–2014 American Community Survey 5-year estimate, B03002: Hispanic or Latino origin by race; 2015. https://factfinder.census.gov/faces/table services/jsf/pages/productview.xhtml?pid=ACS_14_5YR_B03002&prodType=table. Accessed December 13, 2016.

21. US Census Bureau. 2010–2014 American Community Survey 5-year estimate, B15003: educational attainment for the population 25 years and over; 2015. https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_14_5YR_B15003&prodType=table. Accessed December 13, 2016.

22. US Census Bureau. 2010-2014 American Community Survey 5-year estimate, B17001: poverty status in the past 12 months by sex by age; 2015. https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_14_5YR_B17001&prodType=table. Accessed December 13, 2016.

23. US Census Bureau. 2010–2014 American Community Survey 5-year estimate, B19013: Median household income in the past 12 months (in 2014 inflation-adjusted dollars); 2015. https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_14_5YR_B19013&prodType=table. Accessed December 13, 2016.

24. Enterprise GIS Group. Los Angeles County GIS data portal. 2015. http://egis3.lacounty.gov/dataportal/2015/05/11/la-county-cams-address-locator. Accessed January 6, 2017.

25. Los Angeles Times. About Mapping L.A; 2015. http://maps.latimes.com/about/. Accessed January 6, 2017.

26. Urban LE, Weber JL, Heyman MB, Schichtl RL, Verstraete S, Lowery NS, et al. Energy contents of frequently ordered restaurant meals and comparison with human energy requirements and US Department of Agriculture database information: a multisite randomized study. J Acad Nutr Diet 2016;116(4):590–8.e6.

27. Bentzel D, Weiss S, Bucknum M, Shore K. Healthy food and small stores: strategies to close the distribution gap. 2015. http://thefoodtrust.org/uploads/media_items/healthy-food-and-small-stores.original.pdf. Accessed December 13, 2016.

28. Park KP, Khan MA. An exploratory study to identify the site selection factors for U.S. franchise restaurants. J Foodserv Bus Res 2008;8(1):97–114.

29. Carlson A, Frazão E. Food costs, diet quality and energy balance in the United States. Physiol Behav 2014;134:20–31.

30. An R. Fast-food and full-service restaurant consumption and daily energy and nutrient intakes in US adults. Eur J Clin Nutr 2016;70(1):97–103.

Tables

Table 1. Characteristics of Restaurants (N = 24,292) in Los Angeles County, California, 2016[a]

Characteristic	Number (%)
Chain status	
Based on number of locations nationally	
Independent (single location)	15,114 (62.2)
Small chain (2–19 locations)	2,748 (11.3)
Large chain (≥20 locations)	6,430 (26.5)
Based on number of locations in Los Angeles County	
Independent (single location)	15,449 (63.6)
Small chain (2–19 locations)	3,698 (15.2)
Large chain (≥20 locations)	5,145 (21.2)
Industry market segment	
Quick service	9,571 (39.4)
Fast casual	2,548 (10.5)
Midscale dining	4,626 (19.0)
Casual dining	6,483 (26.7)
Fine dining	609 (2.5)
Missing data	455 (1.9)
Type of cuisine	
American/Southern	4,476 (18.4)
Asian	4,438 (18.3)
Latino	3,765 (15.5)
Coffee/bakery/dessert	3,208 (13.2)
Burger	1,943 (8.0)
Pizza	1,692 (7.0)
Sandwich/deli	1,566 (6.5)
European	1,520 (6.3)
Other	1,493 (6.1)
Missing data	191 (0.8)

[a] Information on restaurant characteristics was provided by a market research firm that tracks restaurant industry trends nationally.

Table 2. Characteristics of Neighborhoods (N = 247) in Los Angeles County, California, 2016[a]

Sociodemographic characteristics[b]	
Total population	40,210 (43,989)
Race/ethnicity	
Percentage of Non-Hispanic white residents	31.5 (26.4)
Percentage of Hispanic or Latino residents	43.4 (27.6)
Percentage of black or African American residents	8.7 (14.3)
Percentage of Asian residents	13.5 (14.1)
Percentage of residents aged ≥25 with >high school education	58.0 (21.6)
Percentage of residents below the poverty level in the last 12 months	16.7 (10.0)
Median household income in the last 12 months, $	67,895.7 (31,671.8)
Restaurant characteristics	
Number of restaurants	94.4 (117.5)[c]
Number of restaurants per 1,000 residents	2.3 (1.8)
Proportion of restaurants that are large chains based on the number of locations nationally	26.5 (15.0)[d]
Proportion of restaurants that are large chains based on the number of locations in Los Angeles County	21.7 (12.4)[e]

[a] Neighborhoods and their boundaries were defined according to the *Los Angeles Times*' Mapping L.A. project. All analyses excluded 8 neighborhoods with <1,000 residents and 3 neighborhoods with >15 restaurants per 1,000 residents.
[b] Based on census tract level data drawn from the 2010–2014 American Community Survey 5-Year Estimates, aggregated to the neighborhood level (19–23).
[c] Thirty-two of 247 neighborhoods (13.0%) had <10 restaurants and were excluded in sensitivity analyses examining the relationship between restaurant chain density and neighborhood sociodemographic characteristics.
[d] When we examined neighborhoods with ≥10 restaurants (n = 215), average was 27.2% and standard deviation was 12.5%.
[e] When we examined neighborhoods with ≥10 restaurants (n = 215), average was 22.9% and standard deviation was 10.5%.

Table 3. Relationship Between Neighborhood Sociodemographic Characteristics and Density of Large Chain Restaurants, Los Angeles County, California, 2016[a]

Quartile	Mean (Standard Deviation)			
	Percentage of Non-Hispanic White Residents	Percentage of Residents with >High School Education	Percentage of Residents Below the Poverty Level	Median Household Income, $
Percentage of restaurants that are large chain[b,c] (based on the number of locations nationally)				
Quartile 1 (0%–18%)	36.0 (28.4)	60.9 (21.8)	17.0 (10.7)	68,360.7 (32,674.4)
Quartile 2 (19%–26%)	35.2 (28.8)	58.4 (24.0)	18.4 (9.9)	65,370.3 (32,879.2)
Quartile 3 (27%–35%)	26.8 (23.4)	56.1 (20.4)	15.8 (8.0)	65,832.0 (24,492.2)
Quartile 4 (>35%)	28.4 (24.2)	56.9 (20.4)	15.6 (11.0)	71,683.3 (35,499.6)
Percentage of restaurants that are large chain[b,c] (based on the number of locations in Los Angeles County)				
Quartile 1 (0%–14%)	40.5 (28.8)	64.4 (21.7)	15.4 (10.2)	74,177.7 (36,175.4)
Quartile 2 (15%–22%)	34.2 (28.0)	59.2 (24.1)	18.1 (11.1)	66,282.7 (30,176.5)
Quartile 3 (23%–29%)	27.3 (22.3)[d]	56.0 (18.4)[e]	16.0 (7.0)	64,073.5 (21,896.4)
Quartile 4 (>29%)	24.8 (23.9)[f]	53.0 (20.8)[g]	17.3 (11.0)	67,197.8 (35,946.7)

[a] Neighborhoods (N = 247) and their boundaries were defined according to the *Los Angeles Times*' Mapping L.A. project. All analyses excluded 8 neighborhoods with <1,000 residents and 3 neighborhoods with >15 restaurants per 1,000 residents.
[b] Results did not substantively or significantly change when analyses were conducted on neighborhoods with ≥10 restaurants (n = 215).
[c] Large chain restaurants were defined as restaurants with ≥20 locations.
[d] $P = .005$ for difference between quartile 3 and quartile 1, based on simple linear regression.
[e] $P = .03$ for difference between quartile 3 and quartile 1, based on simple linear regression.
[f] $P = .001$ for difference between quartile 3 and quartile 1, based on simple linear regression.
[g] $P = .003$ for difference between quartile 3 and quartile 1, based on simple linear regression.

RESEARCH BRIEF

Occupational Groups and Environmental Justice: A Case Study in the Bronx, New York

Andrew R. Maroko, PhD[1]; Brian T. Pavilonis, PhD[1]

PEER REVIEWED

Abstract

We used spatial analyses to examine exposure of people in vulnerable occupational groups to neighborhood-level environmental pollutants in the Bronx borough of New York City. Five-year estimates of environmental ambient exposures (derived from land use regression models for PM2.5 [particulate matter with an aerodynamic diameter $\leq$2.5 μm] and black carbon) and demographic and occupational variables were harmonized at the census tract level. Correlations revealed that areas with high environmental exposures also had high proportions of people in service industries and manufacturing and high proportions of socioeconomically vulnerable populations. This combination of vulnerabilities may be cumulative, suggesting residents could have high occupational and residential exposures in addition to sociodemographic-related inequity.

Objective

Socioeconomically disadvantaged populations and racial/ethnic minority populations often live in areas with more environmental hazards than other population groups, an environmental justice issue that may lead to poor health outcomes and worsen differences in health (1–4). However, few studies have examined how occupational groups may be differentially distributed with respect to ambient environmental (neighborhood) exposures. Our ecological study sought to determine whether people in vulnerable occupational groups (ie, those with potentially high exposures to pollutants in the workplace) could be overexposed to environmental pol-

lutants on the basis of their place of residence in the Bronx borough of New York City, thus constituting a potential environmental justice issue.

Methods

Employment information for civilians aged 16 or older at the census tract level were obtained from the US Census Bureau's 2011–2015 *American Community Survey* via the National Historical Geographic Information System (NHGIS.org, IPUMS.org) (5). We collapsed job classifications into 4 categories on the basis of a previous study (6): white collar, service industry, construction (including protective services and agriculture because of a small sample size and similarity in exposure), and manufacturing.

Environmental exposures were derived from 300-meter resolution land use regression model outputs provided by the New York City Department of Health and Mental Hygiene (7). Land use regression uses a statistical model to estimate ambient pollutant concentration as a function of land use (eg, vehicle traffic, building emissions, population density). The environment surrounding monitoring locations in New York City was used to parameterize the regression equation for each year (number of monitors is from 60 to 100, depending on year), which is then applied to locations around the city where no measurements have been taken to create a continuous surface of annual average concentration estimates (8). We resampled land use regression outputs in the Bronx at the census tract level and calculated 5-year (2011–2015) average concentrations of PM2.5 (particulate matter with an aerodynamic diameter $\leq$2.5 μm) and black carbon (a type of particulate pollution often used as a marker for diesel exhaust [9]). Census tract-level demographic and socioeconomic variables (proportion of non-Hispanic white, non-Hispanic black, and Hispanic populations and the population's poverty status) were derived from the *American Community Survey* 5-year data for 2011–2015 (5) (Figure). Associations among pollutant concentration, occupational groups, demographics, and economics were tested by using nonparametric Spearman correlations for census tracts with more than 200 residents (n = 330).
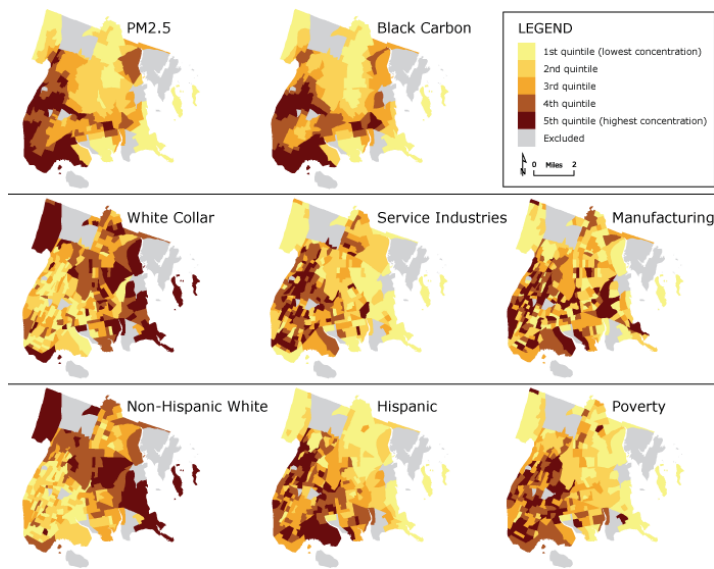
Figure. Spatial distribution of pollutants, occupational groups, and demographics, in quintiles by census tract, Bronx, New York, 2011–2015. The 2 pollutants are PM2.5 (particulate matter with an aerodynamic diameter ≤2.5 μm) and black carbon (a type of particulate pollution often used as a marker for diesel exhaust [9]). The occupational groups are white collar, service industries, and manufacturing (construction not shown but available from authors on request); the demographic groups are non-Hispanic white and Hispanic populations (non-Hispanic black not shown but available from authors on request) and poverty (people living below federal poverty guidelines). Tracts with low populations (200 or fewer residents) were excluded. Sources: *American Community Survey 2011–2015* ACS 5-Year Estimates via the National Historic Geographic Information System (5), New York City Community Air Survey 2011–2015 (10).

Results

Spearman correlations identified significant positive associations between estimated concentrations of black carbon and PM2.5 and proportions of Hispanic residents and people with incomes below federal poverty guidelines ($P < .01$). The proportion of non-Hispanic black residents was not significantly associated with estimated pollutant exposures. Significant positive associations ($P < .01$) were observed between census tracts with high proportions of white-collar workers and non-Hispanic white residents. Conversely, negative associations were found between the proportion of white-collar workers and the proportion of non-Hispanic black and Hispanic residents and people living in poverty ($P < .01$). Census tracts with high proportions of service industry or manufacturing workers were negatively associated with non-Hispanic white populations but positively associated with Hispanic populations and with people living in poverty ($P < .01$). Proportions of

non-Hispanic black residents were positively associated with service industry occupations ($P < .01$) but did not reach significance with respect to manufacturing.

The proportion of workers who identified as being employed in the service industry or manufacturing had significant positive associations with ambient environmental exposure to black carbon and PM2.5 ($P < .01$). Conversely, tracts with high proportions of white-collar workers had significant negative associations with these pollutants ($P < .01$) (Table).

Discussion

The Bronx borough of New York City has often been studied with respect to environmental justice issues because of its high proportion of vulnerable populations, historic settlement patterns, environmental burdens, and poor health outcomes among its residents (11,12). However, occupational exposures to airborne particulate matter are often overlooked in the development of chronic diseases such as cardiovascular disease and, depending on the industry, can be orders of magnitude larger than environmental exposures (13). Occupational sectors such as service industry, construction, and manufacturing have higher mortality rates than white-collar sectors (6).

The results from our study show several spatial relationships among occupational groups, neighborhood environmental exposures, and demographics. The most vulnerable occupational groups (ie, those with the highest likelihood of poor health outcomes or high exposure to pollutants in workplace environments) are positively associated with neighborhoods with higher concentrations of PM2.5 and black carbon. These same neighborhoods also tend to have higher proportions of vulnerable populations on the basis of race/ethnicity and income levels. These sociodemographic characteristics are associated with increased risk of environmental exposures and possibly amplify the effects of these exposures (4). These populations are consistently associated with increased incidence and severity of disease — potentially as a function of psychosocial stressors such as discrimination and social exclusion (4). This combination of vulnerabilities is likely to be cumulative, putting residents of certain neighborhoods in double jeopardy on the basis of traditionally measured environmental injustices as well as environmental injustice as a function of occupational group. Such residents could have high exposures both at work and at home and may suffer from additional socially driven inequity based on racial/ethnic or economic characteristics.

Occupational attributes appear to be important variables, not only with respect to environmental justice work but also more generally in terms of environmental health studies. Although such studies often incorporate either neighborhood exposures or occupa-

tional exposures, they rarely include both simultaneously. The confluence of high-risk occupational groups and environmental neighborhood exposures (physical and social) may further contribute to, or exacerbate, health disparities in regions like the Bronx.

Acknowledgments

Author Information

Corresponding Author: Andrew Maroko, 55 W 125th St, New York, NY 10027. Telephone: 646-364-9508. Email: andrew.maroko@sph.cuny.edu.

Author Affiliations: [1]The CUNY Graduate School of Public Health and Health Policy, New York, New York.

References

1. Chakraborty J, Schweitzer L, Forkenbrock D. Using GIS to assess the environmental justice consequences of transportation system changes. Trans GIS 1999;3(3):239–58.
2. Maantay J, Maroko A. 'At-risk' places: inequities in the distribution of environmental stressors and prescription rates of mental health medications in Glasgow, Scotland. Environ Res Lett 2015;10(11):115003.
3. Maroko AR. Using air dispersion modeling and proximity analysis to assess chronic exposure to fine particulate matter and environmental justice in New York City. Appl Geogr 2012;34:533–47.
4. Morello-Frosch R, Zuk M, Jerrett M, Shamasunder B, Kyle AD. Understanding the cumulative impacts of inequalities in environmental health: implications for policy. Health Aff (Millwood) 2011;30(5):879–87.
5. Manson S, Schroeder J, Van Riper D, Ruggles S. IPUMS National Historical Geographic Information System: version 12.0 [Database]. Minneapolis (MN): University of Minnesota; 2017. Accessed March 29, 2018.
6. Arheart KL, Fleming LE, Lee DJ, Leblanc WG, Caban-Martinez AJ, Ocasio MA, et al. Occupational vs. industry sector classification of the US workforce: which approach is more strongly associated with worker health outcomes? Am J Ind Med 2011;54(10):748–57.
7. Kheirbek I, Johnson S, Ito K, Anan K, Huskey C, Matte T, et al.The New York City Community Air Survey: neighborhood air quality 2008–2014. http://www1.nyc.gov/assets/doh/downloads/pdf/environmental/comm-air-survey-08-14.pdf. Accessed December 16, 2017.
8. Clougherty JE, Kheirbek I, Eisl HM, Ross Z, Pezeshki G, Gorczynski JE, et al. Intra-urban spatial variability in wintertime street-level concentrations of multiple combustion-related air pollutants: the New York City Community Air Survey. J Expo Sci Environ Epidemiol 2013;23(3):232–40.
9. Matte TD, Ross Z, Kheirbek I, Eisl H, Johnson S, Gorczynski JE, et al. Monitoring intraurban spatial patterns of multiple combustion air pollutants in New York City: design and implementation. J Expo Sci Environ Epidemiol 2013; 23(3):223–31.
10. The New York City Department of Health and Mental Hygiene, Queens College, and Zev Ross Spatial Analysis. New York City community air survey. https://www1.nyc.gov/site/doh/data/data-publications/air-quality-nyc-community-air-survey.page. Accessed December 12, 2017.
11. Maantay J. Asthma and air pollution in the Bronx: methodological and data considerations in using GIS for environmental justice and health research. Health Place 2007; 13(1):32–56.
12. Maroko AR, Weiss Riley R, Reed M, Malcolm M. Direct observation of neighborhood stressors and environmental justice in the South Bronx, New York City. Popul Environ 2014;35(4):477–96.
13. Fang SC, Cassidy A, Christiani DC. A systematic review of occupational exposure to particulate matter and cardiovascular disease. Int J Environ Res Public Health 2010;7(4):1773–806.

Table

Table. Spearman Correlations for Occupational Groups, Demographics, and Environmental Exposures, Bronx, New York, 2011–2015

Variables[a]	White Collar	Service Industry	Manufac-turing	Construc-tion[b]	Non-Hispanic White	Non-Hispanic Black	Hispanic	Poverty[c]	PM2.5[d]	Black Carbon
Occupation										
White collar	1	—e	—e	—e	—e	—e	—e	—e	—e	—e
Service industry	−.868[f]	1	—e	—e	—e	—e	—e	—e	—e	—e
Manufacturing	−.554[f]	.310[f]	1	—e	—e	—e	—e	—e	—e	—e
Construction	−.219[f]	−.089	−.041	1	—e	—e	—e	—e	—e	—e
Demographics										
Non-Hispanic white	.594[f]	−.605[f]	−.307[f]	−.017	1	—e	—e	—e	—e	—e
Non-Hispanic black	−.217[f]	.242[f]	.032	.074	−.507[f]	1	—e	—e	—e	—e
Hispanic	−.598[f]	.564[f]	.446[f]	−.011	−.516[f]	−.222[f]	1	—e	—e	—e
Poverty	−.676[f]	.682[f]	.394[f]	−.025	−.619[f]	.167[f]	.648[f]	1	—e	—e
Environmental pollutant										
PM2.5	−.503[f]	.529[f]	.349[f]	−.122[g]	−.495[f]	.010	.636[f]	.579[f]	1	—e
Black carbon	−.505[f]	.511[f]	.356[f]	−.071	−.500[f]	−.024	.638[f]	.584[f]	.949[f]	1
Mean (standard deviation)	48.51 (12.53)	29.39 (9.75)	11.13 (4.46)	10.97 (4.48)	12.84 (20.74)	28.76 (20.53)	52.67 (20.88)	29.78 (15.03)	9.71 (0.60)	1.19 (0.16)

[a] Occupational and demographic values are percentage of the census tract population. Environmental pollutant values are concentrations at the census tract level.
[b] Includes protective services and agriculture.
[c] Percentage of the census tract population with incomes below federal poverty guidelines.
[d] Particulate matter with an aerodynamic diameter ≤2.5 μm.
[e] The correlation matrix is symmetrical. Entries to the right are suppressed for clarity.
[f] Correlation is significant at $P < .01$ level (2-tailed).
[g] Correlation is significant at $P < .05$ level (2-tailed).

ORIGINAL RESEARCH

Estimating County-Level Mortality Rates Using Highly Censored Data From CDC WONDER

Harrison Quick, PhD

PEER REVIEWED

Summary

What is already known on this topic?

Ignoring the impact of suppression due to small counts leads to biased inference.

What is added by this report?

This work describes and compares multiple approaches for analyzing highly suppressed data from CDC WONDER. R and WinBUGS code are provided to conduct the analyses.

What are the implications for public health practice?

The use of spatial Bayesian models can yield improved inference from the analysis of highly suppressed data such as those available on CDC WONDER.

Abstract

Introduction

CDC WONDER is a system developed to promote information-driven decision making and provide access to detailed public health information to the general public. Although CDC WONDER contains a wealth of data, any counts fewer than 10 are suppressed for confidentiality reasons, resulting in left-censored data. The objective of this analysis was to describe methods for the analysis of highly censored data.

Methods

A substitution approach was compared with 1) a simple, nonspatial Bayesian model that smooths rates toward their statewide averages and 2) a more complex Bayesian model that accounts for spatial and between-age sources of dependence. Age group–specific county-level data on heart disease mortality were used for the comparisons.

Results

Although the substitution and nonspatial approach provided age-standardized rate estimates that were more highly correlated with the true rate estimates, the estimates from the spatial Bayesian model provided a superior compromise between goodness-of-fit and model complexity, as measured by the deviance information criterion. In addition, the spatial Bayesian model provided rate estimates with greater precision than the nonspatial approach; in contrast, the substitution approach did not provide estimates of uncertainty.

Conclusion

Because of the ability to account for multiple sources of dependence and the flexibility to include covariate information, the use of spatial Bayesian models should be considered when analyzing highly censored data from CDC WONDER.

Introduction

CDC WONDER (Wide-ranging ONline Data for Epidemiologic Research) is a system developed by the Centers for Disease Control and Prevention (CDC) to promote information-driven decision making by public health practitioners and researchers and provide access to detailed public health information to the general public (1). Although CDC WONDER contains a wealth of data, it has limitations. Per CDC policy (2), any counts fewer than 10 should be suppressed for confidentiality reasons, resulting in left-censored data. Because of high rates of suppression, many chronic disease researchers opt to focus their inference in a few highly populated regions (3) or state- or national-level trends (4), despite known geographic disparities in many chronic disease outcomes (5,6). This suppression may also discourage research on disparities between subsets of the population (eg, race or sex disparities) to avoid reducing already small counts below suppression thresholds. In short, suppression of small counts exacerbates many issues commonly encountered in the field of small area estimation, where the term "small area" refers to a geographic scale (eg,

county, census tract) at which the observed data alone do not provide reliable inference. Thus, when CDC WONDER data are used to conduct surveillance, the ability to estimate rates for rural areas and minority populations — where the chronic disease burden is high (7) — is significantly hindered by data suppression.

To address CDC WONDER's data suppression issue, Tiwari et al (8) proposed an algorithm for estimating age-standardized rates in which suppressed age-specific counts are replaced with estimates based on the county's age-specific population size and the state-wide average rate for that age-group. For example, suppose y_{ik} denotes the number of deaths from age-bracket k in county i of a population of size n_{ik} and our inferential interest lies in λ_{ik}, the corresponding mortality rate. Tiwari et al (8) proposed replacing the suppressed $y_{ik} < 10$ with $y_{ik}^* = \bar{\lambda}_{s_i0k} \times n_{ik}$, where s_i denotes the state that county i belongs to and $\bar{\lambda}_{s_i0k}$ denotes the state-wide average rate for age-bracket k in state s_i such that

$$\bar{\lambda}_{s_i0k} = \sum_{j:s_j = s_i} y_{jk} / \sum_{j:s_j = s_i} n_{jk} \text{ (Equation 1)}$$

Because state-level totals are often 10 or greater, we will assume from this point forward that $\bar{\lambda}_{s_i0k}$ is known and publicly available; when this is not the case, rates could be smoothed toward an alternative value (eg, national estimates).

Although this approach may yield reasonable estimates, it has drawbacks. First and foremost, estimating the uncertainty in age-standardized rate estimates is not an exact science when the data are known (9,10), much less when the data are highly suppressed. Furthermore, the algorithm is not designed to account for heterogeneity in demographic information such as the racial/ethnic make-up and socioeconomic status of the counties' populations. As a result, inference based on these substituted data may be both biased (ie, smoothing toward the wrong values) and too precise (ignoring the uncertainty due to data suppression).

When the goal is to assess geographic disparities in age-standardized rates between regions, overcoming the privacy protections to obtain trustworthy estimates of the age-specific rates and their levels of uncertainty is only half the battle. For instance, Fay (11) followed the work of Fay and Feuer (9) to construct interval estimates for ratios based on F distributions. Tiwari et al (10) modified this work to yield more efficient interval estimation for rates and ratios of rates from nonnested regions, work that was later extended by Tiwari et al (12) for when one subregion is nested within a larger region (eg, a county nested within a state); Zhu et al (13) extended these approaches to more accurately account for spatial autocorrelation. When the age-standardized rates must be estimated from suppressed data, further modifications must be made or

these approaches will fail to adequately account for all sources of uncertainty, yielding interval estimates that may be too narrow (14,15).

Rather than develop the statistical theory to accurately account for substitution-based approaches to overcome CDC WONDER's privacy restrictions in variance calculations, we consider the use of Bayesian statistical models, which rely on data augmentation to make inference on the suppressed counts. As described by Fridley and Dixon (14), data augmentation approaches estimate the suppressed counts via multiple imputation (16) while simultaneously making inference on the parameters of interest — for example, λ_{ik} and the effects of potential risk factors. As noted by Zhu et al (13), Bayesian methods for modeling spatial data (17) can yield improved rate estimates when data are limited while simultaneously providing a mechanism for estimating uncertainty in rate estimates — uncertainty that can be seamlessly propagated into estimates such as age-standardized rates and rate ratios. That said, a key drawback of Bayesian methods is their tendency to rely on computationally burdensome Markov chain Monte Carlo (MCMC) methods.

The objective of this analysis was to illustrate 2 Bayesian approaches for estimating county-level mortality rates, by using heart disease mortality data from 1980 obtained from CDC WON- DER (18), and to compare these results with those generated by the approach of Tiwari et al (8). In particular, we used a simple, nonspatial Bayesian model, which produces estimates similar to those from Tiwari et al (8), along with a more complex Bayesian model that accounts for spatial and between-age sources of dependence.

Methods

The study population for this analysis included all residents of the contiguous United States aged 35 or older during 1980. These data have multiple advantages. Because these data were collected before CDC's suppression guidelines (2) went into effect, the public-use data are complete and free of suppression. Furthermore, because county definitions changed in several ways during the 1980s, the choice of data from 1980 allowed use of readily available shapefiles from the US Census Bureau for the $I = 3,109$ counties (or county equivalents) in the contiguous United States. To replicate the analysis of Tiwari et al (8), the data were separated into $K = 6$ groups: those aged 35 to 44, 45 to 54, 55 to 64, 65 to 74, 75 to 84, and 85 or older. Annual counts of heart disease–related deaths per county per age-group were obtained via CDC WONDER (18) and were defined as those for which the underlying cause of death was "diseases of the heart" according to the *In-*

ternational Classification of Diseases, Ninth Revision (codes 390–398, 402, 404–429). Of the more than 18,000 counts in this data set, nearly half were fewer than 10.

Statistical model

Recall that y_{ik} and n_{ik} denote the number of deaths and the population size in age group k in county i. To model these data, we considered 2 approaches: a simple Poisson-gamma model and a multivariate spatial Bayesian model. Although the former illustrates how a Bayesian model with weakly informative priors can produce estimates similar to those obtained directly from the raw data — but with accurate uncertainty measures — the latter illustrates how Bayesian models can incorporate complex dependence structures to produce more reliable estimates. A formal definition of what constitutes a "reliable" rate and the implications of this definition are provided in the Web Appendix (https://sites.google.com/site/harryq/wonder). Because of the complexity of Bayesian models, the Web Appendix also provides technical details on the methods described in this article and includes R (19) and WinBUGS (20) code.

Poisson-gamma model

Following the advice of Brillinger (21), we assumed

$$y_{ik}|\lambda_{ik} \sim Pois(n_{ik}\lambda_{ik}) \text{ (Equation 2)}$$

for $i = 1, \ldots, I$ and $k = 1, \ldots, K$. Because we wished to fit Equation 2 using a Bayesian framework, we had to specify a prior distribution for each λ_{ik}. A convenient choice was to let

$$\lambda_{ik} \sim Gam(y_{s_i0k}, n_{s_i0k}) \text{ (Equation 3)}$$

As described in the Web Appendix, y_{s_i0k} can be interpreted as the prior number of events and n_{s_i0k} as the prior population size, thereby providing a mechanism for comparing the informativeness of the prior to the amount of information contained in the data. For example, a prior with $n_{s_i0k} = 1,000$ would contain the same amount of information as the data when $n_{ik} = 1,000$, and the posterior mean would be equal to the average of $\lambda_{s_i0k} = y_{s_i0k}/n_{s_i0k}$ (the estimate from the prior) and $\hat{\lambda}_{ik} = y_{ik}/n_{ik}$ (the estimate from the data). Here, we can take an empirical Bayesian approach by letting $\lambda_{s_i0k} = \lambda_{s_i0k}$ from Equation 1 and defining the informativeness of the prior to be such that $\sum_k y_{s_i0k} = 6$ for all states under the restriction that the $n_{s_i0k} = y_{s_i0k}/\lambda_{s_i0k}$ parameters respect the age distribution in the United States. To better accommodate low rates among the younger age groups, which produce a preponderance of zero counts, we modified the prior in Equation 3 based on the suggestion of Kerman (22) by letting

$$\lambda_{ik} \sim Gam\left(y_{s_i0k} + 1/3, n_{s_i0k}\right) \text{ (Equation 4)}$$

This prior specification can be considered relatively noninformative because 96.4% of US counties had more than $\sum_k(y_{s_i0k} + 1/3) = 8$ heart disease–related deaths in 1980. A more complete discussion of this model is provided in the Web Appendix.

Multivariate conditional autoregressive model

Although the prior specification in Equation 4 is a convenient choice, it does not take full advantage of the possibilities of Bayesian modeling. In particular, Equation 4 does not account for spatial relationships or the relationships between different age groups. To allow for such structures to be included in the model, we considered Poisson regression models, where

$$\log \lambda_{ik} = \boldsymbol{x}_{ik}^T \boldsymbol{\beta}_k + \theta_{ik} \text{ (Equation 5)}$$

Here, $\boldsymbol{x}_{ik}$ denotes a vector of county-specific covariates with corresponding age-specific regression coefficients, $\boldsymbol{\beta}_k$; for example, including state-level effects could help account for important health policy differences across state lines. For this analysis, we simply assumed $\boldsymbol{x}_{ik}^T \boldsymbol{\beta}_k = \beta_{0k}$; that is, a model with age-specific intercept parameters. To account for spatial and between-age sources of dependence, we first followed the approach of Besag et al (17) and defined $\theta_{ik} = z_{ik} + \varphi_{ik}$, where z_{ik} accounts for spatial structure within each age-group and φ_{ik} denotes an exchangeable (ie, nonspatial) random effect. More specifically, the conditional autoregressive (CAR) model of Besag et al (17) imposes spatial structure by shrinking each z_{ik} toward the values in neighboring counties (ie, counties that share a border), where the strength of this shrinkage is controlled by the number of neighboring counties.

Although the CAR model is a powerful tool for analyzing spatial data, it does not account for possible correlation between the multiple age groups. To account for this, we instead considered a multivariate extension of the CAR model: the multivariate CAR (MCAR) model of Gelfand and Vounatsou (23). As with the CAR model, the MCAR shrinks estimates toward their neighboring values; unlike the CAR model, however, the MCAR explicitly models the between-group correlation in the data and leverages these correlations to produce more precise age-specific rate estimates. MCAR models were used recently to model spatially referenced survival times in cancer data (24), temporal trends in county-level asthma hospitalization rates (25), temporal trends in heart disease mortality by race and sex (26), and temporal trends in age-specific stroke mortality (27), among many other applications. Full details, including a discussion of the prior distributions used, are provided in the Web Appendix.

Bayesian inference

Fitting the models in Equation 4 or Equation 5 while accounting for the suppression of counts fewer than 10 requires the use of MCMC algorithms. Because of the reliance on MCMC, inference from these Bayesian models is based on samples generated from the posterior distribution — for example, $\lambda_{ik}^{(l)}$ for $l = 1, \ldots, L$, where L denotes the number of samples. These samples can then be used to compute quantities such as the age-standardized mortality rate:

$$\lambda_{i\cdot}^{(l)} = \sum_k \pi_k \lambda_{ik}^{(l)}$$

where π_k denotes a prespecified standard age distribution (eg, based on the 2010 US standard population). To summarize the posterior distribution, it is common to use the posterior median and the 95% credible interval (constructed from the 2.5 and 97.5 percentiles of the posterior samples and analogous to classical 95% confidence intervals).

Comparison of approaches

To compare the various estimation approaches, we first considered simple correlations between the estimates and the rates obtained from the complete data (as considered by Tiwari et al [8]) and correlations between the age-standardized rates and the age-specific rates. The goal of these comparisons was not to demonstrate whether one approach is superior to another but rather to demonstrate the degree to which the approaches are similar to one another. In addition, we also compared the 2 Bayesian approaches by using the deviance information criterion (DIC) (28), which uses the posterior samples to produce a measure that is a compromise between model fit (denoted by $\overline{D}$) and model complexity, p_D. In particular, p_D is often interpreted as the effective number of parameters in the model. Additional details on DIC, including a discussion of its use with censored data, are provided in the Web Appendix.

Creation of maps

Maps were created by using the R statistical software (The R Foundation). Code is available in step 6 of the walkthrough in the Web Appendix (https://sites.google.com/site/harryq/wonder).

Results

The maps of the age-standardized rates generated from the raw data (Figure 1A) and the maps generated by the Poisson-gamma model (Figure 1C) have strong similarities, while artifacts of substituting state-wide averages for suppressed counts based on the

approach of Tiwari et al (8) lead to elevated estimates in many rural counties in the upper Midwest (Figure 1B). In contrast, the map of the estimates from the MCAR model (Figure 1D) preserves the overall trends in the data while producing significantly smoother rate estimates.
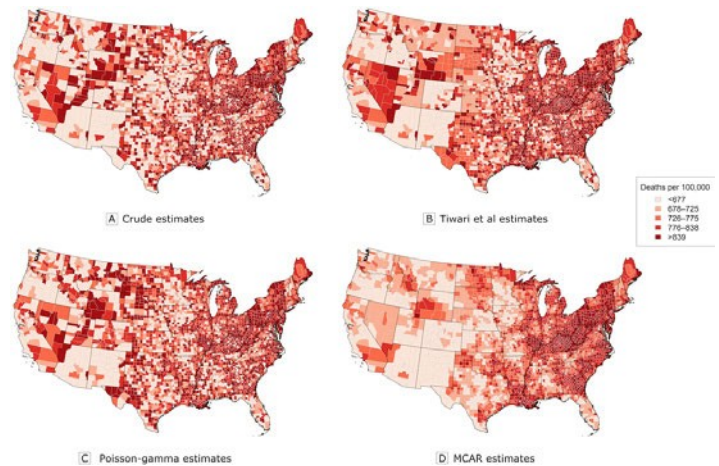

Figure 1. Estimates of age-standardized heart disease mortality rates from 1980. A, Crude age-standardized rates based solely on the data. B, Estimates obtained by using the approach of Tiwari et al (8). C, Estimated posterior medians from the Poisson-gamma model. D, Estimated posterior medians from the multivariate conditional autoregressive model (MCAR). Data source: Centers for Disease Control and Prevention (18).

The correlation results (Table 1) largely support this assessment. The Poisson-gamma approach produced age-standardized rate estimates that were the most highly correlated with the true rates, although the estimates obtained by using the substitution approach of Tiwari et al (8) had nearly an identical correlation. These 2 approaches differed in age-specific rate estimates. In particular, although the Poisson-gamma approach appeared to struggle for adults aged 35 to 44 — producing estimates that were less correlated with the truth — it outperformed the substitution approach for all groups aged 55 or older. Figure 2, which displays the age-specific rate estimates for adults aged 35 to 44 and adults 85 or older, explains how this occurred. Here, although the approach of Tiwari et al (8) gave every suppressed county in each state the same rate (by design), the Poisson-gamma model tended to overestimate rate estimates for those aged 35 to 44. According to Kerman (22), this overestimation of rates when counts are very small was to be expected. Furthermore, unlike the approach of Tiwari et al (8), the Poisson-gamma model produced full posterior distributions for each age-specific rate estimate, thereby allowing quantification of the uncertainty in these estimates. (Figure B.3 in the Web Appendix illustrates how only 4.5% of estimates for those aged 35 to 44 and 42.8% of all age-specific rate estimates from the Poisson-

gamma model were deemed reliable.) When estimating rates for those 85 or older, the Poisson-gamma model permitted heterogeneity within states (Figure 2E); the inability to permit such heterogeneity is a key weakness of the approach of Tiwari et al (8). Further evaluation of the low age-specific correlations is provided in the Web Appendix (Figures B.1 and B.2).
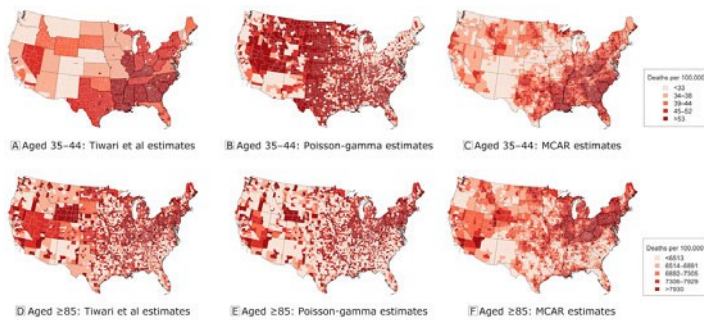

Figure 2. Comparison of 3 approaches for estimating age-standardized heart disease mortality rates for 2 age groups (adults aged 35 to 44 and adults aged ≥85) from 1980. A, Estimates for adults aged 35 to 44 obtained by using the approach of Tiwari et al (8). B, Estimated posterior medians for adults aged 35 to 44 from the Poisson-gamma model. C, Estimated posterior medians for adults aged 35 to 44 from the multivariate conditional autoregressive model (MCAR). D, Estimates for adults aged ≥85 obtained by using the approach of Tiwari et al (8). E, Estimated posterior medians for adults aged ≥85 from the Poisson-gamma model. F, Estimated posterior medians for adults aged ≥85 from the multivariate conditional autoregressive model (MCAR). Data source: Centers for Disease Control and Prevention (18).

Looking at the correlation results (Table 1) and the maps in Figure 1, one may wonder why we bother fitting the complex MCAR model. The DIC results (Table 2) explain why. Here, the MCAR model offered a model fit that is similar to the fit of the Poisson-gamma model (as measured by $\overline{D}$) while doing so with far fewer "effective model parameters" (p_D). To understand how this can be, recall that each λ_{ik} in Equation 4 had its own independent prior distribution; that is, the Poisson-gamma model did not shrink the λ_{ik} toward each other, producing estimates of the (p_D) for older age groups that approach the full $I = 3,109$ number of parameters. In contrast, the MCAR model explicitly imposed dependence between its model parameters, resulting in estimates of the (p_D) that were nearly 80% less than those from the Poisson-gamma model (eg, 10,785 vs 2,307). In addition, the estimates produced by the MCAR model were more precise (Web Appendix), and the smooth geographic patterns in Figure 1D, Figure 2C, and Figure 2F may provide clearer insight into the underlying trends in heart disease mortality.

Discussion

This analysis highlighted some of the benefits of using Bayesian methods to account for left-censored data like those encountered in CDC WONDER. Although the Poisson-gamma model is a relatively simple approach, models (such as the MCAR model) that explicitly account for multivariate spatial dependence structures can lead to better inference by leveraging other sources of information to produce more reliable estimates.

The strengths of the MCAR model described in this analysis extend beyond modeling censored data to the broader field of small area estimation. As alluded to in the discussion of Equation 5, many benefits are associated with using the MCAR model in conjunction with covariate information when modeling chronic disease outcomes. Combining covariate information with spatial structure can produce more reliable estimates of the rates themselves, which is beneficial for disease surveillance, while simultaneously conducting inference on the potential risk factors that are included as covariates. When the covariates in the analysis are themselves spatially structured, it can be unclear if the covariate is effecting change in the outcome or vice versa, or if an unmeasured spatial confounder is influencing both the covariate *and* the outcome. In these settings, including a spatial random effect can lead to a phenomenon referred to as "spatial confounding" (29) and increase the standard errors associated with these covariates. Although the notion of spatial confounding has historically been considered a drawback of spatial models (29), others have argued (30) that inference from such models can help protect against type 1 error (ie, incorrectly rejecting the null hypothesis).

Finally, although we analyzed age-specific heart disease mortality as an illustration, the MCAR model is also well suited for analyzing rarer event data via its ability to jointly model multiple outcomes. This analysis leveraged information from older age groups with higher death counts to produce more reliable estimates for those aged 35 to 44. Similarly, one could jointly model a chronic disease outcome for multiple race/ethnicities, exploiting the shared factors that may lead to increased rates for non-Hispanic white persons and racial/ethnic minorities alike. Alternatively, one could use MCAR models to simultaneously analyze multiple chronic disease outcomes with similar etiologies to improve the reliability of all estimates.

Although the suppression of data creates an obstacle to conducting chronic disease surveillance, Bayesian statistical methods such as those described in this analysis can overcome these challenges while also producing more reliable estimates with valid uncertainty measures. By illustrating the benefits of and providing code

for their implementation, we hope to ease the burden of using Bayesian models and broaden their application to censored data sets available from sources like CDC WONDER, thereby improving the inference made from public-use data.

Acknowledgments

Author Information

Corresponding Author: Harrison Quick, PhD, Department of Epidemiology and Biostatistics, Drexel University, Philadelphia, PA 19104. Email: hsq23@drexel.edu.

References

1. Centers for Disease Control and Prevention. CDC WONDER, 2017. http://wonder.cdc.gov. Accessed March 17, 2017.
2. Centers for Disease Control and Prevention. CDC/ATSDR policy on releasing and sharing data. Manual; guide CDC-02. 2003. http://www.cdc.gov/maso/Policy/ReleasingData.pdf. Accessed June 30, 2015.
3. Rust G, Zhang S, Malhotra K, Reese L, McRoy L, Baltrus P, et al. Paths to health equity: local area variation in progress toward eliminating breast cancer mortality disparities, 1990–2009. Cancer 2015;121(16):2765–74.
4. Wilmot KA, O'Flaherty M, Capewell S, Ford ES, Vaccarino V. Coronary heart disease mortality declines in the United States from 1979 through 2011: evidence for stagnation in young adults, especially women. Circulation 2015; 132(11):997–1002.
5. Casper M, Kramer MR, Quick H, Schieb LJ, Vaughan AS, Greer S. Changes in the geographic patterns of heart disease mortality in the United States 1973 to 2010. Circulation 2016; 133(12):1171–80.
6. Vaughan AS, Quick H, Pathak EB, Kramer MR, Casper M. Disparities in temporal and geographic patterns of declining heart disease mortality by race and sex in the United States, 1973–2010. J Am Heart Assoc 2015;4(12):e002567.
7. National Center for Health Statistics. Health, United States, 2017: with special feature on mortality. https://www.cdc.gov/nchs/data/hus/hus17.pdf. Accessed February 27, 2019.
8. Tiwari C, Beyer K, Rushton G. The impact of data suppression on local mortality rates: the case of CDC WONDER. Am J Public Health 2014;104(8):1386–8.
9. Fay MP, Feuer EJ. Confidence intervals for directly standardized rates: a method based on the gamma distribution. Stat Med 1997;16(7):791–801.
10. Tiwari RC, Clegg LX, Zou Z. Efficient interval estimation for age-adjusted cancer rates. Stat Methods Med Res 2006; 15(6):547–69.
11. Fay MP. Approximate confidence intervals for rate ratios from directly standardized rates with sparse data. Commun Stat Theory Methods 1999;28(9):2141–60.
12. Tiwari RC, Li Y, Zou Z. Interval estimation for ratios of correlated age-adjusted rates. J Data Sci 2010;8:471–82.
13. Zhu L, Pickle LW, Pearson JB Jr. Confidence intervals for rate ratios between geographic units. Int J Health Geogr 2016; 15(1):44.
14. Fridley BL, Dixon P. Data augmentation for a Bayesian spatial model involving censored observations. Environmetrics 2007; 18(2):107–23.
15. Quick H, Groth C, Banerjee S, Carlin BP, Stenzel MR, Stewart PA, et al. Exploration of the use of Bayesian modeling of gradients for censored spatiotemporal data from the *Deepwater Horizon* oil spill. Spat Stat 2014;9:166–79.
16. Rubin DB. Multiple imputation for nonresponse in surveys. New York (NY): John Wiley and Sons; 1987.
17. Besag J, York J, Mollié A. Bayesian image restoration, with two applications in spatial statistics. Ann Inst Stat Math 1991; 43(1):1–20.
18. Centers for Disease Control and Prevention, National Center for Health Statistics. Compressed mortality file 1979–1998. CDC WONDER on-line database, compiled from compressed mortality file CMF 1968–1988, series 20, no. 2A, 2000 and CMF 1989–1998, series 20, no. 2E, 2003. https://wonder.cdc.gov/controller/saved/D16/D10F745. Accessed March 3, 2017.
19. R Core Team. R: A Language and Environment for Statistical Computing. Vienna (AT): R Foundation for Statistical Computing; 2015.http://www.R-project.org/. Accessed February 27, 2019.
20. Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS — a Bayesian modelling framework: concepts, structure, and extensibility. Stat Comput 2000;10(4):325–37.
21. Brillinger DR. The natural variability of vital rates and associated statistics. Biometrics 1986;42(4):693–734.
22. Kerman J. Neutral noninformative and informative conjugate beta and gamma prior distributions. Electron J Stat 2011; 5(0):1450–70.
23. Gelfand AE, Vounatsou P. Proper multivariate conditional autoregressive models for spatial data analysis. Biostatistics 2003;4(1):11–25.

24. Carlin BP, Banerjee S. Hierarchical multivariate CAR models for spatio-temporally correlated survival data (with discussion). In: JM Bernardo, M Bayarri, JO Berger, AP Dawid, D Heckerman, AFM Smith, and M West, editors. Bayesian statistics 7: Proceedings of the Seventh Valencia International Meeting. 45–63. Oxford (UK): Oxford Science Publications; 2003. P. 45–63.

25. Quick H, Banerjee S, Carlin BP. Modeling temporal gradients in regionally aggregated California asthma hospitalization data. Ann Appl Stat 2013;7(1):154–76.

26. Quick H, Waller LA, Casper M. A multivariate space-time model for analysing county-level heart disease death rates by race and sex. J R Stat Soc 2018;67(1):291–304.

27. Quick H, Waller LA, Casper M. Multivariate spatiotemporal modeling of age-specific stroke mortality. Ann Appl Stat 2017; 11(4):2165–77.

28. Spiegelhalter DJ, Best N, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit (with discussion). J R Stat Soc B 2002;64(4):583–639.

29. Reich BJ, Hodges JS, Zadnik V. Effects of residual smoothing on the posterior of the fixed effects in disease-mapping models. Biometrics 2006;62(4):1197–206.

30. Hanks EM, Schliep EM, Hooten MB, Hoeting JA. Restricted spatial regression in practice: geostatistical models, confounding, and robustness under model misspecification. Environmetrics 2015;26(4):243–54.

Tables

Table 1. Comparison of the Correlation Results of 3 Estimation Approaches, Analysis of County-Level Mortality Rates Using Highly Censored Data From CDC WON-DER[a]

Approach	Age Group						Age-Standardized
	35–44	45–54	55–64	65–74	75–84	≥85	
Tiwari et al (8)	0.15	0.73	0.16	0.07	−0.01	0.08	0.73
Poisson-gamma	0.09	0.74	0.23	0.25	0.24	0.27	0.74
Multivariate conditional autoregressive model	0.15	0.65	0.18	0.15	0.05	0.14	0.65

[a] Age-standardized correlation results were based on all 3,109 US counties, whereas age-specific correlation results were based only on the suppressed counties (counties with counts <10). Data source: Centers for Disease Control and Prevention (18).

Table 2. Comparison of the Deviance Information Criterion[a] Results of 3 Estimation Approaches, Analysis of County-Level Mortality Rates Using Highly Censored Data From CDC WONDER[b]

Approach	Age Group						Overall
	35–44	45–54	55–64	65–74	75–84	≥85	
Poisson-gamma							
DIC	2,204	6,108	12,393	17,866	19,005	16,956	74,533
$\overline{D}$	1,663	5,006	10,509	15,447	16,506	14,616	63,748
p_D	542	1,102	1,884	2,419	2,499	2,339	10,785
Multivariate conditional autoregressive model							
DIC	1,558	5,242	11,245	16,201	17,417	15,904	67,568
$\overline{D}$	1,478	5,030	10,842	15,743	16,887	15,281	65,260
p_D	80	213	403	458	530	624	2,307

[a] Spiegelhalter et al (28).
[b] Where $\overline{D}$ is a measure of model fit (lower is better), p_D is a measure of model complexity (lower indicating fewer effective model parameters), and $DIC = \overline{D} + p_D$. Data source: Centers for Disease Control and Prevention (18).

TOOLS FOR PUBLIC HEALTH PRACTICE

The Rate Stabilizing Tool: Generating Stable Local-Level Measures of Chronic Disease

Harrison Quick, PhD[1]; Joshua Tootoo, MS[2]; Ruiyang Li, MS[2]; Adam S. Vaughan, PhD, MPH, MS[3];

Linda Schieb, MSPH[3]; Michele Casper, PhD[3]; Marie Lynn Miranda, PhD[2]

PEER REVIEWED

Summary

What is already known on this topic?

Existing methods of generating small area estimates often require advanced statistical knowledge, programming and coding skills, and extensive computing power.

What is added by this report?

We created an ArcGIS Tool — the Rate Stabilizing Tool (RST) — that produces age-adjusted rate estimates from record-level data and indicates which rates should be considered statistically reliable. This tool is particularly important for generating estimates when the population size or the number of events is small. With the RST, estimates can be generated for a wide range of geographic units, including subcounty levels.

What are the implications for public health practice?

With its ease of use, the RST addresses the need to produce stable local estimates of chronic disease measures to improve chronic disease surveillance, prevention, and treatment.

Abstract

Accurate and precise estimates of local-level epidemiologic measures are critical to informing policy and program decisions, but they often require advanced statistical knowledge, programming/coding skills, and extensive computing power. In response, we developed the Rate Stabilizing Tool (RST), an ArcGIS-based tool that enables users to input their own record-level data to generate more reliable age-standardized measures of chronic disease (eg, prevalence rates, mortality rates) or other population health outcomes at the county or census tract levels. The RST uses 2 forms of empirical Bayesian modeling (nonspatial and spatial) to estimate age-standardized rates and 95% credible intervals for user-specified geographic units. The RST also provides indicators of the reliability of point estimates. In addition to reviewing the RST's statistical techniques, we present results from a simulation study that illustrates the key benefit of smoothing. We demonstrate the dramatic reduction in root mean-squared error (rMSE), indicating a better compromise between accuracy and stability for both smoothing approaches relative to the unsmoothed estimates. Finally, we provide an example of the RST's use. This example uses heart disease mortality data for North Carolina census tracts to map the RST output, including reliability of estimates, and demonstrates a subsequent statistical test.

Introduction

Public health professionals are increasingly using spatial analysis and geographic information systems (GIS) to document and address geographic disparities in the burden of chronic disease (1–8). Maps of local-level disparities in chronic disease morbidity, mortality, risk factors, and treatments are critical to informing policy and program decisions and enhancing partnerships to address the disparities (9–12). One important component in the use of GIS for chronic disease prevention and health promotion is the availability of data at the local level (eg, county, census tract) that yield stable estimates that are both accurate and precise. Here, our focus is the ability to produce stable *event* rates (eg, death rates), which depend primarily on the number of events that occur in a place of interest for a designated period. These event counts in turn depend on the prevalence or incidence of the event and the population size. In general, the smaller the population size, the smaller the event counts and the greater the instability in population measures of chronic disease. In particular, small counts are often encountered when analyzing small geographic areas (eg, census tracts) or examining population subgroups (eg, race/ethnicity, sex) or sparsely populated regions (eg, rural areas). In this article, we use the term "small area" to refer to areas for which the

data alone do not provide stable estimates for a given population measure, regardless of the physical size of the geographic area itself.

Recent advances in computing and in the field of small-area estimation — specifically Bayesian methods (13–17) — have provided avenues for generating more reliable local-level population measures of chronic disease when the number of events are small. In particular, these approaches often involve *smoothing* observed rates toward a common mean (eg, the national average) or toward neighboring values. However, these methods typically require knowledge of advanced statistics, programming/coding skills, and extensive computing power — resources that may be challenging to obtain for many public health professionals in need of stable small-area estimates.

In response to the need for local-level measures of chronic disease and recognizing the challenges that often exist in generating reliable estimates, we developed the Rate Stabilizing Tool (RST). The RST is an ArcGIS-based tool that enables users to input their own record-level data to generate more reliable age-standardized measures of chronic disease (eg, prevalence, rates) or other population health outcomes at the local level. Bayesian modeling techniques are built into the tool, enabling users to better evaluate measures of statistical uncertainty for each population subgroup and locale.

In this article, we describe the statistical techniques that are built into the Rate Stabilizing Tool, review the results from a simulation study, provide an overview for how to use the RST, and discuss its strengths and limitations. Files needed to install the RST and detailed instructions are available at https://www.cdc.gov/dhdsp/maps/gisx/rst.html. Statistical and technical details of the Rate Stabilizing Tool are available in a Web Appendix (https://sites.google.com/site/harryq/rst).

Statistical Techniques of the Rate Stabilizing Tool

Bayesian modeling

The Rate Stabilizing Tool employs Bayesian modeling techniques to generate local-level estimates of the prevalence of chronic disease (or other outcomes). These estimates are more stable than those generated by conventional methods. Bayesian modeling techniques are used because 1) they are well-equipped to maximize the information gained from available data in situations where data are sparse, thereby yielding estimates with greater precision than crude estimates, and 2) they generate accompanying measures of uncertainty, the benefits of which will be discussed shortly. Bayesian methods generate estimates by combining information from the observed data (via the *likelihood* [ie, the distribution of

the observed data given various model parameters]) and so-called *prior information* (often expressed in the form of model structure [eg, spatial correlation]). The result of this combination is referred to as the *posterior distribution*. From the posterior distribution, we can then generate summaries such as the mean and 95% credible interval (the Bayesian equivalent of classical confidence intervals) for each of the region-specific rate estimates and make statistical comparisons with other values. An extended introduction to Bayesian methods is available in the Web Appendix; a more thorough introduction to Bayesian methods can be found in the text by Carlin and Louis (18).

Two forms of Bayesian modeling are incorporated into the RST — a nonspatial approach and a spatial approach. In the nonspatial approach, local-level rates are smoothed toward the observed rate from the overarching spatial domain (eg, the rate for a selected state). In contrast, the spatial approach smooths each local-level rate toward the crude rate of the combined neighboring geographic units (and is similar to the approach of Clayton and Kaldor [17]). Complete details on these approaches, including justifications for the selected likelihood and prior distributions and derivations of the posterior distributions, are available in the Web Appendix.

Age-standardization of local-level rates

Age-standardization of local-level chronic disease rates is important because differences in age-distributions across regions can contribute to stark differences in measures of the burden of chronic disease, even if the underlying rates in each age-group are comparable. Generally speaking, age-standardized rates for a given region are obtained by computing the weighted average of the region's age-specific rates, where the weights used are based on the age distribution of a standard population (eg, the 2010 US standard [19,20]). Directly using these age-specific rates poses challenges, however, because crude estimates of these rates are often based on small counts. Not only can these small counts lead to age-specific rate estimates that are unstable, but the instability in the age-specific rates can seep into the age-standardized estimates. As such, a key feature of the RST is that we first obtain smoothed estimates of the age-specific rates by using one of the aforementioned Bayesian methods, and then these smoothed age-specific rates are used to compute the age-standardized rates for each region. This process allows the uncertainty in the smoothed age-specific rates to propagate through to the age-standardized rates; in contrast, estimates of the age-standardized rates based solely on the data may require complex equations to approximate these variance estimates (21,22).

Simulation Study

We conducted a simulation study to compare smoothed age-standardized rates (both spatial and nonspatial smoothing) with unsmoothed age-standardized rates to demonstrate the RST's effectiveness. The simulation study was based on heart disease death data from US counties for 1979–1988 and multiple age groups (35–44, 45–54, 55–64, 65–74, 75–84, and ≥85) obtained from CDC WONDER (23). From these data, we calculated an estimate of the age-group–specific mortality rate for each county; these are henceforth considered the "true rates" and were used to generate 100 data sets of simulated death count. We then analyzed the simulated death data by using the spatial and nonspatial smoothing methods of the RST and compared the estimates from the RST to the unsmoothed age-standardized mortality rates. We compared estimates from all 3 approaches by using root mean square error (rMSE), a measure that combines the bias of an estimate and its variance, and we estimated coverage probabilities (ie, the proportion of the 95% credible interval that contains the true rates) for both smoothing approaches. Complete details of the simulation study are available in the Web Appendix.

Figure 1 compares the rMSE of the age-standardized rate estimates from the spatial and nonspatial smoothing approaches with the rMSE of the unsmoothed rates, where a lower rMSE indicates better compromise between accuracy (ie, bias) and stability (ie, variance). Here, we see the key benefit of smoothing, namely a dramatic reduction in the rMSE for both smoothing approaches when compared with the unsmoothed estimates. A comparison of the age-standardized and age-group specific estimates from the 2 smoothing approaches shows only minor differences. A more thorough comparison of these 2 approaches, including maps of the rMSEs of the age-group specific and the age-standardized rates, can be found in the Web Appendix. In addition to improvements in rMSE, both smoothing approaches achieved coverage probabilities approximately equal to 0.95 as desired (ie, the 95% credible intervals contain the true values approximately 95% of the time).
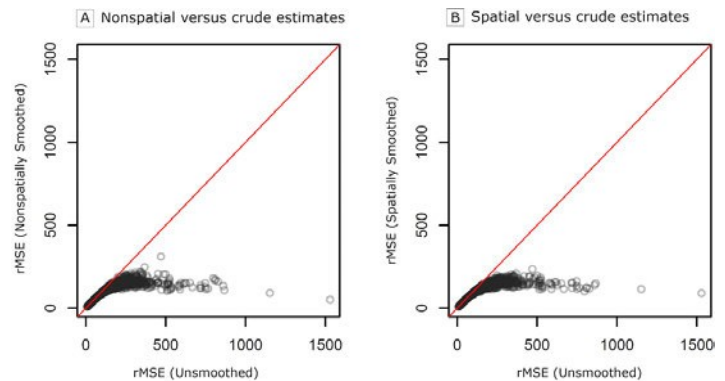

Figure 1. Comparison of the root mean square error (rMSE) of the age-standardized rates from the 2 smoothing approaches (A, nonspatial vs crude estimates and B, spatial vs crude estimates) of the Rate Stabilizing Tool to the unsmoothed rates estimated directly from the raw data in the simulation study.

An Overview of How to Use the Rate Stabilizing Tool (RST)

The RST operates as a set of tools within an ArcToolbox toolset; no installation or administrative privileges are required to run this tool. After inputting individual-level data into ArcGIS, users specify their desired age structure, and then the RST produces 3 sets of age-standardized rates: unsmoothed; nonspatially smoothed; and spatially smoothed. The RST also generates 95% credible intervals and alerts on the reliability of each smoothed rate estimate. An overview of the use of the tool is as follows:

1. Input individual-level data. The user loads a table where each record represents a single event (eg, death) and contains the individual's age and a geographic identifier (eg, census tract, county).

2. Choose age structure. The user then selects age groups that will be used for age-standardization. For age standardization, the RST connects to the US Census Data web API (https://census.gov/data/developers/data-sets/acs-5year.html) and downloads the age-specific population sizes for each census geography of interest, along with the age distribution for the US standard population.

3. Import US Census areal unit boundary definitions (24) (eg, a shapefile) for map creation and spatial smoothing. In addition to facilitating the creation of maps, the tool will use the boundary definitions to create a neighborhood dictionary for the geographic units in the spatial domain. The neighborhood dictionary is required for RST's spatial smoothing approach. This dictionary de-

scribes which geographic units are adjacent to one another, thus defining the neighbor pairs. Once constructed, the neighborhood dictionary is saved and can be re-used for future analyses with the same shapefile.

4. Examine and evaluate the output. The RST generates an output text file, with one record for each geographic unit. Each record contains the following information:

- Age-standardized rate, unsmoothed
- Age-standardized rate, smoothed (nonspatial) and corresponding 95% credible intervals
- Age-standardized rate, smoothed (spatial) and corresponding 95% credible intervals

In addition to providing rate estimates and 95% CIs, the RST also provides an alert when the estimate for a given geographic unit is deemed unreliable (ie, when the width of the 95% credible interval is larger than the estimate). The RST generates 3 types of alerts:

- Unreliable nonspatial Bayesian estimate, when the nonspatial Bayesian estimate is not reliable for a given geographic unit;
- Unreliable spatial Bayesian estimate, when the spatial Bayesian estimate is not reliable for a given geographic unit; and
- Unreliable estimate, when neither of the Bayesian estimates are reliable for a given geographic unit.

5. Mapping the results. After evaluating the output from the RST and deciding which values are appropriate to display on a map, users can create maps by joining the output from the RST to their US Census areal unit boundary definition shapefile for the area of interest. Users can easily make maps comparing the display of the 3 types of rates generated by the RST.

6. Using the tool: an example. To illustrate the use of the RST, we analyzed data on heart disease deaths in Charlotte, North Carolina, for 2006–2011. We used the RST to age standardize the mortality rates to the 4 age groups (0–34, 35–44, 45–64, and ≥65 y) and generated heart disease mortality rates at the census tract level. We obtained shapefiles corresponding to these boundaries from 2010 US Census Topologically Integrated Geographic Encoding and Referencing reference files. These boundaries were used in the RST's spatial Bayesian smoothing approach.

The map on the left side of Figure 2A displays unsmoothed age-standardized heart disease mortality rates in Charlotte and the surrounding area. Although this map highlights census tracts with high and low observed mortality rates, it obscures the degree of statistical uncertainty in these rates. For example, if a priority is to

target public health interventions to areas with elevated rates, how would one differentiate between census tracts with truly high rates and census tracts with high rates that are unreliable because of small population sizes? To address this challenge, we mapped the smoothed rates (nonspatially smoothed and spatially smoothed) and found census tracts with unreliable mortality rates (2 maps on right side of Figure 2A). These 2 maps indicate that the rates for many of the census tracts are unreliable (33.4% with nonspatial smoothing and 34.1% with spatial smoothing) and should be considered with caution.
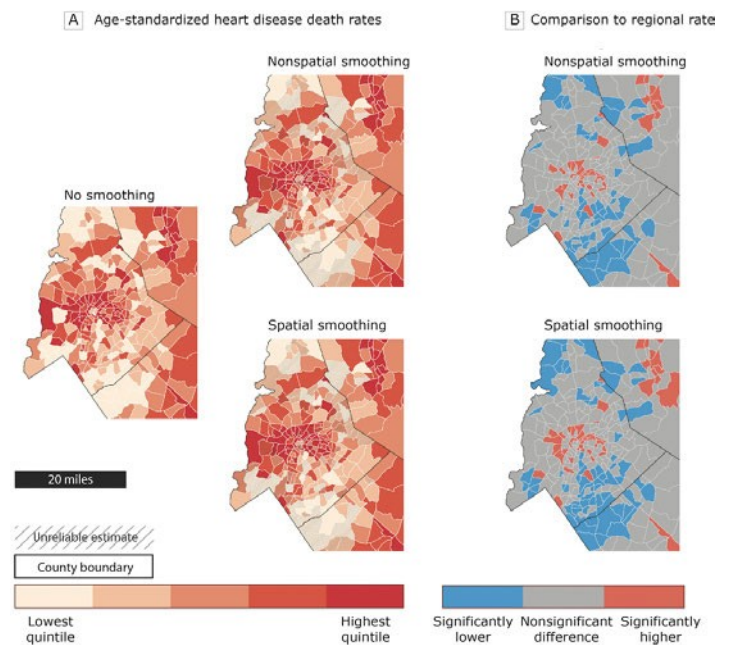

Figure 2. Illustration of the functionality of the Rate Stabilizing Tool using heart disease mortality data from the region surrounding Charlotte, North Carolina. A, Age-standardized heart disease death rates by census tract using 3 methods, with hatch marks indicating unreliable rates based on the 2 Bayesian smoothing approaches. B, census tracts with death rates that are significantly higher or lower than the state rate using the 2 smoothing methods.

An additional way to use the information on statistical uncertainty generated by the RST is to compare the rate for each census tract to a regional standard. The maps in Figure 2B display census tracts that have age-standardized heart disease death rates that are significantly higher or significantly lower than the regional average rate based on the 95% credible intervals generated by the RST for spatially and nonspatially smoothed rates. Census tracts where the 95% credible intervals do not include the mean rate for the region were classified as having rates that were significantly higher or significantly lower than the rate for the region. For several census tracts — such as those in the southern part of the Charlotte,

North Carolina region (right side of Figure 2A) — the rates were determined to be unreliable because of the wide 95% credible intervals, but we can conclude that those rates are significantly below the regional average because the entire range of the 95% credible intervals is below the regional rate.

Strengths and Limitations

An important strength of the RST is that it combines 2 tasks — rate smoothing and age-standardization — into a single tool. By doing so, the RST avoids the potential pitfall of estimating age-standardized rates from extreme age-specific rates (eg, rates based on zero deaths). The RST overcomes this pitfall by first smoothing the age-specific rates, producing age-specific rates that are more reliable than those calculated directly from the data. By calculating the age-standardized rates on the basis of these smoothed rates, we can improve the stability of our estimates (Figure 1).

In addition to the ease-of-use attributable to combining these 2 tasks, the RST offers inferential improvements. We demonstrated through our simulation study that both approaches for computing smoothed age-standardized rates dramatically improve the quality of the estimates compared with the estimates generated solely from the observed data based on the rMSE. In addition, the RST provides 95% credible intervals for smoothed age-standardized rates: this is a notable strength given the complexity of producing uncertainty estimates when calculating age-standardized rates according to standard methods (21,22). Furthermore, the 95% credible intervals produced by the RST yield coverage probabilities (ie, the probability that the 95% credible interval contains the true value) near the desired 0.95 for both the age-specific and the age-standardized rate estimates. This indicates that convenience of the RST does not compromise statistical validity.

This version of the RST has several limitations. First, although many public-use data sets consist of aggregate, tabular data that comprise the number of events and the population sizes stratified by geographic unit and age group, the RST is designed only to analyze *record-level* data. To mitigate this limitation, we developed instructions to generate *synthetic* individual-level data from a table of aggregate data. Future iterations of the tool will allow users to import record-level or tabular data directly. In addition to added flexibility, future updates to the tool will facilitate the analysis of public-use data sets from sources such as CDC WONDER, which are subject to various privacy protections that result in data *suppression* (eg, CDC WONDER suppresses counts of ≤9 to protect data privacy [25]); ignoring (or inappropriately accounting for) these protective measures may result in biased rate estimates (26). After this functionality is added, the RST will be able to seamlessly account for such privacy protections to pro-

duce rate estimates for small areas that are both reliable and valid; Quick et al (27) explained how this can be done. The RST is also not currently equipped to analyze survey data, where accommodating sample sizes and survey weights adds layers of complexity that must be carefully considered.

A final limitation of the RST is that it relies on empirical Bayesian methods rather than fully Bayesian methods. The approaches used by the RST smooth toward estimates determined by the data and the degree of smoothing is predetermined. In contrast, a fully Bayesian approach would include prior distributions on the values each region is smoothed toward *and* the degree of smoothing, thereby learning from the data what each region should be smoothed toward and how strong the smoothing should be. The conditional autoregressive model of Besag et al (13) is a popular approach for this type of analysis. Unfortunately, fully Bayesian methods have one key drawback: computational burden. In particular, fully Bayesian models are typically fitted by using complex Markov chain Monto Carlo algorithms that must be run until convergence has been achieved. That is, the algorithm needs to iteratively learn about each of the model parameters until their estimates stabilize, a process which often requires thousands of iterations and can take minutes or hours to complete depending on the size of the data set. Because convergence is often diagnosed visually, designing the RST to diagnose convergence in an automated and efficient fashion is much more challenging. Despite these computational challenges, however, the inferential benefits of fully Bayesian models necessitate their consideration in future iterations of the RST.

Conclusion

The Rate Stabilizing Tool is an add-on tool for ArcGIS that produces accurate and precise estimates of event rates for geographic areas with small population sizes or small counts. The RST imports record-level event data and uses an empirical Bayesian model to estimate age-standardized rates and 95% credible intervals for user-specified geographic units. In addition, users are alerted if a point estimate is deemed unreliable for a given geographic unit. With its ease of use, the RST addresses the need to produce stable local estimates of chronic disease measures to improve chronic disease surveillance, prevention, and treatment.

Acknowledgments

Author Information

Corresponding Author: Harrison Quick, PhD, Department of Epidemiology and Biostatistics, Drexel University, 3215 Market St, Philadelphia, PA 19104. Email: hsq23@drexel.edu.

Author Affiliations: [1]Department of Epidemiology and Biostatistics, Drexel University, Philadelphia, Pennsylvania. [2]Children's Environmental Health Initiative, Rice University, Houston, Texas. [3]Division for Heart Disease and Stroke Prevention, Centers for Disease Control and Prevention, Atlanta, Georgia.

References

1. Centers for Disease Control and Prevention. Chronic disease GIS exchange. 2017. https://www.cdc.gov/dhdsp/maps/gisx/index.html. Accessed March 1, 2018.

2. Dubowitz T, Williams M, Steiner ED, Weden MM, Miyashiro L, Jacobson D, et al. Using geographic information systems to match local health needs with public health services and programs. Am J Public Health 2011;101(9):1664–5.

3. Tootoo J, Castrucci BC, Maxson P, Casper M, Miranda M. Using geographic health information systems for population health. In: Michener J, Koo D, Castrucci B, Sprague J, editors. The practical playbook: public health and primary care together. New York (NY): Oxford University Press; 2015. p. 287.

4. Vaughan AS, Ritchey MD, Hannan J, Kramer MR, Casper M. Widespread recent increases in county-level heart disease mortality across age groups. Ann Epidemiol 2017; 27(12):796–800.

5. Korycinski RW, Tennant BL, Cawley MA, Bloodgood B, Oh AY, Berrigan D. Geospatial approaches to cancer control and population sciences at the United States cancer centers. Cancer Causes Control 2018;29(3):371–7.

6. Quick H, Waller L, Casper M. Multivariate spatiotemporal modeling of age-specific stroke mortality. Ann Appl Stat 2017; 11(4):2165–77.

7. Quick H, Waller L, Casper M. A multivariate space-time model for analyzing county-level heart disease death rates by race and sex. J R Stat Soc Ser C Appl Stat 2018; 67(1):291–304.

8. Vaughan AS, Quick H, Pathak EB, Kramer MR, Casper M. Disparities in temporal and geographic patterns of declining heart disease mortality by race and sex in the United States, 1973–2010. J Am Heart Assoc 2015;4(12):e002567.

9. Eayres D, Williams ES. Evaluation of methodologies for small area life expectancy estimation. J Epidemiol Community Health 2004;58(3):243–9.

10. Spratt SE, Batch BC, Davis LP, Dunham AA, Easterling M, Feinglos MN, et al. Methods and initial findings from the Durham Diabetes Coalition: integrating geospatial health technology and community interventions to reduce death and disability. J Clin Transl Endocrinol 2015;2(1):26–36.

11. Stephens AS, Purdie S, Yang B, Moore H. Life expectancy estimation in small administrative areas with non-uniform population sizes: application to Australian New South Wales local government areas. BMJ Open 2013;3(12):e003710.

12. Woods LM, Rachet B, Riga M, Stone N, Shah A, Coleman MP. Geographical variation in life expectancy at birth in England and Wales is largely explained by deprivation. J Epidemiol Community Health 2005;59(2):115–20.

13. Besag J, York J, Mollié A. Bayesian image restoration, with two applications in spatial statistics (with discussion). Ann Inst Stat Math 1991;43(1):1–20.

14. Gelfand A, Smith A. Sampling-based approaches to calculating marginal densities. J Am Stat Assoc 1990;85(410):398–409.

15. Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion). J R Stat Soc B 2009; 71(2):319–92.

16. Lunn D, Thomas A, Best N, Spiegelhalter D. WinBUGS — a Bayesian modelling framework: concepts, structure, and extensibility. Stat Comput 2000;10(4):325–37.

17. Clayton D, Kaldor J. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. Biometrics 1987;43(3):671–81.

18. Carlin B, Louis T. Bayesian methods for data analysis. Boca Raton (FL): CRC Press; 2009.

19. Klein R, Schoenborn C. Age adjustment using the 2000 projected U.S. population. Atlanta (GA): Centers for Disease Control and Prevention, Healthy People 2010 Stat Notes; 2001. https://www.cdc.gov/nchs/data/statnt/statnt20.pdf. Accessed February 19, 2019.

20. US Census Bureau. Section 1: population. in statistical abstract of the United States: 2011, 130 edition. Washington (DC): US Census Bureau; 2010. p. 1–62.

21. Fay MP, Feuer EJ. Confidence intervals for directly standardized rates: a method based on the gamma distribution. Stat Med 1997;16(7):791–801.

22. Tiwari RC, Clegg LX, Zou Z. Efficient interval estimation for age-adjusted cancer rates. Stat Methods Med Res 2006; 15(6):547–69.

23. Centers for Disease Control and Prevention, National Center for Health Statistics. Compressed mortality file 1979–1998. CDC WONDER on-line database, compiled from compressed mortality file CMF 1968–1988, series 20, no. 2A, 2000; and CMF 1989–1998, series 20, no. 2E, 2003.

24. US. Census Bureau. 2015 TIGER/Line shapefiles [machine-readable data files]. 2015. https://www.census.gov/geo/maps-data/data/tiger-line.html. Accessed February 19, 2019.

25. Centers for Disease Control and Prevention. CDC/ATSDR policy on releasing and sharing data. 2003. https://www.cdc.gov/maso/Policy/ReleasingData.pdf. Accessed June 30, 2015.

26. Helsel D. Much ado about next to nothing: incorporating nondetects in science. Ann Occup Hyg 2010;54(3):257–62.

27. Quick H, Holan S, Wikle C. Zeros and ones: a case for suppressing zeros in sensitive count data with an application to stroke mortality. Stat 2015;4(1):227–34.
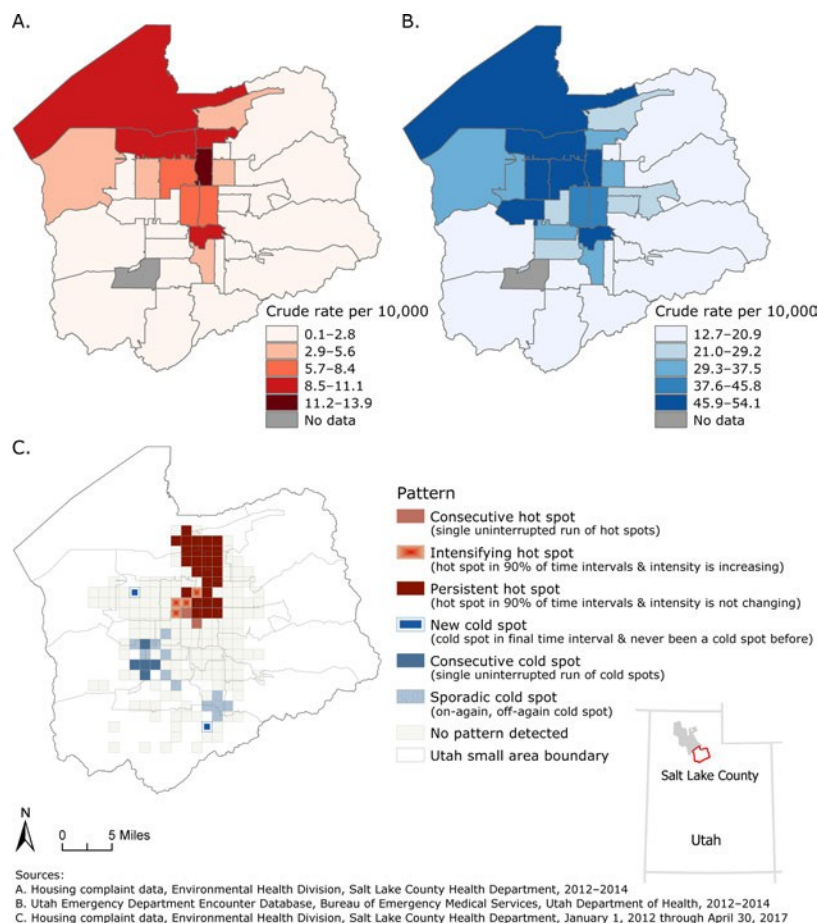
GIS SNAPSHOTS

Using Asthma-Related Housing Complaints to Target Residents With Uncontrolled Asthma in Salt Lake County, Utah

Jiyoung Byun, MPH[1]; Sophie McDonnell, BS[2]; Jenny Robertson, MSPH[1]

PEER REVIEWED

Sources:
A. Housing complaint data, Environmental Health Division, Salt Lake County Health Department, 2012–2014
B. Utah Emergency Department Encounter Database, Bureau of Emergency Medical Services, Utah Department of Health, 2012–2014
C. Housing complaint data, Environmental Health Division, Salt Lake County Health Department, January 1, 2012 through April 30, 2017

Maps A and B compare rates of asthma-related housing complaints and rates of asthma-related emergency department encounters, by small-area boundaries, Salt Lake County, Utah, 2012–2014. Map C depicts hot spots of asthma-related housing complaints that were identified in north-central Salt Lake County, Utah, January 1, 2012, through April 30, 2017.

Background

The prevalence of asthma is high in Salt Lake County, Utah; 9.5% of adults aged 18 years or older (1) and 6.7% of children and adolescents aged 0 to 17 years have asthma (2). On average, 1,800 adults and 1,200 children visit an emergency department (ED) (3) and 400 adults and 400 children are hospitalized (4) with a primary diagnosis of asthma each year.

In 2015, the Utah Asthma Program and partners from the Utah Asthma Task Force developed the Utah Asthma Home Visiting Program (UAHVP). This program serves families with uncontrolled asthma and is only available in Salt Lake and Utah counties (5). The Salt Lake County Health Department (SLCoHD) collaborated with the Utah Asthma Program to explore using recent and up-to-date housing complaint data to more efficiently target the UAHVP. Currently, the UAHVP is targeted in areas by using ED data, which have a reporting lag time of several years. In comparison, housing complaint data are collected in real time and readily accessible from the SLCoHD Environmental Health Division.

The goals of this project were to retrospectively identify asthma-related housing complaints, geocode these complaints, assess their relationship to the rate of asthma ED encounters, and analyze emerging hot spots to determine whether these data and methods could identify communities that may benefit from the UAHVP.

Methods

Housing complaints reported to the SLCoHD Environmental Health Division from January 1, 2012, to April 30, 2017, were manually reviewed and categorized as asthma-related if the complaint described an asthma trigger defined by the Centers for Disease Control and Prevention (CDC) (eg, smoke, mites, mold, pets, cockroaches, rodents, strong odors, cigarettes, birds, pollution) (6). The final data set included 1,959 asthma-related housing complaints. Geocoding and spatial analyses were performed by using ArcGIS Pro 2.0 (Esri). Ninety-nine percent of complaints were geocoded with a match score of 90 or higher, aggregated to a small area as defined by the Utah Department of Health (7), and used to calculate and map crude incidence rates.

Crude rates of asthma ED encounters from 2012 through 2014 were mapped by Utah small area and compared visually with crude rates of asthma-related housing complaints from 2012 through 2014 (8). The Pearson correlation coefficient between rates was calculated by using Microsoft Excel (Microsoft Corp) to determine the strength of the relationship.

We analyzed emerging hot spots (9) of asthma-related housing complaints by aggregating cases into space-time cubes of 6 months and 5,500 feet and evaluating trends over time by using a neighborhood distance of 11,000 feet and a time-step interval of 2. The appropriate distance band for the space-time cube was determined by plotting global Moran's I z-scores from spatial autocorrelation analysis using 1,000-foot intervals from 1,000 to 20,000 feet and identifying the distance with the highest z-score peak (5,500 feet).

Findings

Visual comparison suggested that the rate of asthma-related housing complaints was positively correlated with the rate of asthma ED encounters by small area. Correlation analysis supported this finding and indicated a strong positive relationship ($r = 0.77$). Analysis of emerging hot spots of asthma-related housing complaints identified consecutive, intensifying, and persistent hot spots in communities of north central Salt Lake County. These hot spots may reflect communities with older housing that may benefit from the resources provided by the UAHVP.

Action

Our findings demonstrate the potential of using asthma-related housing complaints as a current, proxy data source for measuring asthma burden and of analyzing emerging hot spots to target or expand the UAHVP. Next steps include investigating factors that explain the spatial pattern of asthma-related housing complaints. If the pattern can be explained by factors addressed in the UAHVP or by participating partners, such as Green and Healthy Homes, a national initiative to create safe and healthy homes for low-income families, the findings would provide additional support for the use of these data and methods to guide program decisions. Further exploratory work could investigate the types, number, and causes of asthma triggers occurring in hot spots, which could be useful for measuring severity and customizing asthma control strategies in neighborhoods.

This project had several limitations. First, we could not confirm that the positive relationship of asthma-related housing complaints with asthma ED encounters existed in recent years because we lacked recent data on ED encounters. Second, housing complaints are reported predominantly by renters, so asthma-related housing issues that may exist for homeowners were not captured. Third, the high rates of asthma ED encounters were likely influenced by the underlying spatial distribution of social determinants that contribute to asthma burden, such as low household income and barriers to health care access.

Author Information

Corresponding Author: Jenny Robertson, MSPH, Epidemiology Bureau, Salt Lake County Health Department, 610 S 200 E, Salt Lake City, UT 84111. Telephone: 385-468-4203. Email: JRobertson@slco.org.

Author Affiliations: [1]Epidemiology Bureau, Salt Lake County Health Department, Salt Lake City, Utah. [2]Health Promotion Bureau, Salt Lake County Health Department, West Jordan, Utah.

References

1. Public Health Indicator Based Information System (IBIS). 2017Health indicator report of asthma: adult prevalence. https://ibis.health.utah.gov/indicator/view/AsthAdltPrev.LHD.html. Accessed April 11, 2019.
2. Public Health Indicator Based Information System (IBIS). 2016–2017Health indicator report of asthma: child prevalence. https://ibis.health.utah.gov/indicator/view/AsthChiPrev.LHD.html. Accessed April 11, 2019.
3. Public Health Indicator Based Information System (IBIS). 2014Query results for emergency department encounter query module for Utah counties and local health districts-count-ED encounters. https://ibis.health.utah.gov/query/result/ed/EDCntyHospED/Count.html. Accessed August 13, 2018.
4. Public Health Indicator Based Information System (IBIS). 2014Query results for inpatient hospital discharge query module for Utah counties and local health districts-count. https://ibis.health.utah.gov/query/result/hddb/HDDBCntyHosp/CountHosp.html. Accessed August 13, 2018.
5. Utah Department of Health. Utah asthma home visiting program. http://health.utah.gov/asthma/homevisit/index.html. Accessed August 6, 2018.
6. Centers for Disease Control and Prevention. Common asthma triggers. https://www.cdc.gov/asthma/triggers.html. Accessed August 6, 2018.
7. Utah Department of Health. Small areas. https://ibis.health.utah.gov/pdf/resource/SmallAreas.pdf. Accessed August 6, 2018.
8. Public Health Indicator Based Information System (IBIS). 2014Query results for emergency department encounter query module for Utah small areas-crude rates-ED encounters. https://ibis.health.utah.gov/query/result/ed/EDSareaHospED/CrudeRate.html. Accessed August 6, 2018.
9. Environmental Systems Research Institute. Emerging hot spot analysis. http://pro.arcgis.com/en/pro-app/tool-reference/space-time-pattern-mining/emerginghotspots.htm. Accessed January 3, 2019.

ORIGINAL RESEARCH

Spatial Clustering of Suicide and Associated Community Characteristics, Idaho, 2010–2014

Ahmed M. Kassem, MBBCh, PhD[1,2]; Kris K. Carter, DVM, MVPM[2,3];

Christopher J. Johnson, MPH[4]; Christine G. Hahn, MD[2]

PEER REVIEWED

Summary

What is already known on this topic?

A few studies in the southeastern United States have identified spatial clusters of suicide at the county and census tract levels.

What is added by this report?

This study identified spatial clusters of suicide at the census block group level in Idaho, a northwestern rural state.

What are the implications for public health practice?

Because all of Idaho is federally designated as having a shortage of mental health providers, this study will inform stakeholders targeting Idaho communities with disproportionately high suicide rates at a more detailed level.

Abstract

Introduction

In 2015, Idaho had the fifth highest suicide rate in the United States. Little is known about the characteristics of areas in Idaho with high suicide rates. To aid suicide prevention efforts in the state, we sought to identify and characterize spatial clusters of suicide.

Methods

We obtained population data from the 2010 US Census and the 2010–2014 American Community Survey, analyzed data on sui-cides from death certificates, and used a discrete Poisson model in SaTScan to identify spatial clusters of suicide. We used logistic regression to examine associations between suicide clustering and population characteristics.

Results

We found 2 clusters of suicide during 2010–2014 that accounted for 70 (4.7%) of 1,501 suicides in Idaho. Areas within clusters were positively associated with the following population characteristics: median age $\leq$31.1 years versus >31.1 years (multivariable-adjusted odds ratio [aOR] = 2.4; 95% confidence interval [CI], 1.04–5.6), >53% female vs $\leq$53% female (aOR = 2.7; 95% CI, 1.3–5.8; P = .01), >1% American Indian/Alaska Native vs $\leq$1% American Indian/Alaska Native (aOR = 2.9; 95% CI, 1.4–6.3), and >30% never married vs $\leq$30% never married (aOR = 3.4; 95% CI, 1.5–8.0; P = .004).

Conclusion

Idaho suicide prevention programs should consider using results to target prevention efforts to communities with disproportionately high suicide rates.

Introduction

In the United States, suicide is the tenth leading cause of death; more than 44,000 suicides were reported in 2015 (1). From 2000 to 2015, the US age-adjusted suicide rate increased by 28%, from 10.4 per 100,000 population to 13.3 per 100,000 population (2). Suicide results in substantial medical and work-loss costs; lifetime costs were estimated to exceed $56 billion in 2015 (1); this conservative estimate did not account for underreporting of suicides and other societal costs (eg, pain and suffering, justice system). Beyond the economic burden, suicide negatively affects families and community members, who may have long-lasting mental health problems and other life-changing difficulties (3).

Suicide rates vary in the United States by geographic location. During 2011–2015, the age-adjusted suicide rate was higher in the West than in the Northeast (14.0 per 100,000 population [West census region] vs 9.8 per 100,000 population [Northeast census region]) (1). Although suicide rates increased across all levels of urbanization in the United States during 1999–2015, rates were higher in less urban areas than in more urban areas (4). Because geographic differences are not fully explained by demographic patterns (5), they could be attributed to other factors, such as lack of access or poor access to quality mental health care, low socioeconomic status, and weak social cohesion in areas with high suicide rates (6–8). Increased access to lethal means could be another explanatory factor in areas with higher suicide rates (9).

A comprehensive public health approach to suicide prevention, in contrast to an approach that focuses on mental health treatment, can address multiple risk factors across the lifespan (10). Although a public health suicide prevention approach is warranted in communities nationwide (10), it is essential to focus on communities with disproportionately high suicide rates to eliminate geographic disparities and reduce suicide altogether (11,12). Furthermore, examination of suicide data at a fine-scale geographic level is needed to identify these communities for efficient planning and targeting effective prevention efforts, especially when resources are limited.

Several types of suicide clusters have been reported, including mass clusters, space–time clusters, and spatial clusters (13). Spatial cluster analysis has been used to identify communities with disproportionately high suicide rates, because spatial cluster analysis overcomes the "small numbers problem" (in which rates for areas with small populations have wider variability and less reliability than rates for areas with large populations) inherent in spatial analysis and allows for statistical assessment of rates across geographic units (14). A study in 2012 found 2 high-risk spatial clusters of suicide during 1999–2008 that comprised 15 of 120 counties in Kentucky (15). Another study, in 2017, found 24 high-risk spatial clusters of suicide during 2001–2010 that comprised 491 of 3,154 census tracts in Florida (16). Studies of suicide in Scotland, Australia, São Paulo, and Québec used the same methodology (17–20). To our knowledge, no study of suicide using spatial cluster analysis has been conducted in rural or western parts of the United States.

In Idaho, a northwestern rural state with a population of 1.7 million, suicide is a major public health problem (21). Idaho consistently ranks among the top 10 states with the highest suicide rates, with an age-adjusted suicide rate of 22.2 per 100,000 population, compared with 13.3 per 100,000 population nationally in 2015 (1). Eighteen of 44 counties in Idaho had an age-adjusted suicide rate of 22.0 per 100,000 population or more during 2010–2014 (21).

However, these rates are likely unstable because of the small numbers problem (21). Because all of Idaho is federally designated as having a shortage of mental health providers (22), targeting Idaho communities with disproportionately high suicide rates at a more detailed level than the county level (because some counties are very large in area) is crucial. Therefore, we sought to identify and characterize areas with spatial clusters of suicide at the neighborhood level in Idaho. We examined whether there are geographic areas in Idaho that have statistically significant higher rates of suicide than expected, compared with other geographic areas in the state, and we explored their characteristics. For a complete representation of suicide in Idaho, we also described the epidemiology of residents who died by suicide.

Methods

We used a retrospective ecological study design to investigate suicides among Idaho residents during 2010–2014. We did not include suicides occurring in Idaho among out-of-state persons, because an objective of our study was to examine the characteristics of communities in which Idaho residents who died by suicide lived at the time of death. We used the census block group as a proxy for neighborhood. A census block group is a statistical division of a census tract that covers a contiguous area and generally has a population size of 600 to 3,000 people, whereas a census tract is a relatively permanent statistical subdivision of a county and generally has a population size of 1,200 to 8,000 people (23). Our study was deemed nonresearch public health practice by the Idaho Division of Public Health's Research Determination Committee.

We obtained individual-level data on suicides from death certificates stored by the Idaho Bureau of Vital Records and Health Statistics, and for the spatial cluster analysis, we aggregated data on suicides to census block group. Although some suicide reporting systems and research exclude suicides among persons younger than 10 years, we did not exclude any age group, in accordance with the standard practice in Idaho (21). We identified suicides by the established *International Classification of Diseases, Tenth Revision*, codes as follows: X60.0–X84.9, Y87.0, and U03.9 (24). Death certificates included information on sex, age, ethnicity, race, education, marital status, military status (based on the question "Ever in US Armed Forces?"), occupation, and mechanism of injury. We geocoded residential addresses from death certificates to obtain 15-digit census block group identifiers. We completed geocoding by using the Automated Geospatial Geocoding Interface Environment System (25). In total, 98.5% of residential addresses were matched to a census block group identifier; we excluded 23 suicides without a matched census block group identifier. We used

census block group identifiers to merge suicide data with other data sources.

We obtained data on population estimates from the 2010 US Census and data on community characteristics from the 2010–2014 American Community Survey's 5-year estimates (26). We measured the following community characteristics, suggested by previous studies (16–20), in proportions as appropriate: female; median age; American Indian or Alaska Native; Hispanic or Latino; persons never married; persons in single-parent families; persons with less than a high school education (ie, did not receive a regular high school diploma, GED, or alternative credential); unemployed persons, median household income; persons in poverty; persons in renter-occupied housing units; persons with disabilities; and persons with no health insurance. In addition to showing demographic patterns, these characteristics capture dimensions of social cohesion and economic deprivation that could be associated with suicide (16–20).

Data analysis

Using information from death certificates, we first calculated descriptive statistics of residents who died by suicide and stratified these data by sex. We used the Pearson χ^2 test for categorical variables (or Fisher exact test for <5 expected cell counts) and t tests for continuous variables. Next, we conducted spatial cluster analysis by using SaTScan version 9.4 (Martin Kulldorff and Information Management Services Inc), free software that uses scan statistics to identify clusters (27). We used the discrete Poisson model (28) to scan for nonoverlapping geographical areas (census block groups) with significantly high rates of suicide. In SaTScan, we used population size (default of 50% of the total population at risk) to specify the maximum spatial cluster size; circular spatial window shape, adjusted for sex and age distributions; and the default of 999 Monte Carlo replications. We selected the spatial clusters with $P < .10$ for the subsequent analyses.

We used logistic regression models to examine associations between community characteristics and suicide clustering. Suicide clustering was constructed as a binary outcome variable indicating whether a census block group belonged to a spatial cluster of suicide (with $P < .10$). To simplify interpretation and use of findings for a wider audience, we dichotomized each variable for community characteristics into high and low levels. Except for age and income, we constructed the variables to compare the highest quartile with the lowest 3 quartiles for each variable. For age and income, we constructed the variables to compare the lowest quartile with the highest 3 quartiles for each variable. We fit a series of univariable models to examine association of each community characteristic with suicide clustering. Community characteristics that were significant at $P < .05$ in the univariable models were in-

cluded in a multivariable model to identify the most important community characteristics related to suicide clustering. We performed model diagnostics, including goodness of fit and multicollinearity assessments, which did not indicate problems. We used SAS version 9.3 (SAS Institute Inc) for all statistical analyses other than spatial cluster analysis, and we used ArcGIS version 10 (Environmental Systems Research Institute, Inc) for cartographic displays of spatial clusters.

Results

During 2010–2014, 1,501 Idaho residents died by suicide. Most residents who died by suicide were male (78.5%), aged 35 to 64 years (53.7%), non-Hispanic (95.8%) and white (97.0%) (Table 1). Overall, male and female residents who died by suicide did not significantly differ by the demographic characteristics examined. However, they significantly differed by marital status, military status, occupational status, and suicide method. The proportion of divorced persons was higher among females (32.6%) than males (25.0%), and the proportion of persons never married was higher among males (32.9%) than females (27.0%). The proportion of those who served in the military was higher among males (26.8%) than among females (3.1%). The proportion of those who were homemakers and those who had never worked or were disabled was higher among females (19.7% and 5.0%, respectively) than males (0.2% and 2.7%, respectively). For mechanism of injury, males (67.6%) were more likely than females (34.4%) to die by a firearm, and females (36.2%) were more likely than males (11.1%) to die by poisoning.

Spatial clusters of census block groups with high suicide rates

SaTScan identified a "most likely" cluster and 9 secondary clusters (Table 2). The 2 identified spatial clusters (with $P < .10$) of census block groups with disproportionately high suicide rates during 2010–2014 accounted for 70 (4.7%) of 1,501 deaths by suicide (Figure). The "most likely" spatial cluster, comprising 25 census block groups and a population of 30,405, was found in southeastern Idaho. During 2010–2014, 54 suicides occurred in this spatial cluster, whereas 28 suicides were expected, indicating that the suicide rate was 90% higher inside the cluster than outside (relative risk = 1.9, $P = .04$). A secondary spatial cluster with $P < .10$ was identified in northeastern Idaho. This secondary spatial cluster, comprising 6 census block groups and a population of 4,391, had 16 suicides, whereas 4 suicides were expected. The suicide rate was more than 3 times higher inside this cluster than outside (relative risk = 3.6, $P = .06$).
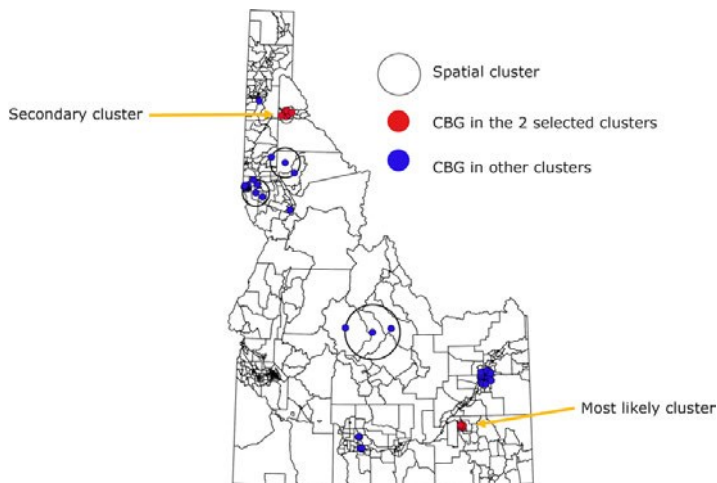
Figure. Spatial clusters of Idaho resident suicides by census block group, 2010–2014. A dot is a centroid of a census block group (CBG); 1 dot might represent 1 or more suicides that occurred in that CBG during the study period.

Characteristics of census block groups in spatial clusters

Compared with census block groups outside spatial clusters of suicide, census block groups in spatial clusters were more likely to have a higher proportion of females, American Indians or Alaska Natives, never married persons, and persons in poverty, and a lower proportion of persons with less than a high school education (Table 3). Census block groups within spatial clusters had populations with a younger median age and a lower median household income. We observed no significant differences between census block groups within spatial clusters and outside spatial clusters in proportion Hispanic or Latino ethnicity, single-parent families, unemployment, renter-occupied housing, disability, or health insurance coverage. In the multivariable model that included significant characteristics from the univariable models, the following community characteristics remained significant: median age ≤31.1 years (multivariable-adjusted odds ratio [aOR] = 2.4; 95% confidence interval [CI], 1.04–5.6; P = .04), >53% female (aOR = 2.7; 95% CI, 1.3–5.8; P = .01), >1% American Indian or Alaska Native (aOR = 2.9; 95% CI, 1.4–6.3; P = .006), and >30% never married (aOR = 3.4; 95% CI, 1.5–8.0).

Discussion

This ecological study identified geographic areas with disproportionately high suicide rates at the census block group level in 2 parts of Idaho. The communities in areas with suicide clustering had a unique demographic and socioeconomic profile. To our knowledge, this is the first study to investigate spatial clustering of suicide in the western region of the United States.

The 2 spatial clusters of census block groups identified were in 2 of the 18 counties where high rates of suicide had been reported (21). Identifying these clusters provides a more detailed view of geographic areas in these counties: 25 census block groups in a county with 60 census block groups, and 6 census block groups in a county with 18 census block groups (21). Our findings on spatial clusters of suicide at the census block group level cannot be fully compared with findings from previous studies, because those studies used different geographic units (counties and census tracts, not census block groups) (15–20). The proportion of geographic units that were part of the identified clusters was smaller in Idaho (3%) than they were in Kentucky (13%) (15) and Florida (16%) (16). Despite different levels of geography with varying population compositions, this finding might be attributed to differences in suicide risk levels in each state; a state where suicide risk has less geographic variation (eg, Idaho) is less likely to have many clusters. Our study spanned 5 years, which is half of the study period of other US studies (15,16); a longer study including more suicides might have identified more or fewer areas or same or different areas within spatial clusters.

Our findings are generally consistent with findings of other studies reporting that areas of lower socioeconomic status are associated with higher rates of suicide (7). We found a positive association between suicide clustering and both low household income and high proportion of persons in poverty; however, we found a negative association between suicide clustering and low educational attainment. This finding is consistent with at least 1 previous study that found the proportion of the population without a diploma is less likely to be included in a suicide cluster (20). Our finding that suicide clustering was associated with a higher proportion of never-married persons is consistent with research on the influence of social support and family structure on suicide (8). Community characteristics related to housing, unemployment, disability, and health insurance coverage that were not significantly associated with suicide clustering in our study might be investigated in future studies to confirm our findings. Overall, the unique demographic and socioeconomic profile of areas with suicide clustering in Idaho should be viewed as a potential way to depict an environmental context that is conducive to suicide, rather than a direct cause of suicide clustering.

The literature identified 2 possible explanations for suicide clustering. First, concentrations of persons at high risk for suicide might live in areas that could be identified as a cluster (compositional effects) (8). Second, place of residence might influence suicide risk

by being less supportive (eg, because of social or economic isolation) of persons at high risk (contextual effect) (8). Our objective was not to investigate causation, and we did not incorporate individual-level data to assess individual risk of suicide after controlling for contextual effect.

Our study demonstrates the feasibility of a state health department investigation of spatial clusters of suicide using multiple data sources. Strengths of this study include the use of population-based suicide data; use of the census block group as a granular, detailed unit of geographic analysis; and consideration of a broad range of community characteristics that covered the same period as the suicides. Spatial cluster analysis using SaTScan has many advantages, including adjusting for population inhomogeneity, adjusting for multiple comparisons, adjusting for covariates, and limiting preselection bias by not specifying cluster size a priori (27).

This study has several limitations. First, incorrectly not classifying suicide as a cause of death on death certification could have resulted in underreporting of suicide. Second, missing information on residential addresses resulted in incomplete geocoding; however, less than 2% of suicides were missing information on residential addresses. Third, we did not have information on how long the decedents lived in their homes; thus, we could not determine how duration of exposure to communities could affect results. Fourth, our cluster analysis was driven by the settings we selected in SaTScan; however, we followed the standard settings and those used in previous studies. Finally, our findings might not reflect current high-risk areas because data were from 2010–2014. However, retrospective analysis of mortality data is a fundamental tool for community health assessment, and we used the most recent available data. Although the contextual factors conducive to suicide in the identified clusters have probably not changed greatly since our study period, continuous evaluation and data triangulation to determine whether high-risk areas remain at high risk over time could increase confidence in public health programs that target prevention efforts to those areas. Although a study from Australia found that historical suicide clusters, detected during a 5-year period, predicted only 36% of suicide clusters detected during a subsequent 5-year period (29), our findings are better suited to inform current planning and response needs of suicide prevention programs rather than to predict future suicides.

Our findings could help public health practitioners and policy makers prioritize resources and target efforts for suicide prevention. The Centers for Disease Control and Prevention developed a technical package of prevention strategies to help communities use the best available evidence for suicide prevention (30). These strategies include strengthening economic supports; strengthening access and delivery of suicide care; creating protective environments; promoting connectedness; teaching coping and problem-

solving skills; and identifying and supporting people at risk (30). A multicomponent public health suicide prevention approach should address the needs of communities at the highest risk of suicide, such as communities we found in our study. In Idaho, a public health approach that strengthens economic supports and strengthens access and delivery of suicide care in the identified areas might be most effective in preventing suicide.

Acknowledgments

Author Information

Corresponding Author: Ahmed M. Kassem, MBBCh, PhD, 1600 Clifton Rd NE, Mailstop H24-2, Atlanta, GA 30329. Telephone: 404-718-6781. Email: akassem@cdc.gov.

Author Affiliations: [1]Epidemic Intelligence Service, Division of Scientific Education and Professional Development, Centers for Disease Control and Prevention, Atlanta, Georgia. [2]Division of Public Health, Idaho Department of Health and Welfare, Boise, Idaho. [3]Center for Preparedness and Response, Centers for Disease Control and Prevention, Atlanta, Georgia. [4]Cancer Data Registry of Idaho, Boise, Idaho.

References

1. US Department of Health and Human Services, Centers for Disease Control and Prevention. Web-Based Injury Statistics Query and Reporting System (WISQARS). http://www.cdc.gov/injury/wisqars/index.html. https://www.cdc.gov/injury/wisqars/cost/index.html. Accessed August 13, 2018.
2. QuickStats: age-adjusted rate* for suicide,† by sex — National Vital Statistics System, United States, 1975–2015. MMWR Morb Mortal Wkly Rep 2017;66(10):285.
3. Centers for Disease Control and Prevention. Understanding suicide: fact sheet. Atlanta (GA): National Center for Injury Prevention and Control; 2015. https://www.cdc.gov/violenceprevention/pdf/suicide_factsheet-a.pdf. Accessed February 14, 2019.
4. Kegler SR, Stone DM, Holland KM. Trends in suicide by level of urbanization — United States, 1999–2015. MMWR Morb Mortal Wkly Rep 2017;66(10):270–3.

5. Centers for Disease Control and Prevention. Regional variations in suicide rates — United States, 1990–1994. MMWR Morb Mortal Wkly Rep 1997;46(34):789–93.

6. Tondo L, Albert MJ, Baldessarini RJ. Suicide rates in relation to health care access in the United States: an ecological study. J Clin Psychiatry 2006;67(4):517–23.

7. Rehkopf DH, Buka SL. The association between suicide and the socio-economic characteristics of geographical areas: a systematic review. Psychol Med 2006;36(2):145–57.

8. Denney JT, Wadsworth T, Rogers RG, Pampel FC. Suicide in the city: do characteristics of place really influence risk? Soc Sci Q 2015;96(2):313–29.

9. Miller M, Barber C, White RA, Azrael D. Firearms and suicide in the United States: is risk independent of underlying suicidal behavior? Am J Epidemiol 2013;178(6):946–55.

10. David-Ferdon C, Crosby AE, Caine ED, Hindman J, Reed J, Iskander J. CDC Grand rounds: preventing suicide through a comprehensive public health approach. MMWR Morb Mortal Wkly Rep 2016;65(34):894–7.

11. National Action Alliance for Suicide Prevention, Office of the Surgeon General. 2012 National strategy for suicide prevention: goals and objectives for action. Washington (DC): US Department of Health and Human Services, Office of the Surgeon General; 2012.

12. National Prevention Council. National Prevention Strategy, Washington (DC): US Department of Health and Human Services, Office of the Surgeon General, 2011.

13. Haw C, Hawton K, Niedzwiedz C, Platt S. Suicide clusters: a review of risk factors and mechanisms. Suicide Life Threat Behav 2013;43(1):97–108.

14. Rezaeian M, Dunn G, St Leger S, Appleby L. Geographical epidemiology, spatial analysis and geographical information systems: a multidisciplinary glossary. J Epidemiol Community Health 2007;61(2):98–102.

15. Saman DM, Walsh S, Borówko A, Odoi A. Does place of residence affect risk of suicide? A spatial epidemiologic investigation in Kentucky from 1999 to 2008. BMC Public Health 2012;12(1):108.

16. Johnson AM, Woodside JM, Johnson A, Pollack JM. Spatial patterns and neighborhood characteristics of overall suicide clusters in Florida From 2001 to 2010. Am J Prev Med 2017; 52(1):e1–7.

17. Exeter DJ, Boyle PJ. Does young adult suicide cluster geographically in Scotland? J Epidemiol Community Health 2007;61(8):731–6.

18. Qi X, Hu W, Page A, Tong S. Spatial clusters of suicide in Australia. BMC Psychiatry 2012;12(1):86.

19. Bando DH, Moreira RS, Pereira JC, Barrozo LV. Spatial clusters of suicide in the municipality of São Paulo 1996–2005: an ecological study. BMC Psychiatry 2012;12(1):124.

20. Ngamini Ngui A, Apparicio P, Moltchanova E, Vasiliadis HM. Spatial analysis of suicide mortality in Québec: spatial clustering and area factor correlates. Psychiatry Res 2014; 220(1-2):20–30.

21. Idaho Department of Health and Welfare. Idaho resident suicide deaths 2010-2014 report. Boise (ID): Idaho Department of Health and Welfare, Division of Public Health; 2016.

22. Idaho Department of Health and Welfare. Get Healthy Idaho: measuring and improving population health. Boise (ID): Idaho Department of Health and Welfare, Division of Public Health; 2015.

23. US Census Bureau. 2010Geographic terms and concepts. https://www.census.gov/geo/reference/terms.html. Accessed February 14, 2019.

24. Centers for Disease Control and Prevention, National Center for Health Statistics. International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM). https://www.cdc.gov/nchs/icd/icd10cm.htm. Accessed February 19, 2019.

25. Goldberg DW, Kohler B, Kosary C. The Texas A&M, NAACCR, NCI Geocoding Service. 2016. http://geo.naaccr.org. Accessed February 19, 2019.

26. US Census Bureau. Our surveys and programs. https://www.census.gov/programs-surveys.html. Accessed February 14, 2019.

27. Kulldorff M, Information Management Services, Inc. SaTScan v9.4: software for the spatial and space-time scan statistics 2016. https://www.satscan.org. Accessed February 14, 2019.

28. Kulldorff M. A spatial scan statistic. Commun Stat Theory Methods 1997;26(6):1481–96.

29. Too LS, Pirkis J, Milner A, Robinson J, Spittal MJ. Clusters of suicidal events among young people: do clusters from one time period predict later clusters? Suicide Life Threat Behav 2018.

30. Stone DM, Holland KM, Bartholow BN, Crosby AE, Jack SPD, Wilkins N. Preventing suicide: a technical package of policies, programs and practices. Atlanta (GA): US Department of Health and Human Services, Centers for Disease Control and Prevention; 2017.

Tables

Table 1. Characteristics of Residents Who Died by Suicide, Stratified by Sex, Idaho, 2010–2014[a]

Characteristic	Total[b] (n = 1,501)	Male (n = 1,178)	Female (n = 323)	P Value[c]
Age, mean (SD), y	45.6 (18.7)	45.9 (19.4)	44.4 (16.3)	.15
Age group, n (%), y				
<15	20 (1.3)	18 (1.5)	2 (0.6)	
15–24	231 (15.4)	182 (15.5)	49 (15.2)	
25–34	212 (14.1)	171 (14.5)	41 (12.7)	
35–44	255 (17.0)	192 (16.3)	63 (19.5)	
45–54	303 (20.2)	226 (19.2)	77 (23.8)	.008
55–64	248 (16.5)	186 (15.8)	62 (19.2)	
65–74	117 (7.8)	99 (8.4)	18 (5.6)	
75–84	69 (4.6)	62 (5.3)	7 (2.2)	
≥85	46 (3.1)	42 (3.6)	4 (1.2)	
Ethnicity, n (%)				
Hispanic	63 (4.2)	47 (4.0)	16 (5.0)	.45
Non-Hispanic	1,437 (95.8)	1,130 (96.0)	307 (95.1)	
Race, n (%)				
White	1,456 (97.0)	1,146 (97.3)	310 (96.0)	
Black	4 (0.3)	4 (0.3)	0	
American Indian	24 (1.6)	15 (1.3)	9 (2.8)	.30
Asian Pacific Islander	8 (0.5)	6 (0.5)	2 (0.6)	
Other or mixed race	9 (0.6)	7 (0.6)	2 (0.6)	
Education, n (%)				
<High school[d]	264 (17.7)	210 (18.0)	54 (16.9)	
High school	607 (40.8)	488 (41.8)	119 (37.2)	.18
>High school	617 (41.5)	470 (40.2)	147 (45.9)	
Marital status, n (%)				
Married, including married but separated	537 (35.9)	423 (36.1)	114 (35.4)	
Widowed	87 (5.8)	71 (6.1)	16 (5.0)	.03
Divorced	398 (26.6)	293 (25.0)	105 (32.6)	
Never married	472 (31.6)	385 (32.9)	87 (27.0)	
Military status, n (%)				
Yes	323 (21.6)	313 (26.8)	10 (3.1)	<.001
No	1,170 (78.4)	857 (73.3)	313 (96.9)	

[a] Individual-level data on suicides obtained from death certificates stored by the Idaho Bureau of Vital Records and Health Statistics.
[b] The total number of participants for each variable varies because of missing values.
[c] Based on the Pearson χ^2 test or Fisher exact test for categorical variables and t test for continuous variables.
[d] Did not receive a regular high school diploma, GED, or alternative credential.
[e] Based on *International Classification of Diseases, Tenth Revision* (24). No death using the U03.9 ICD-10 code was reported.

(continued)

Table 1. Characteristics of Residents Who Died by Suicide, Stratified by Sex, Idaho, 2010–2014[a]

Characteristic	Total[b] (n = 1,501)	Male (n = 1,178)	Female (n = 323)	P Value[c]
Occupational status, n (%)				
Student	127 (8.6)	97 (8.3)	30 (9.4)	<.001
Homemaker, housewife	65 (4.4)	2 (0.2)	63 (19.7)	
Never worked, disabled	47 (3.2)	31 (2.7)	16 (5.0)	
Other occupational groups	1,246 (83.9)	1,035 (88.8)	211 (65.9)	
Mechanism of injury,[e] n (%)				
Poisoning (X60–X69)	248 (16.5)	131 (11.1)	117 (36.2)	<.001
Hanging, strangulation, suffocation, drowning and submersion (X70–X71)	294 (19.6)	216 (18.3)	78 (24.2)	
Firearm (X72–X74)	907 (60.4)	796 (67.6)	111 (34.4)	
Other methods (X75–X84, Y87)	52 (3.5)	35 (3.0)	17 (5.3)	

[a] Individual-level data on suicides obtained from death certificates stored by the Idaho Bureau of Vital Records and Health Statistics.
[b] The total number of participants for each variable varies because of missing values.
[c] Based on the Pearson χ^2 test or Fisher exact test for categorical variables and t test for continuous variables.
[d] Did not receive a regular high school diploma, GED, or alternative credential.
[e] Based on *International Classification of Diseases, Tenth Revision* (24). No death using the U03.9 ICD-10 code was reported.

Table 2. Spatial Clusters of Suicide by Residential Location, Idaho, 2010–2014[a]

Cluster No.	Cluster	No. of Census Block Groups	Population[b]	Observed No. of Suicide Deaths	Expected No. of Suicide Deaths	Annual Deaths per 100,000	Relative Risk	Log-Likelihood Ratio	P
1	Most likely	25	30,405	54	28.4	35.9	1.9	9.4	.04
2	Secondary	6	4,391	16	4.5	67.6	3.6	8.9	.06
3	Secondary	11	14,084	28	13.3	39.8	2.1	6.3	.55
4	Secondary	22	25,347	44	25.3	32.8	1.8	5.8	.69
5	Secondary	1	1,600–1,700	7	1.7	78.7	4.1	4.7	.95
6	Secondary	3	2,947	10	3.2	58.1	3.1	4.5	.97
7	Secondary	3	2,040	8	2.3	67.1	3.6	4.4	.98
8	Secondary	5	4,896	12	4.4	51.0	2.7	4.4	.98
9	Secondary	30	60,471	72	51.0	26.6	1.4	4.0	>.99
10	Secondary	1	500–600	4	0.6	117.9	6.3	4.0	>.99

[a] Individual-level data on suicides obtained from death certificates stored by the Idaho Bureau of Vital Records and Health Statistics. Data on population estimates obtained from the 2010 US Census and data on community characteristics from the 2010–2014 American Community Survey's 5-year estimates (26).
[b] Adjusted for sex and age.

Table 3. Characteristics of Census Block Groups Within and Outside Spatial Clusters of Suicide, Idaho, 2010–2014[a]

Characteristic[b]	Census Block Groups Within Spatial Clusters Of Suicide, No. (%) (n = 31)	Census Block Groups Outside Spatial Clusters Of Suicide, No. (%) (n = 932)	Odds Ratio (95% Confidence Interval)[c]
Proportion female			
>0.53	14 (45.2)	226 (24.3)	2.6 (1.3–5.3)
≤0.53	17 (54.8)	706 (75.8)	1 [Reference]
Median age			
≤31.1 y	16 (51.6)	223 (23.9)	3.4 (1.7–7.0)
>31.1 y	15 (48.4)	709 (76.1)	1 [Reference]
Proportion American Indian or Alaska Native			
>0.01	15 (48.4)	225 (24.1)	3.0 (1.4–6.1)
≤0.01	16 (51.6)	707 (75.9)	1 [Reference]
Proportion Hispanic or Latino			
>0.16	5 (16.1)	235 (25.2)	0.6 (0.2–1.5)
≤0.16	26 (83.9)	697 (74.8)	1 [Reference]
Proportion of never-married persons			
>0.30	19 (61.3)	221 (23.7)	5.1 (2.4–10.7)
≤0.30	12 (38.7)	711 (76.3)	1 [Reference]
Proportion of persons in single-parent families			
>0.24	12 (38.7)	228 (24.5)	2.0 (0.9–4.1)
≤0.24	19 (61.3)	704 (75.5)	1 [Reference]
Proportion of persons with <high school education[d]			
>0.16	2 (6.5)	238 (25.5)	0.2 (0.1–0.9)
≤0.16	29 (93.6)	694 (74.5)	1 [Reference]
Proportion of unemployed persons			
>0.07	11 (35.5)	229 (24.6)	1.7 (0.8–3.6)
≤0.07	20 (64.5)	703 (75.4)	1 [Reference]
Median household income, $			
≤35,345	14 (45.2)	226 (24.3)	2.6 (1.3–5.3)
>35,345	17 (54.8)	706 (75.8)	1 [Reference]
Proportion of persons in poverty			
>0.22	13 (41.9)	227 (24.4)	2.2 (1.1–4.7)
≤0.22	18 (58.1)	705 (75.6)	1 [Reference]
Proportion of persons in renter-occupied housing unit			
>0.41	12 (38.7)	228 (24.5)	2.0 (0.9–4.1)

[a] Individual-level data on suicides obtained from death certificates stored by the Idaho Bureau of Vital Records and Health Statistics. Data on population estimates obtained from the 2010 US Census and data on community characteristics from the 2010–2014 American Community Survey's 5-year estimates (25).
[b] Each variable for community characteristics was dichotomized into high and low levels. Except for age and income, we constructed the variables to compare the highest quartile with the lowest 3 quartiles for each variable. For age and income, we constructed the variables to compare the lowest quartile with the highest 3 quartiles for each variable.
[c] Based on Wald method from univariable logistic regression models.
[d] Did not receive a regular high school diploma, GED, or alternative credential.

(continued)

Table 3. Characteristics of Census Block Groups Within and Outside Spatial Clusters of Suicide, Idaho, 2010–2014[a]

Characteristic[b]	Census Block Groups Within Spatial Clusters Of Suicide, No. (%) (n = 31)	Census Block Groups Outside Spatial Clusters Of Suicide, No. (%) (n = 932)	Odds Ratio (95% Confidence Interval)[c]
≤0.41	19 (61.3)	704 (75.5)	1 [Reference]
Proportion of persons with disability			
>0.22	8 (25.8)	232 (24.9)	1.1 (0.5–2.4)
≤0.22	23 (74.2)	700 (75.1)	1 [Reference]
Proportion of persons with no health insurance coverage			
>0.23	7 (22.6)	233 (25.0)	0.9 (0.4–2.1)
≤0.23	24 (77.4)	699 (75.0)	1 [Reference]

[a] Individual-level data on suicides obtained from death certificates stored by the Idaho Bureau of Vital Records and Health Statistics. Data on population estimates obtained from the 2010 US Census and data on community characteristics from the 2010–2014 American Community Survey's 5-year estimates (25).
[b] Each variable for community characteristics was dichotomized into high and low levels. Except for age and income, we constructed the variables to compare the highest quartile with the lowest 3 quartiles for each variable. For age and income, we constructed the variables to compare the lowest quartile with the highest 3 quartiles for each variable.
[c] Based on Wald method from univariable logistic regression models.
[d] Did not receive a regular high school diploma, GED, or alternative credential.

ORIGINAL RESEARCH

Identifying County-Level All-Cause Mortality Rate Trajectories and Their Spatial Distribution Across the United States

Peter Baltrus, PhD[1,2]; Khusdeep Malhotra, BDS, MPH [3]; George Rust, MD[4]; Robert Levine, MD[5];

Chaohua Li, MPH[1]; Anne H. Gaglioti, MD[1,6]

PEER REVIEWED

Summary

What is already known about this topic?

All-cause mortality in the United States declined from 1935 through 2014, with a recent uptick in 2015. This national trend is composed of disparate local trends.

What is added by this report?

By using a novel methodology, we detected 8 unique county-level mortality rate trajectory groups. Disparities widened from 1999 to 2016. Differences existed in the demographic and socioeconomic profiles across the trajectory groups, with favorable mortality trajectories in the Northeast, in the Midwest, and on the West Coast and unfavorable trajectories concentrated in the Southeast.

What are the implications for public health practice?

Further investigation of the determinants of the trajectory groupings and the geographic outliers identified could inform interventions to achieve equitable distribution of county mortality rates.

Abstract

Introduction

All-cause mortality in the United States declined from 1935 through 2014, with a recent uptick in 2015. This national trend is composed of disparate local trends. We identified distinct groups of all-cause mortality rate trajectories by grouping US counties with similar temporal trajectories.

Methods

We used all-cause mortality rates in all US counties for 1999 through 2016 and estimated discrete mixture models by using county level mortality rates. Proc Traj in SAS was used to detect how county trajectories clustered into groups on the basis of similar intercepts, slopes, and higher order terms. Models with increasing numbers of groups were assessed on the basis of model fit. We created county-level maps of mortality trajectory groups by using ArcGIS.

Results

Eight unique trajectory groups were detected among 3,091 counties. The average mortality rate in the most favorable trajectory group declined 29.4%, from 592.3 deaths per 100,000 in 1999 to 418.2 in 2016. The least favorable mortality trajectory group declined 3.4% over the period, from 1,280.3 deaths per 100,000 to 1,236.9. We saw significant differences in the demographic and socioeconomic profiles and geographic patterns across the trajectory categories, with favorable mortality trajectories in the Northeast, Midwest, and on the West Coast and unfavorable trajectories concentrated in the Southeast.

Conclusions

County-level disparities in all-cause mortality rates widened over the past 18 years. Further investigation of the determinants of the trajectory groupings and the geographic outliers identified by our research could inform interventions to achieve equitable distribution of county mortality rates.

Introduction

The all-cause mortality rate is an indicator of general population health. The age-adjusted all-cause mortality rate declined in the US general population from 1935 (1) to a record low in 2014 (2). A notable 1.1% increase occurred in the age-adjusted all-cause mortality rate in 2015 (3). Overall declines in mortality rates did

not occur in all geographic areas (4); southeastern states had higher rates overall and lower rates of decline compared with the national trend (3).

Although mortality rate trends differ by state, it is important to study mortality and mortality trends at smaller geographic levels. Although use of counties as a geographic unit of analysis has limitations (5,6) and county-level infrastructure is variable, counties are the smallest unit of analysis for which stable mortality rates can be calculated and for which infrastructure exists for implementing and administering health and social policies. County mortality rates vary by geography (7,8), but few analyses of all-cause mortality rate trends have been done at the county level. Although some methods are available to compare and analyze long-term trends in mortality rates that include joinpoint regression, spatial and aspatial generalized linear mixed models, and Bayesian space–time models, all these approaches rely on the change in the rates being compared to exhibit linear or log linear changes over time and rely on a series of changes between small intervals over the entire time period (9).

We sought to group and examine common trends in county-level mortality for the most recently available mortality data (1999–2016) by using a new statistical method called group-based trajectory modeling (GBTM). Although trends in US counties were previously reported by examining the difference in rates at 2 time points and linear or log linear changes in rates over time, GBTM incorporates information from all time points and allows for examination of nonlinear (quadratic, cubic, and other higher order) rate trends. GBTM determines if groups of study units with similar trajectory shapes exist and has been used to determine whether the health outcome trends of individual units group together into patterns (10–13). To our knowledge, this method has not been used to examine mortality rates in US counties.

We sought to identify patterns of county mortality rate trajectories and to determine if any positive (exceptionally low initial rates decreasing rapidly) or negative (exceptionally high rates decreasing slowly or not at all) deviant trajectories existed. We also estimated the extent to which trajectories clustered geographically. Finally, we identified geographic deviants: counties whose mortality rate trajectory group patterns were significantly different than the trajectories of surrounding counties.

Methods

County-level, age adjusted mortality data from the Compressed Mortality File was obtained for years 1999 through 2016 from the National Center for Health Statistics through a data use agreement

(14). We included all deaths across the entire age spectrum. We included rates for each year in which the number of deaths in a county was greater than or equal to 20. Counties were included in the analysis if they had at least 2 years of stable mortality rate data.

The yearly, age-adjusted, all-cause mortality rate of the county was the outcome measure used to generate rate trajectories using Proc Traj for SAS, version 9.4 (SAS Institute, Inc) (15,16). Group-based trajectory modeling assumes that a certain number of discrete underlying groups in the population each have their own population prevalence, intercept, and slope and possibly higher order terms (17). These subpopulations are not directly observable but are estimated (latent).

Proc Traj requires specification of the number of groups the model will fit. We estimated a quadratic model with a dependent variable of mortality rate and an independent variable of time in years with a single group and kept adding groups and assessing the change in the Bayesian information criterion (BIC) as an evaluation of model fit (15,18). We simultaneously assessed the percentage of counties in each group and the shape of the trajectories when plotted. The fit of the model increased with the addition of more groups. The model with 8 groups produced both a negative deviant group and a positive deviant group (defined as being less than 2% of the counties and substantially different upon visual inspection from the other trajectories). Group 1 was the positive deviant group whereas group 8 was the negative deviant group, both having trajectories with substantially lower rates (group 1) or higher rates (group 8) than the rest of the trajectories (Figure 1). Identification of such groups was one of the aims of our study; adding a greater number of groups did not affect the composition of these 2 groups, nor did it identify any new deviant groups. Including more than 8 groups only created more roughly parallel groups between group 2 and 7, some with very small numbers of counties. The BIC continued to increase with the addition of more groups beyond 8 (Appendix A), but on the basis of the foregoing considerations we stopped at 8 groups for ease of interpretation of the data. For sensitivity analysis, we repeated the process with linear models as the starting point. Trajectory groups looked similar to linear models, but the quadratic models produced a better fit according to the BIC. We next added or removed second and higher-order terms from each group's model on the basis of significance ($P < .05$). This process yielded quadratic models for trajectories 1, 2, and 8. Trajectories 3 through 7 included a cubic term.
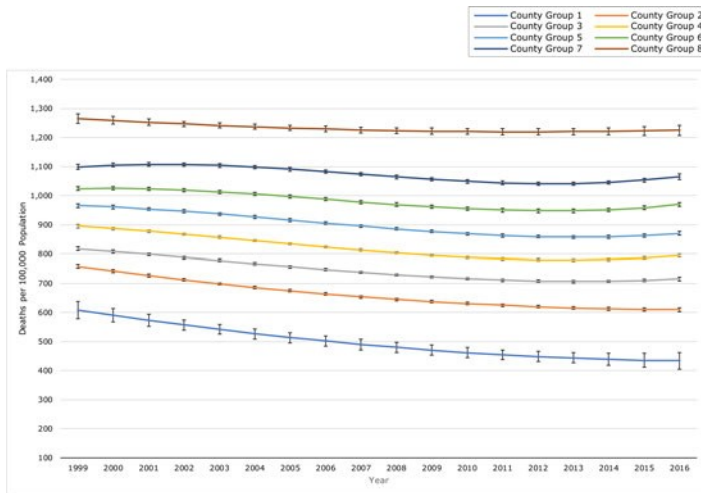
Figure 1. Age-adjusted mortality rate trajectories for US counties for 8 groups of counties based on group-based trajectory models, 1999–2016. The outcome measure used to generate rate trajectories was the yearly, age-adjusted, all-cause mortality rate of the county. Solid lines correspond to model-predicted values for rates; dotted lines are confidence intervals for the predicted values.

We used US census data for 2000 and 2010 to describe the changes in sociodemographic composition of the county trajectory groups. Variables included total population, population density (population per square mile), median age, percentage of county population living below the federal poverty level, median household income, percentage white population, percentage black population, percentage American Indian/Alaska Native population, percentage Asian population, and percentage Hispanic (any race) population. We reported means for each year and changes of means between the years.

We created a choropleth map of the county trajectory groups (Figure 2). Thematic mapping of county trajectories showed clear evidence of spatial autocorrelation. This simply means that observations that are located next to each other are related to each other, that is, there is no spatial independence between observations. We measured the degree of spatial autocorrelation (ie, the degree to which neighboring observations are related to each other) by using the Global Moran's I statistic of ArcGIS Pro (Esri). We used 2 method to determine the number of neighbors for each observation: polygon contiguity (based on neighbors sharing borders) and inverse distance (which means the farther away a neighbor is, the less influence the neighbor has) (19,20). Once we determined the number and relationship of neighbors, we identified local clusters by using the local indicators of spatial association (LISA) technique (19). The LISA technique generates a statistic named Getis-Ord Gi* (Esri), which specifies where features with high or low values cluster. Significant clusters were those where a feature and

its neighbors all had high Getis-Ord Gi* values. Geographic deviants were defined as counties that had much higher or much lower values than their neighboring counties. On the basis of a county's relative position within a cluster, counties were grouped into 4 categories of significant spatial clusters ($P < .05$): 1) high–high clusters representing all counties with high mortality, the worst trajectory group; 2) high–low clusters representing counties in the worst trajectory groups near counties in the most favorable trajectory groups (at-risk counties doing worse than those around them); 3) low–high clusters representing counties in the best trajectory groups near counties in the worst trajectory groups (resilient counties doing better than those around them); and 4) low–low clusters of counties in the most favorable trajectory groups. Of 3,144 counties, 3,091 counties and county equivalents were included in the analysis.
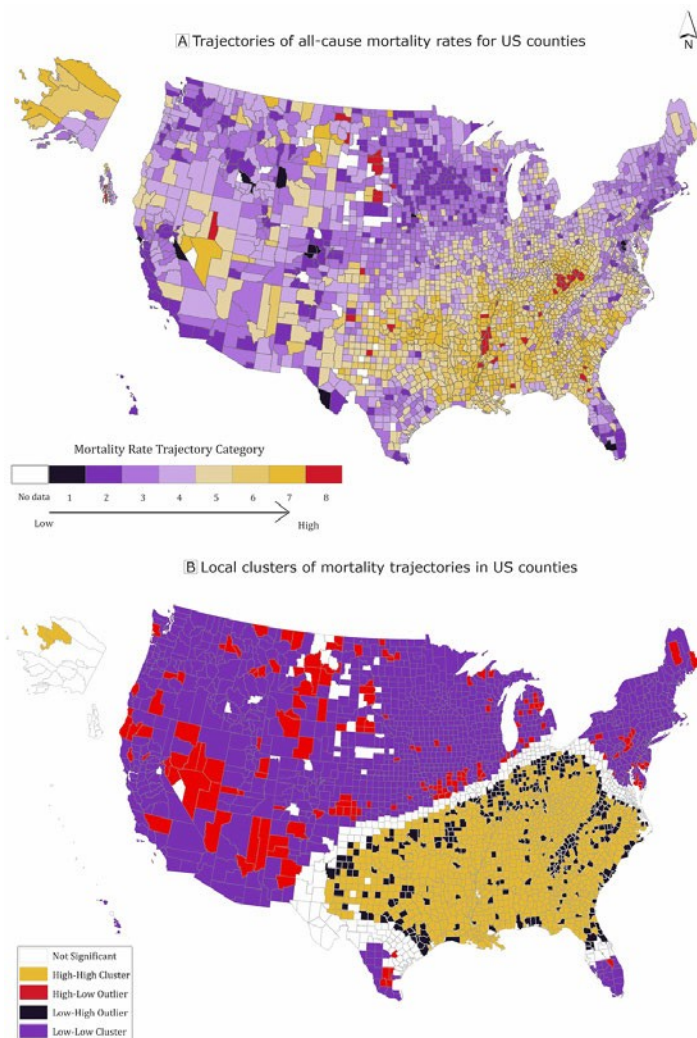
Figure 2. Trajectories of age-adjusted all-cause mortality in US counties using group-based trajectory models, 1999–2016. The outcome measure used to generate rate trajectories was the yearly, age-adjusted, all-cause mortality rate of the county. Panel A: Trajectories of all-cause mortality rates for US counties. Panel B: Local clusters of mortality trajectories in US counties detected by using local indicators of spatial association (LISA). The 4 categories of significant spatial clusters (*P* < .05): 1) high–high clusters representing all counties with high mortality, the worst trajectory group; 2) high–low outliers representing counties in the worst trajectory groups near counties in the most favorable trajectory groups (at-risk counties doing worse than those around them); 3) low–high outliers representing counties in the best trajectory groups near counties in the worst trajectory groups (resilient counties doing better than those around them); and 4) low–low clusters of counties in the most favorable trajectory groups. Source: 1999–2016 Compressed Mortality File, Centers for Disease Control and Prevention (14).

Results

The equations for trajectories 1, 2, and 8 included quadratic terms, which produced trajectories with slower mortality rates decline over time (Figure 1). The equations for trajectories 3 through 7 contained a cubic term and produced trajectories that had a slowing rate of decline in rates with increasing rates near the end of the study period (Table 1). The numeric ordering of trajectories reflects mortality rate trajectories from most favorable to least favorable. Trajectory 1 had the lowest average mortality rate at the beginning of the study (1999) and at the end of the study (2016) and the steepest decline over the study period. Trajectory 8 had the highest mortality rates at both time points and only a modest decline over the study period. The trajectories did not overlap, which indicates that disparities in mortality rates across the trajectory groups persisted throughout the study period.

Disparities between trajectory groups increased over the study period. At baseline, the average mortality rate for Trajectory 1 was 592.3 deaths per 100,000, decreasing by 29.4% to 418.2 deaths per 100,000, whereas Trajectory 8 had a baseline rate of 1,280.3 deaths per 100,000 and decreased by 3.4% over the 18-year period to 1,236.9 deaths per 100,000 (Table 2). These 2 groups had a difference of 688 deaths per 100,000 in 1999 that increased to a difference of 818.7 deaths per 100,000 in 2016. There was a graded association in the amount of change in rates across the trajectory groups; as baseline rates increased, the rate decline decreased.

Sociodemographic characteristics of county trajectory groupings were similar for 2000 and 2010. A graded association with median income and poverty was noted across trajectory groups. Median income decreased and percentage of county population living below the federal poverty level increased as health trajectories worsened (Table 2). A more complex relationship was observed with racial composition of mortality trajectory groupings. The county percentage of black population increased from trajectory 1 to trajectory 7. Percentage of white population increased across Trajectories 1 to 2, peaked at Trajectory 3, and then decreased from Trajectory 3 to 8. The percentage of American Indian/Alaska Native population increased across trajectory groups, peaking in Trajectory 8 (2000,11.8%; 2010,12.7%). The percentage of Asian and Hispanic populations in county trajectory groups increased as trajectories became more favorable.

Panel A of Figure 2 depicts the geographic variation of mortality rate trajectory groups. The Southeast was characterized by counties in high mortality rate trajectory groups, whereas counties in low mortality trajectory groups tended to be in the Northeast, the upper Midwest, and the West Coast. This pattern was reflec-

ted in the clustering of counties detected by using LISA (Panel B of Figure 2). Clusters of the most favorable trajectory counties (low–low) and counties with worse trajectories than neighboring counties (high–low) were in the northern, midwestern, and western regions. Clusters of the least favorable trajectory counties (high–high) and clusters of counties with significantly better trajectories than their neighboring counties (low–high) were predominantly in the south. This indicates that while some regions of the country may be doing well or poorly in terms of mortality there are counties with substantially different mortality trajectory patterns than their geographic neighbors.

We identified positive and negative deviant county groups. Trajectory 1 (positive deviant, n = 14) had substantially lower mortality rates than the middle 6 trajectories, and trajectory 8 (negative deviant, n = 50) had substantially higher mortality rates than the middle 6 trajectories during the study period. Positive deviant counties tended to be wealthy except for Presidio County, Texas, a small, West Texas county bordering the Rio Grande River with a largely Hispanic (83.4%) population. Two trajectory 1 counties were in the Washington, District of Columbia, metropolitan area and 3 in Colorado; the remaining trajectory 1 counties were dispersed throughout the country. Several negative deviant counties were identified with differing demographic characteristics, but most had high poverty rates. Counties in North Dakota (n = 1) and South Dakota (n = 5) had large Native American populations, counties in the Mississippi Delta (n = 8) had large black populations, and an Appalachian cluster in Kentucky (n = 14) and West Virginia (n = 4) was predominantly white.

Discussion

We used a new application of group-based trajectory modeling to identify groups of US counties with similar temporal trajectories of all-cause mortality rates. This national analysis over an 18-year period identified 8 distinct trajectory groups. Within those trajectories, we identified groups of positive and negative deviant counties. This work presents a new approach to identifying and quantifying spatiotemporal trends in health disparities that addresses limitations of current approaches. First, this approach overcomes the limitation of relying on linear or log-linear rate changes over time by allowing for higher order terms in the equations used to generate trajectories. Second, this approach allows the use of all rates in a period instead of relying on change between rates at 2 points within an overall period. Third, this approach groups trajectory patterns that emerge from the data used to support the analysis instead of relying on an a priori trend categorization. Our main findings show substantial and widening inequities in mortality rates and mortality rate trends across groups of US counties. We saw geographic clustering of the trajectories, with

worse trajectories clustering in the Southeast and better trajectories clustering in the Northeast, the upper Midwest, and the West Coast. Local-area variation in mortality has been well-documented in the United States (7,8). However, identification of clusters of counties with similar mortality rate trajectories over time contributes to understanding the factors that drive such differences. Demographic factors such as racial composition and socioeconomic status have been demonstrated and partially explain high mortality trajectories and less favorable mortality trajectories in the South (21).

In our analysis, several counties in trajectory 8, the worst mortality trajectory group, have disproportionately large American Indian populations. For example, Sioux County, North Dakota, rests entirely within the Standing Rock Indian Reservation. Buffalo County, South Dakota, where the Cow Creek Sioux Tribe resides, had the highest 2016 all-cause mortality rate and the lowest per capita income in the United States. This may be because American Indians have higher rates of mortality across the lifespan than other racial/ethnic groups (22–24). Additionally, the economic and social conditions on reservations may contribute to a higher mortality rate and a less favorable temporal mortality rate trajectory for American Indians living on reservations compared with those living in other areas of the country.

Historically disenfranchised places in the Mississippi Delta, where there were high concentrations of slavery followed by the structural inequities of sharecropping and segregation (25), and in Appalachia, where poverty and environmental and occupational injustice is entrenched (26), had a disproportionate number of trajectory 8 counties. One study found similar spatial clustering of poor physical and mental health and food insecurity in these areas (27). Counties in trajectory 8 that were not part of geographic clusters may have unique factors that explain their poor mortality rates and rate trajectories that warrant further exploration. The size of the rate gap between trajectory 8 counties and the other trajectory groupings is cause for concern, further study, and action.

Counties in the best trajectory group, trajectory 1, had generally higher socioeconomic conditions than other parts of the country, but not uniformly so. Marin County, California; Los Alamos, New Mexico; Montgomery County, Maryland; and Fairfax County, Virginia, ranked in the top 20 counties in the nation by median income. No other county in the top 25 median-income counties for the nation was found in this best outcome group, so high socioeconomic status may not be enough to predict favorable mortality trajectory trends. Other counties in the group had a less affluent socioeconomic profile. For example, although Collier County, Florida, includes affluent communities such as Naples and Marco Is-

land, it also included vast rural areas with large numbers of migrant farmworkers and had an overall median income less than half that of the most affluent counties in the nation.

Multilevel influences potentially contribute to the differences we observed across groups of mortality rate trajectories. Changes in socioeconomic status, demographic composition, health care infrastructure, patterns of health care use, health behaviors, and changes in state and federal health, housing, education, and social policy could all be contributing factors. One demographic compositional change we noted was that the largest percentage and change in percentage of Hispanic populations occurred in counties with the best mortality outcomes. This may be due to the documented "Hispanic paradox" in health outcomes (28,29).

Although we saw a significant geographic clustering of counties in each trajectory, some counties with low mortality rate trajectories were in the same geographic area as those with high rate trajectories (and vice versa). These counties may be considered positive deviants, having achieved more optimal mortality rates and rate trends despite being surrounded by counties with worse mortality rates and less improvement over time. If these positive deviance communities have common characteristics amenable to intervention, they could reveal a path toward achieving improved outcomes in counties with unfavorable trajectory patterns. Alternatively, these positive deviant counties may be surrounded by counties with significantly different demographic composition, health care access, or rurality, and such differences also may account for the differences in mortality trajectories observed in our analysis.

Our study has several limitations. We chose to use age-adjusted mortality rates for everyone without stratifying by age, sex, or race to create an overall indicator of public health in US counties because of the large amount of space required to present a description of this novel methodology for the first time. Preliminary analysis of different age and racial/ethnic groups has indeed revealed nonuniform trends (Appendix B), which we intend to discuss in future articles. By studying all-cause mortality, differences in specific causes of death would possibly cause different trajectory groupings and geographic patterns. On the other hand, all-cause mortality is less subject to many of the known limitations of death certificate data. We have only begun to tease out the myriad explanatory factors for these differences in outcomes. Although geographic granularity is limited in this county-level analysis, smaller neighborhood-level analyses may produce unstable rates and may be difficult to interpret on a national level. There are also limitations in interpreting the results of the statistical models. Trajectory 1 contained only 14 counties, but these counties had a greater than 98% probability of belonging to group 1, indicating that they are true outliers. All counties had a greater than 50% probability

of membership in their assigned group, and misclassification would likely result in being assigned to the trajectory above or below the one reported. More groups could have been added to the model, but this would have improved the model fit minimally without providing more information to inform interventions.

Further research should examine what county level factors are associated with the observed patterns in county groupings of mortality rate trajectories identified here. Demographic, socioeconomic, and health system variables as well as social variables such as social capital and social cohesion should be examined. Although traditional regression models will be helpful, we suggest that a more comprehensive approach be taken to determine how these variables interact to produce the observed patterns. Such an approach will require the use of longitudinal data on the predictor variables and modeling approaches including multilevel modeling, structural equations, and system dynamic models.

That county disparities in temporal, all-cause mortality rate trends are worsening suggests that we need to quickly learn the reasons why some counties succeed in reducing mortality rates while others fail. The lessons learned from successful counties could be applied to those that are failing. The identification of positive geographic outliers may provide an opportunity to learn what factors may be driving exceptional outcomes. Hopefully, investigating these special cases will lead to knowledge to help improve the health outcomes of lagging counties and thereby reduce county level disparities in the all-cause mortality trends observed here.

Acknowledgments

Author Information

Corresponding Author: Peter Baltrus, PhD, Morehouse School of Medicine, 720 Westview Dr, SW, Atlanta, GA 30310. Telephone: 404-752-1189. Email: pbaltrus@msm.edu.

Author Affiliations: [1]National Center for Primary Care, Morehouse School of Medicine, Atlanta, Georgia. [2]Department of Community Health and Preventive Medicine, Morehouse School

of Medicine, Atlanta, Georgia. [3]Geography and Urban Studies, Temple University, Philadelphia, Pennsylvania. [4]Department of Behavioral Sciences and Social Medicine, Florida State University School of Medicine, Tallahassee, Florida. [5]Department of Family and Community Medicine, Baylor College of Medicine, Houston, Texas. [6]Department of Family Medicine, Morehouse School of Medicine, Atlanta, Georgia.

References

1. Hoyert D. 75 years of mortality in the United States, 1935–2010. NCHS Data Brief .Vol 88. Hyattsville (MD): National Center for Health Statistics; 2012.

2. Kochanek KD, Murphy SL, Xu J, Tejada-Vera B. Deaths: final data for 2014. Natl Vital Stat Rep 2016;65(4):1–122.

3. Murphy SL, Xu J, Kochanek KD, Curtin SC, Arias E. Deaths: final data for 2015. 2017. https://stacks.cdc.gov/view/cdc/50011. Accessed March 22, 2019.

4. Levine R, Briggs N, Husaini B, Hennekens C. Geographic studies of black–white mortality. In: Satcher D, Pamies R, editors. Multicultural medicine and health disparities. New York (NY): MacMillan and Co; 2006.

5. Krieger N, Chen JT, Waterman PD, Soobader M-J, Subramanian SV, Carson R. Geocoding and monitoring of US socioeconomic inequalities in mortality and cancer incidence: does the choice of area-based measure and geographic level matter?: the Public Health Disparities Geocoding Project. Am J Epidemiol 2002;156(5):471–82.

6. Chen JT, Rehkopf DH, Waterman PD, Subramanian SV, Coull BA, Cohen B, et al. Mapping and measuring social disparities in premature mortality: the impact of census tract poverty within and across Boston neighborhoods, 1999–2001. J Urban Health 2006;83(6):1063–84.

7. Murray CJ, Kulkarni SC, Michaud C, Tomijima N, Bulzacchelli MT, Iandiorio TJ, et al. Eight Americas: investigating mortality disparities across races, counties, and race-counties in the United States. PLoS Med 2006;3(9):e260. Erratum in: PLoS Med 2006;3(12):e545. .

8. Cullen MR, Cummins C, Fuchs VR. Geographic and racial variation in premature mortality in the U.S.: analyzing the disparities. PLoS One 2012;7(4):e32930.

9. Vaughan AS, Kramer MR, Waller LA, Schieb LJ, Greer S, Casper M. Comparing methods of measuring geographic patterns in temporal trends: an application to county-level heart disease mortality in the United States, 1973 to 2010. Ann Epidemiol 2015;25(5):329–335.e3.

10. Ben-Assuli O, Padman R, Bowman M, Leshno M, Shabtai I. On analyzing readmissions using a trajectory model: evidence from Israel. Stud Health Technol Inform 2015;216:1063.

11. Hybels CF, Bennett JM, Landerman LR, Liang J, Plassman BL, Wu B. Trajectories of depressive symptoms and oral health outcomes in a community sample of older adults. Int J Geriatr Psychiatry 2016;31(1):83–91.

12. Kwon S, Lee J, Carnethon MR. Developmental trajectories of physical activity and television viewing during adolescence among girls: National Growth and Health Cohort Study. BMC Public Health 2015;15(1):667.

13. Mostazir M, Jeffery A, Voss L, Wilkin T. Childhood obesity: evidence for distinct early and late environmental determinants a 12-year longitudinal cohort study (EarlyBird 62). Int J Obes 2015;39(7):1057–62. Erratum in: Int J Obes (Lond) 2016; 40(2):380.

14. National Center for Health Statistics. Compressed mortality file, 1999–2016 (machine readable data file and documentation, CD-ROM Series 20, No. 2V) as compiled from data provided by the 57 vital statistics jurisdictions through the Vital Statistics Cooperative Program. Hyattsville (MD): National Center for Health Statistics; 2017.

15. Jones B, Nagin D, Roeder K. A SAS procedure based on mixture models for estimating developmental trajectories. Sociol Methods Res 2001;29(3):374–93.

16. Jones B, Nagin D. Advances in group-based trajectory modeling and an SAS procedure for estimating them. Sociol Methods Res 2007;35(4):542–71.

17. Nagin D. Analyzing developmental trajectories: semi-parametric, group-based approach. Psychol Methods 1999; 4(2):139–57.

18. Nagin D. Group-based modeling of development. Boston (MA): Harvard University Press; 2005.

19. Getis A. A geographic approach to identifying disease clusters. In: Janelle DG WB, Hansen K, editors. Worldminds: geographical perspectives on 100 problems: commemorating the 100th anniversary of the Association of American Geographers 1904–2004. Norwell (MA): Kluwer Academic Publishers; 2004.

20. Zhou Y, Hallisey EJ, Freymann GR. Identifying perinatal risk factors for infant maltreatment: an ecological approach. Int J Health Geogr 2006;5(1):53.

21. 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):600–13.

22. Wingate MS, Barfield WD, Smith RA, Petrini J. Perinatal disparities between American Indians and Alaska Natives and other US populations: comparative changes in fetal and first day mortality, 1995–2008. Matern Child Health J 2015; 19(8):1802–12.

23. Howard G, Peace F, Howard VJ. The contributions of selected diseases to disparities in death rates and years of life lost for racial/ethnic minorities in the United States, 1999–2010. Prev Chronic Dis 2014;11:E129.

24. Hutchinson RN, Shin S. Systematic review of health disparities for cardiovascular diseases and associated factors among American Indian and Alaska Native populations. PLoS One 2014;9(1):e80973.

25. Compiled from the census of 1860. (n.d.). Washington (DC): Library of Congress. https://www.loc.gov/resource/g3861e.cw0013200/?r=-0.195,-0.065,1.428,0.887,0and https://www.loc.gov/resource/g3861e.cw0013200/?r= -0.195,-0.065,1.428,0.887,0. Accessed August 24, 2018.

26. Isserman A. Appalachia then and now: update of "The Realities of Deprivation" reported to the president in 1964. J Appalach Stud 1997;3(1):43–69. http://www.arc.gov/assets/research_reports/socioeconomicreviewofappalachiathenandnow.pdf. Accessed March 21, 2019

27. Leonard T, Hughes AE, Donegan C, Santillan A, Pruitt SL. Overlapping geographic clusters of food security and health: where do social determinants and health outcomes converge in the U.S? SSM Popul Health 2018;5:160–70.

28. Franzini L, Ribble JC, Keddie AM. Understanding the Hispanic paradox. Ethn Dis 2001;11(3):496–518.

29. Palloni A, Arias E. Paradox lost: explaining the Hispanic adult mortality advantage. Demography 2004;41(3):385–415.

Tables

Table 1. Coefficients for Estimated Trajectories From Group-Based Trajectory Models Using 1999–2016 US County Annual All-Cause Mortality Data[a]

Trajectory[b]	Intercept[c] (*P* Value)	Slope[d] (*P* Value)	Quadratic[e] (*P* Value)	Cubic[f] (*P* Value)	% of US Counties	No. of Counties
1	628.51 (<.001)	−19.92 (<.001)	0.51 (.02)	NA	0.5	14
2	775.94 (<.001)	−17.78 (<.001)	0.47 (<.001)	NA	9.4	290
3	826.74 (<.001)	−8.19 (<.001)	−0.48 (.03)	0.03 (<.001)	19.7	608
4	901.42 (<.001)	−5.31 (.002)	−0.96 (<.001)	0.05 (<.001)	25.2	780
5	968.73 (<.001)	−1.53 (.42)	−1.25 (<.001)	.05 (<.001)	20.3	626
6	1,020.59 (<.001)	5.95 (.007)	−1.88 (<.001)	0.08 (<.001)	15.1	467
7	1,087.72 (<.001)	13.58 (<.001)	−2.55 (<.001)	0.10 (<.001)	8.2	252
8	1,273.51 (<.001)	−7.73 (.002)	0.28 (.03)	NA	1.7	54

Abbreviation: NA, not applicable.

[a] Coefficients are from an 8-group model; coefficients were added or removed from models if *P* < .05 for the coefficient. Note that if a term became nonsignificant when a higher-order term was added to the model and significant, the nonsignificant lower-order term remained in the model. Data are from the Centers for Disease Control and Prevention's Compressed Mortality File (14).

[b] The numeric ordering of trajectories reflects mortality rates from most favorable (lowest baseline rate/largest decline in rate) to least favorable (highest baseline rate/largest decline in rate).

[c] Baseline mortality rate estimated by the model.

[d] First order term estimated by model; represents the linear component of change in rate per year.

[e] Second order term estimated by model; represents the quadratic component of change in rate per year.

[f] Third order term estimated by model (if significant); represents the cubic component of change in rate per year.

Table 2. Age-Adjusted, All-Cause Mortality Rates and Demographic Characteristics of the 1999–2016 County Mortality Trajectory Groups[a]

County Group	1	2	3	4	5	6	7	8	Overall
Counties, n (%)	14 (0.5)	290 (9.4)	608 (19.7)	780 (25.2)	626 (20.3)	467 (15.1)	252 (8.2)	54 (1.7)	3091 (100)
Mortality rate per 100,000									
1999	592.3	748.6	813.4	893.2	961.6	1,026.7	1,104.9	1,280.3	920.1
2016	418.2	604.7	709.5	789.1	863.3	964.3	1,059.9	1,236.9	825.2
Change	−173.9	−143.9	−103.9	−104.1	−98.3	−62.4	−45.0	−43.4	−94.9
% Change	−29.4	−19.2	−12.8	−11.7	−10.2	−6.1	−4.1	−3.4	−10.3
Population									
2000	180,108	231,865	106,542	99,164	62,933	46,473	30,796	20,059	91,196
2010	202,341	255,498	120,595	109,220	68,043	47,651	30,859	18,939	100,053
Change	22,234	24,324	13,888	10,087	5,111	1,088	633	−857	8,853
Population density[b]									
2000	370.3	839.7	224.3	212.8	167.7	144.1	146.0	68.1	247.3
2010	413.0	903.5	247.3	227.8	181.2	143.8	143.0	64.5	264.2
Change	42.7	63.8	22.6	15.0	13.5	−0.3	−3.1	−2.4	17.2
Median age, y									
2000	36.8	37.4	37.9	37.3	37.4	36.9	36.4	33.4	37.2
2010	39.9	40.6	41.1	40.4	40.1	39.5	39.3	37.3	40.2
Change	3.1	3.2	3.2	3.0	2.7	2.6	2.9	3.8	2.9
Persons with income below federal poverty level, %									
2000	8.0	9.0	10.1	11.9	14.1	16.6	19.6	23.9	13.3
2010	10.5	11.7	13.1	15.3	17.8	20.5	24.4	28.8	16.8
Change	2.4	2.8	3.0	3.3	3.7	3.9	4.8	4.9	3.5
Median annual household income[c], $									
2000	83,248	66,912	59,120	55,160	50,636	45,790	41,360	37,317	53,400
2010	76,184	62,960	55,007	50,575	46,311	42,078	37,987	36,653	49,308
Change	−7,064	−3,952	−4,113	−4,585	−4,325	−3,712	−3,373	−664	−4,092
Race/ethnicity, %									
White									
2000	86.6	88.2	91.1	87.9	83.4	77.8	71.2	70.1	84.4
2010	83.4	86.1	89.5	86.6	82.0	76.4	69.4	68.4	82.9
Change	−3.2	−2.1	−1.7	−1.3	−1.4	−1.4	−1.7	−1.4	−1.5
Black									
2000	2.5	2.9	2.4	5.2	10.9	16.5	22.3	16.3	8.9
2010	3.0	3.2	2.8	5.4	10.9	16.3	22.5	16.2	9.0
Change	0.4	0.3	0.4	0.2	−0.1	−0.2	0.2	−0.2	0.1

[a] Data are from the Centers for Disease Control and Prevention's Compressed Mortality File (14). Sociodemographic data are from the 2000 and 2010 US Census.
[b] People per square mile.
[c] 2018 dollars.
[d] Any race.

(continued)

Table 2. Age-Adjusted, All-Cause Mortality Rates and Demographic Characteristics of the 1999–2016 County Mortality Trajectory Groups[a]

County Group	1	2	3	4	5	6	7	8	Overall
Asian									
2000	2.8	2.6	1.0	0.7	0.5	0.4	0.3	0.3	0.8
2010	3.8	3.4	1.4	1.0	0.7	0.6	0.4	0.3	1.2
Change	1.0	0.8	0.4	0.3	0.2	0.2	0.1	0.1	0.3
American Indian/Alaska Native									
2000	0.6	0.8	1.2	1.4	1.2	2.2	3.9	11.8	1.8
2010	0.6	0.9	1.3	1.5	1.3	2.4	4.1	12.7	1.9
Change	0.1	0.1	0.1	0.1	0.1	0.2	0.2	0.4	0.1
Hispanic[d]									
2000	16.3	8.3	7.0	7.3	6.2	3.8	2.3	2.00	6.1
2010	21.2	10.8	9.3	9.5	8.5	5.8	3.5	2.7	8.3
Change	4.9	2.5	2.3	2.2	2.3	2.0	1.2	0.7	2.1

[a] Data are from the Centers for Disease Control and Prevention's Compressed Mortality File (14). Sociodemographic data are from the 2000 and 2010 US Census.
[b] People per square mile.
[c] 2018 dollars.
[d] Any race.

Appendix A. Bayesian Information Criterion For Models With 1 to 30 Groups of Counties[a]

Number of Group	BIC	ΔBIC[b]
1	−353,621.1	Not applicable
2	−33,8347.9	15,273.2
3	−33,2614.9	5,733
4	−32,9808.4	2,806.5
5	−32,8178.7	1,629.7
6	−32,7229.1	949.6
7	−32,6665.9	563.2
8	−32,6316.9	349.0
9	−32,5754.4	562.5
10	−32,5551.7	202.7
11	−32,5567.8	−16.1
12	−32,5230.3	337.5
13	−32,5012.4	217.9
14	−32,4841.7	170.7
15	−32,4743.4	98.3
16	−32,4601.8	141.6
17	−32,4544.4	57.4
18	−32,4481.9	62.5
19	−32,4438.9	43.0
20	−32,4460.2	−21.3
21	−32,4318.0	142.2
22	−32,4311.6	6.4
23	−32,4204.7	106.9
24	−32,4214.5	−9.8
25	−32,4179.4	35.1
26	−32,4121.1	58.3
27	−32,4084.9	36.2
28	−32,4188.9	−104.0
29	−32,4123.2	65.7
30	−32,4152.1	−28.9

[a] All models are quadratic.

[b] $\Delta BIC = BIC_{group = k+1} - BIC_{group = k}$.

Appendix B. Trajectories for County All-Cause Mortality Rates, 1999–2016, by Race/Ethnicity, Age Groups, and Sex.

This appendix is available for download at

https://www.msm.edu/Research/research_centersandinstitutes/NCPC2/documents/publications/Preventing-Chronic-Disease-Appendix-B.pdf

GIS SNAPSHOTS

Tracking Senior Fall and Fall-Related Injury EMS Calls to Target Fall Prevention Programs, Salt Lake County, Utah

Jiyoung Byun, MPH[1]; Jenny Robertson, MSPH[1]

PEER REVIEWED

Counts and Rates of Fall-Related EMS Calls for Adults Aged ≥65 Years, by Census Tract, Salt Lake County, Utah, January 1, 2014–April 30, 2017
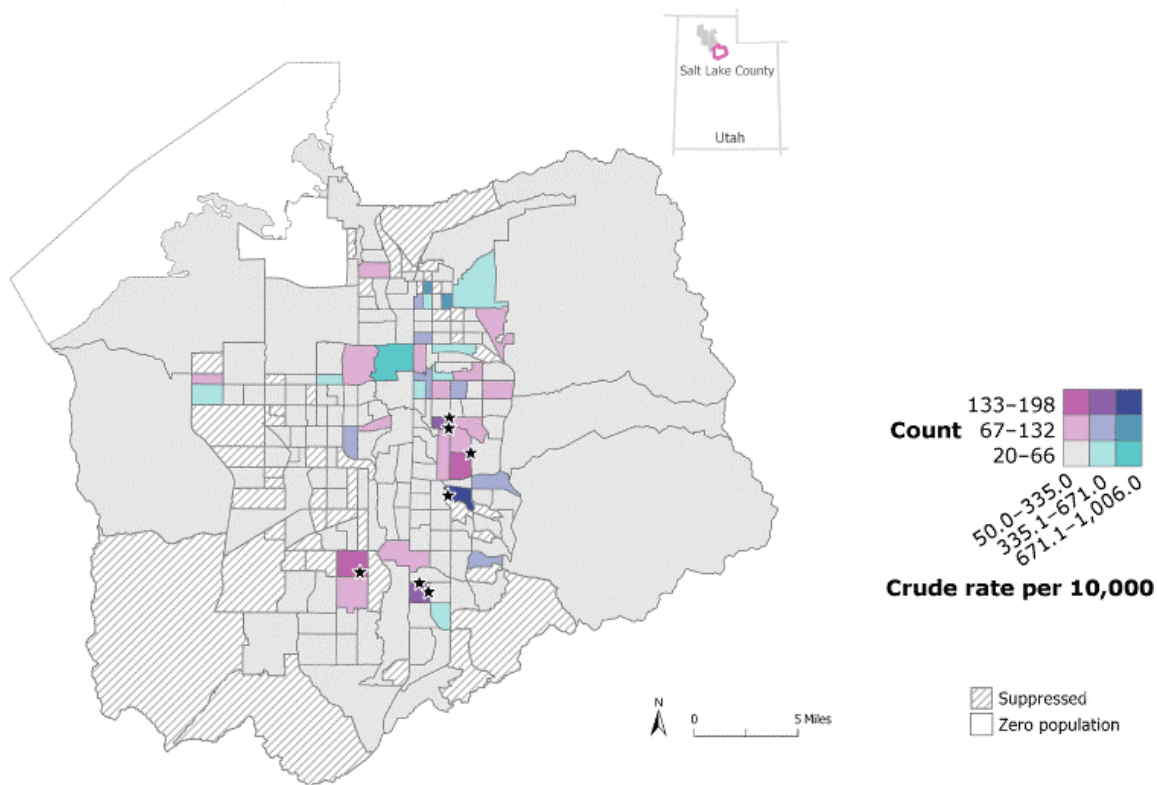
*<20 falls or relative standard error of crude rate of fall-related EMS calls >30%
Source: Prehospital data, Bureau of Emergency Medical Services and Preparedness, Utah Department of Health

The map displays counts and rates, by census tract, of fall-related EMS calls among adults aged 65 or older in Salt Lake County, Utah, from January 1, 2014, through April 30, 2017. Stars indicate senior living residences with high fall burden. Data were suppressed in areas that had fewer than 20 falls or the relative standard error of the crude rate of fall-related EMS calls was higher than 30%. Abbreviation: EMS, emergency medical services.

Background

Every year in the United States, more than 25% of people aged 65 or older fall at least once (1), and these injuries are associated with high rates of illness and death (2). Physiological age-related changes such as reduction of sight, hearing, and muscle strength are major causes of falls among older people (3). Among Salt Lake County residents aged 65 or older, in 2014 the fall injury emergency department encounter rate was 458.6 per 10,000 (4) and the fall injury hospitalization rate was 130.0 per 10,000 (5), and in 2016 the fall mortality rate was 5.4 per 10,000 (6).

As part of ongoing fall prevention activities, the Salt Lake County Health Department hosts the evidence-based Stepping On program for seniors. To better target this and other fall prevention programs to reduce rates of fall-related illness and death, we mapped dispatched emergency medical services (EMS) calls for falls and fall-related injuries among adults aged 65 years or older and identified areas with high prevalence.

Methods

We extracted data on EMS calls from the Utah prehospital reporting system that had an incident address in Salt Lake County; a date of incident from January 1, 2014, through April 30, 2017; and a dispatch report of 1) a fall or 2) an unconscious/fainting or unknown problem/person down. Those with a dispatch report of a fall were assumed to be correctly classified as fall-related. Narratives of those with a dispatch report of an unconscious/fainting or unknown problem/person down were searched by using SAS version 9.4 (SAS Institute, Inc) for terms suggesting evidence of a fall, such as "fall," "fell," or "GLF" (ground level fall). Calls were excluded if the narrative included terms indicating no fall, such as "no fall," "negative fall," "not sustain a fall," "denies (any) fall," or "not suffer a GLF."

The final data set included 14,824 fall-related injuries. Of those, 93% could be geocoded (96% of those geocoded with match score ≥90) and aggregated to the census tract level by using ArcGIS Pro 2.0 (Esri). Crude incidence rates were calculated by using American Community Survey 5-year population estimates for adults aged 65 years or older from 2014 through 2016, and mapped by census tract. Both counts and rates were classified by using equal intervals. Fall injury points were overlaid on census tract rates, and Google Maps (Google LLC) was used to explore neighborhoods with high counts or crude rates to identify facilities where fall prevention activities may be targeted.

Institutional review board (IRB) approval was obtained for the EMS data set from the Utah Department of Health IRB Committee.

Findings

Fall injury counts among adults aged 65 years or older were highest in census tracts in southeast and southwest Salt Lake County. Fall injury rates among adults aged 65 years or older were highest in census tracts in north-central and southeast Salt Lake County. Seven facilities were identified as locations of falls in these high-count or high-rate areas; all were mixed-level senior living residences (eg, independent living, assisted living, memory care). One census tract with a high rate did not have a senior living facility; most of these falls occurred in individual homes because of safety hazards. All census tracts with a high count had at least one senior living facility in which most falls occurred.

Action

Results were used by community partners to secure pilot funding for the Otago Exercise Program, and they are currently being used to target Stepping On and Otago programs, collaborate with Salt Lake County Aging and Adult Services' Meals on Wheels program to better reach the senior population vulnerable to falls, develop one-on-one prevention programs at sites with a high prevalence of falls, implement collaborative fall prevention programs with EMS community paramedicine programs, and evaluate program interventions. Interventions target individuals and include easily modifiable risk factors such as muscle strengthening and balance retraining exercises, medication review, vision and hearing checks, and improving safety around the home. However, public health would do well to partner with other sectors, such as city planning, to improve the built environment for seniors.

Geospatial data can be challenging to interpret, and various analyses and visualizations should be assessed together for the most accurate picture. Assessing rates without also examining counts may lead to inappropriate resource allocation because of the size of the population aged 65 or older in certain census tracts. For example, one census tract had a high rate but low count because the population aged 65 years or older in this tract was small. Similarly, census tracts with a low rate and high count indicate areas where the population aged 65 or older is large. Resources allocated to high-rate/low-count tracts may have a lesser impact in reducing the burden of falls than resources allocated to low-rate/high-count tracts. Ultimately, program managers found count data most useful for targeting resources to locations. In the future, it would be useful to compare locations by calculating rates by facility.

Limitations of this project include incomplete 2017 prehospital data resulting from a reporting delay, missing 2015–2016 prehospital data from one EMS agency because of data submission issues, and potential misclassification of incidents as fall-related or not fall-related.

Author Information

Corresponding Author: Jenny Robertson, MSPH, Epidemiology Bureau, Salt Lake County Health Department, 610 S 200 E, Salt Lake City, UT 84111. Telephone: 385-468-4203. Email: JRobertson@slco.org.

Author Affiliations: [1]Epidemiology Bureau, Salt Lake County Health Department, Salt Lake City, Utah.

References

1. Centers for Disease Control and Prevention. Important facts about falls. https://www.cdc.gov/homeandrecreationalsafety/falls/adultfalls.html. Accessed August 13, 2018.
2. Sterling DA, O'Connor JA, Bonadies J. Geriatric falls: injury severity is high and disproportionate to mechanism. J Trauma 2001;50(1):116–9.
3. Ungar A, Rafanelli M, Iacomelli I, Brunetti MA, Ceccofiglio A, Tesi F, et al. Fall prevention in the elderly. Clin Cases Miner Bone Metab 2013;10(2):91–5.
4. Public Health Indicator Based Information System (IBIS). 2014Query results for injury emergency department encounter query module for Utah counties and local health districts. Crude rates. ED encounters. https://ibis.health.utah.gov/query/result/ed/InjEDCntyHospED/CrudeRate.html. Accessed August 13, 2018.
5. Public Health Indicator Based Information System (IBIS). 2014Query results for inpatient hospital discharge query module for Utah counties and local health districts, crude rates. https://ibis.health.utah.gov/query/result/hddb/InjHDDBCntyHosp/CrudeRateHosp.html. Accessed August 9, 2018.
6. Public Health Indicator Based Information System (IBIS). 2016Query results for mortality ICD-10 query module for Utah counties and local health districts, crude rates, deaths per 100,000 population. https://ibis.health.utah.gov/query/result/mort/InjMortCntyICD10/CrudeRate.html. Accessed August 9, 2018.
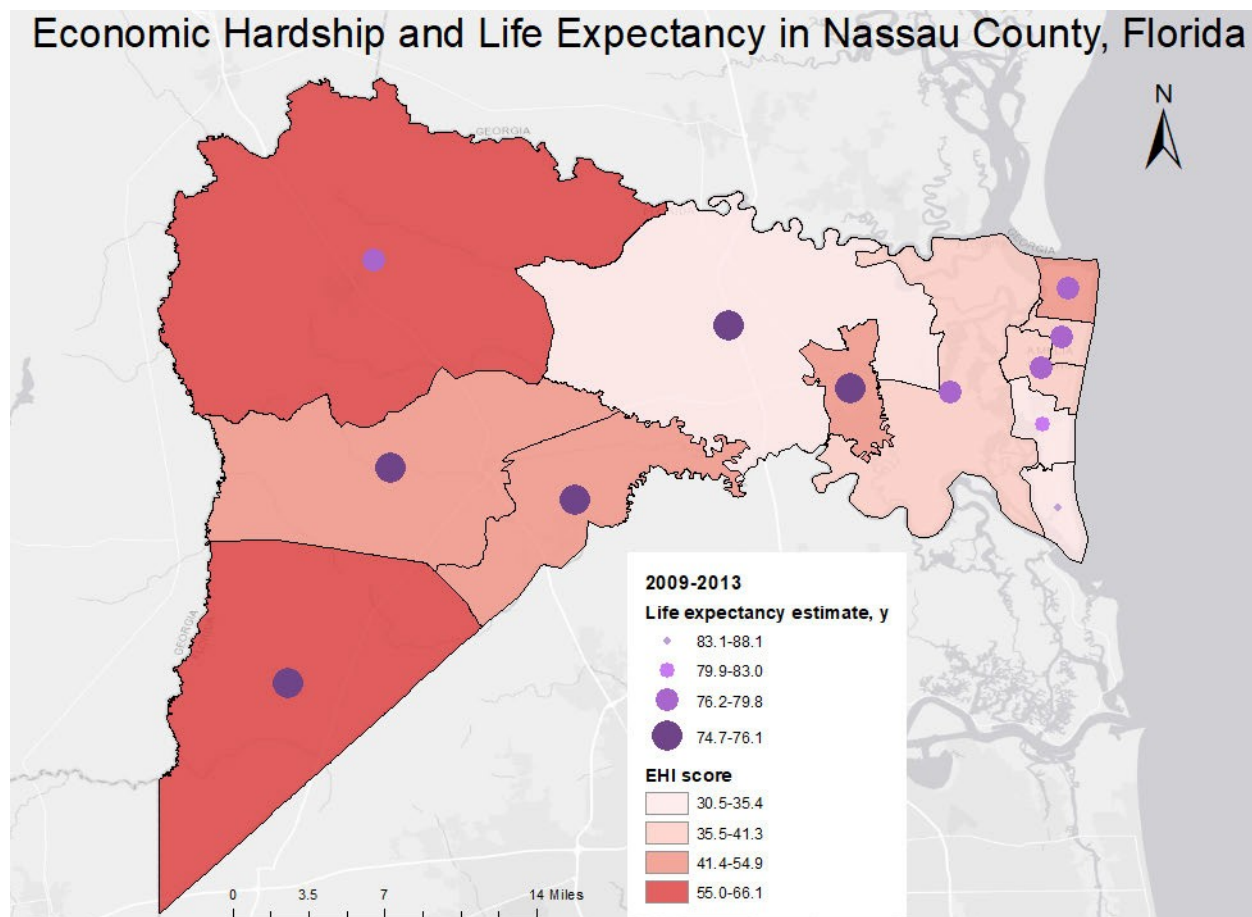
GIS SNAPSHOTS

Economic Hardship and Life Expectancy in Nassau County, Florida

Jessica Joiner, MPH[1]; Melissa Jordan, MS, CPH[1]; Keshia Reid, PhD[1];

Kristina Kintziger, PhD[2]; Chris Duclos, MS[1]

PEER REVIEWED

Economic Hardship Index (EHI) scores and life expectancy estimates at birth, by census tract, in Nassau County, Florida. EHI scores and life expectancy estimates were grouped into 4 classes by using Jenks natural breaks and combined to produce a bivariate map. EHI scores range from 0 to 100, with higher scores indicating worse economic conditions. The map highlights marked differences in economic conditions and life expectancy between census tracts. The map was shared with county health department staff members in response to calls for locally relevant data to address health disparities.

Background

The Economic Hardship Index (EHI) combines 6 social and economic measures to provide a more complete picture of socioeconomic conditions in a community than any one measure alone (1,2). It has the advantage of comprising data that operate at the community level, rather than the individual or family level, and it allows for a comparison of one community relative to another (or itself) over time (3). Socioeconomic conditions in a community are strongly associated with health (4). Economic hardship can affect a person's ability to access important health care services and lead a healthy lifestyle. Local estimates of life expectancy are useful in understanding the contributions of socioeconomic conditions to population health (2). Life expectancy data allow the examination of health disparities by place, because it reflects the combined effect of major illnesses and injuries and their underlying causes, including social and environmental determinants of health (5).

In response to calls for relevant and timely community-level indicators that address underlying causes of illness and injury and advance health equity, the Florida Department of Health used the EHI to test a bivariate mapping technique that combined data on economic hardship and life expectancy at the census-tract level in Nassau County, Florida. Nassau County is in northeastern Florida along the Atlantic coast. It has a land area of 725.9 square miles and a population of 73,314 (6). The average population per census tract in Nassau County is 5,640 (6).

Methods

We calculated the EHI by using 6 indicators from the US Census Bureau's 2014 American Community Survey (ACS) 5-year estimates: unemployment (percentage of the population aged ≥16 who were unemployed), population dependency (percentage of the population aged <18 or >64), educational attainment (percentage of the population aged ≥25 with less than a high school diploma), per capita income, crowded housing (percentage of occupied housing units with >1 person per room), and poverty (percentage of persons living below the federal poverty level). We standardized these indicators within each tract to give them equal weight and combined them into a composite score. Scores range from 0 to 100, with higher scores indicating worse economic conditions. Additional methodologic details about the index are reported elsewhere (3).

We calculated life expectancy estimates by using 5 years (2009–2013) of aggregated mortality data geocoded to 2010 census tract areas. We used the adjusted Chiang II method to gen- erate life expectancy estimates for all tracts (7). This method uses

the life table approach and assumes that deaths are spread evenly throughout each age period. It also handles zero deaths in a given age category and is adjusted to account for variance in the last age interval — all of which are important considerations in calculating life expectancy estimates (7). We suppressed life expectancy estimates with a standard error of 2 years or more because of low numbers of deaths or small populations.

We used ArcMap 10.3.1 for Desktop (Esri) to join EHI scores and life expectancy estimates to the 2010 census tract shapefile and produce a bivariate map that displays life expectancy as graduated circles and EHI scores as a choropleth map. Each variable was categorized into 4 classes by using the Jenks natural breaks method. We performed correlation analysis in SAS version 9.4 (SAS Institute Inc).

Findings

Average life expectancy in Nassau County was 77.9 years (95% confidence interval [CI], 77.4–78.5 y), comparable to the national average of 78.7 years (8). We observed a gap of approximately 13 years between the tract with the shortest life expectancy (74.7 [95% CI, 73.1–76.2] y) and the tract with the longest life expectancy (88.1 [95% CI, 84.8–91.5] y). EHI scores ranged from a low of 30.5 (least hardship) to a high of 66.1 (greatest hardship). Most tracts followed the expected pattern, such that areas with higher levels of economic hardship generally had lower life expectancy. Overall, tracts with the highest levels of economic hardship and lower life expectancy were concentrated on the eastern side of the county. The 2 tracts with the lowest level of economic hardship and highest life expectancy were along the coast and shared a boundary. A simple correlation analysis showed a moderate negative association between life expectancy and economic hardship (r = −0.494, P = .10), although this association was not significant because of the small sample size (n = 12).

Action

The Florida Department of Health's Health Equity Program Council (HEPC) Data and Assessment Subcommittee has piloted maps for several counties. These maps were shared with county health administrators and the state health department's central office staff members who seek to link community-level data to Florida's State Health Improvement Plan. Linking EHI scores with life expectancy estimates provides county health departments a more complete picture of neighborhood conditions than any one measure alone. The Nassau County map was distributed as part of a press release by the local health department and received news coverage in which residents shared their thoughts on the map and life expectancy disparities in their community. In the news story, com-

munity members expressed surprise at the disparities, but they also felt that areas with higher economic hardship were indeed areas of poorer health. We hope this publicity will expand beyond Nassau County and generate more interest in addressing health disparities by using locally relevant data and bivariate mapping techniques.

Acknowledgments

Author Information

Corresponding Author: Jessica Joiner, MPH, Public Health Research Unit, Division of Community Health Promotion, Florida Department of Health, 4052 Bald Cypress Way Bin #A24, Tallahassee, FL 32399-1712. Telephone: 850-558-9662. Email: jessica.joiner@flhealth.gov.

Author Affiliations: [1]Florida Department of Health, Public Health Research Unit, Division of Community Health Promotion, Tallahassee, Florida. [2]University of Tennessee, Department of Public Health, Knoxville, Tennessee.

References

1. Chicago Department of Public Health, Epidemiology and Public Health Informatics. Selected socioeconomic indicators in Chicago, 2008–2012. https://data.cityofchicago.org/Health-Human-Services/Census-Data-Selected-socioeconomic-indicators-in-C/kn9c-c2s2. Accessed November 14, 2018.
2. Boothe VL, Fierro LA, Laurent A, Shih M. Sub-county life expectancy: a tool to improve community health and advance health equity. Prev Chronic Dis 2018;15:E11.
3. Montiel LM, Nathan RP, Wright DJ. An update on urban hardship. Albany (NY): The Nelson A. Rockefeller Institute of Government; 2004.
4. Adler NE, Stewart J, Cohen S, Cullen M, Diez Roux AV, Dow W, et al.Reaching for a healthier life: facts on socioeconomic status and health in the U.S. San Francisco (CA): The John D. and Catherine T. MacArthur Foundation Research Network on Socioeconomic Status and Health; 2007.
5. Los Angeles County Department of Public Health, Office of Health Assessment and Epidemiology. Life expectancy in Los Angeles County: how long do we live and why? A cities and communities report. Los Angeles (CA); 2010.
6. US Census Bureau. 2010Census summary file 1: table GCT-PH1. https://factfinder.census.gov/faces/nav/jsf/pages/community_facts.xhtml. Accessed November 14, 2018.
7. Chiang CL. A stochastic study of the life table and its applications. II. Sample variance of the observed expectation of life and other biometric functions. Hum Biol 1960;32:221–38.
8. National Center for Health Statistics. Health, United States, 2016: with chartbook on long-term trends in health. Hyattsville (MD); National Center for Health Statistics; 2017.

GIS SNAPSHOTS

Using Local Data on Adults Aged 18 to 64 to Tailor Interventions for Blood Pressure Medication Adherence in Maine
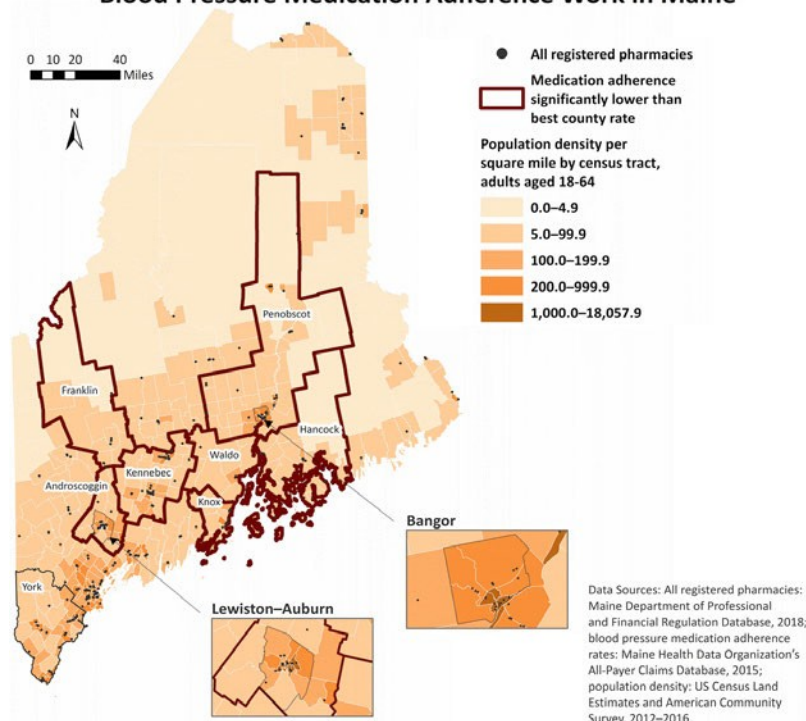
Caitlin Pizzonia, MPH, CPH[1,2]; David Pied, BS[2]; Sara L. Huston, PhD[1,2];

Pamela Foster Albert, MPH[1,2]; Gregory Parent, BS[2]; Nathan Morse, BS[2]

PEER REVIEWED

Maine licensed pharmacy locations, blood pressure medication adherence rates, and population density in 2018. Medication adherence in 2015 among Maine adults aged 18 to 64, calculated for renin–angiotensin system antagonists by using the proportion-of-days-covered method, was 83.8% (95% confidence interval, 83.4%–84.1%). York County had the highest adherence rate (85.2%; 95% confidence interval, 84.3%–86.0%). Counties with medication adherence rates significantly lower than the York County rate indicate where to focus interventions. Adult census tract–level population density for 2012 through 2016 indicates where to implement rural-specific interventions.

Background

Maps showing US state-level data help researchers understand the geographic distribution of chronic disease burden. As public health analysts refine spatial analysis skills, sub-state analyses are sought to determine how to tailor interventions to specific popula- tions so that public health programs can use limited funds most effectively. A workgroup at the Division of Disease Prevention at the Maine Center for Disease Control and Prevention (Maine CDC) created this map to determine where and how to administer public health programs among adults aged 18 to 64 to increase adherence to antihypertensive medication regimens, ultimately influencing hypertension control rates.

In 2015, one in 3 Maine adults (33.4%) had diagnosed hypertension (1), and in 2015–2016 only half of Americans with diagnosed hypertension had controlled hypertension (2). Adherence to antihypertensive medication is associated with controlled hypertension and reduced risk of cardiovascular events (2). US costs for hypertension without heart disease, including health care services, medications, and missed work, totaled $55.9 billion in 2014–2015 dollars (3). Recent evidence also shows that adults aged 35 to 64 are less likely than adults aged 65 or older to take blood pressure medication and have controlled hypertension, thereby increasing their risk for heart disease and stroke (4). Reducing hypertension and cardiovascular events are public health priorities and Healthy People 2020 indicators (5). The Million Hearts program, a national initiative to prevent 1 million heart attacks and strokes, promotes efforts to control hypertension through increasing medication adherence and self-measured blood pressure monitoring (6).

Because half of Maine's land area is uninhabited and 40.8% of the state population lives in rural counties, population density is a critical component in understanding rural health needs. Rural areas tend to have more veterans, older adults (≥65 y), and residents living in poverty than urban areas (7), and rural residents may face unique challenges in controlling hypertension, such as living long distances from pharmacies or physicians.

Data Sources and Map Logistics

Licensed pharmacy locations were received from the 2018 Maine Department of Professional and Financial Regulation Database and geocoded (8). Adult population density, representing adults aged 18 to 64, was calculated from the 2012–2016 US Census population estimates by dividing the total population of adults aged 18 to 64 by the census tract land area (square miles) (9,10). Five manual breaks were used to show variation in adult population density.

Pharmacy claims from the 2015 Maine Health Data Organization's All-Payer Claims Database (APCD), used to calcu- late blood pressure medication adherence rates, represent claims from private and Medicaid beneficiaries (11). Though some private beneficiaries aged 65 to 85 may be Medicare Part D beneficiaries, not all Medicare Part D claims are included in the APCD. Medication adherence was calculated according to Centers for Disease Control and Prevention (CDC) 1305 grant guidance, which uses the proportion-of-days-covered method (having medication for ≥80% of total enrollment days) (12).

Because pharmacy claims could not be linked to medical claims in the Maine APCD, we limited our analysis to renin–angiotensin system antagonists (RASAs), which are used exclusively for hypertension (RASA medications include angiotensin converting enzyme inhibitors, angiotensin II receptor blockers, and direct renin inhibitors and are used only for hypertension, unlike other antihypertensive medications). Total enrollment days were calculated from the patient's prescription start date through December 31, 2015 (12). Antihypertensive medication adherence was calculated among Maine adults aged 18 to 85 who filed 2 or more prescription claims for RASAs and had medication for at least 90 continuous days in 2015. Medication adherence rates were calculated among 2 age groups, 18 to 64 and 65 to 85, and Pearson χ^2 tests ($P < .05$) were used to determine significance between the 2 age groups. County rates among adults aged 18 to 64 were analyzed, and a Pearson χ^2 test ($P < .05$) was used to compare county-specific medication adherence rates to the Maine rate. Then, significance was determined by using a one-way ANOVA and least significant differences test ($P < .05$), comparing the best county rate to all other counties.

SAS version 9.4 (SAS Institute Inc) was used to calculate medication adherence rates and perform statistical analyses. No institutional review board approval was required, but we completed a data use agreement with the Maine Health Data Organization. The map was produced in ArcGIS version 10.6 (Esri).

Highlights

Antihypertensive medication adherence rates increased with age and were significantly lower among adults aged 18 to 64 (83.8%; 95% confidence interval [CI], 83.4%–84.1%) than among adults aged 65 to 85 (86.9%; 95% CI, 86.6%–87.1%). Adherence rates among adults aged 18 to 64 were significantly higher in York County (85.2%; 95% CI, 84.3%–86.0%) than in Maine overall (83.8%; 95% CI, 83.4%–84.1%). Androscoggin, Franklin, Hancock, Kennebec, Knox, Penobscot, and Waldo counties had significantly lower medication adherence rates ($P < .05$) than York County, indicating where to tailor interventions (Table).

Population density and pharmacy locations were concentrated in southern coastal regions but varied substantially within counties with lower medication adherence. Inset maps of Lewiston–Auburn (Androscoggin County) and Bangor (Penobscot County) were included in the map accompanying this article to show greater detail for population-dense towns in counties with lower medication adherence rates. A limitation of the map is that medication adherence rates may be slightly overestimated because of study inclusion criteria, the proportion-of-days-covered method, and restricting analyses to RASAs only. The map does not display medication adherence rates and is best used alongside data tables or an internal interactive ArcGIS online web application that the Division of Disease Prevention workgroup created.

Action

Maine CDC could use data presented in the map to focus future tailored interventions in pharmacies located in counties with significantly lower adherence rates and replicate successful practices in York County pharmacies to improve medication adherence. Maine CDC can help implement self-measured blood pressure monitoring, lifestyle change programs, or telehealth interventions on the basis of local population density. Self-measured blood pressure monitoring and telehealth may be beneficial in areas with low population density because in-person lifestyle change programs may be less effective if patients live far away from a retail pharmacy. Applying all 3 approaches in densely populated areas could improve medication adherence and hypertension control for a high proportion of the state's younger adult population (18–64 y). If the interventions increased medication adherence in identified lower adherence counties, 495 adults aged 18 to 64 taking RASA medications would be adherent to blood pressure medications, increasing their chances of controlled hypertension. This conservative estimate was calculated by multiplying the difference between the county rate and the York County rate (85.2%) by the county-level study populations (Table) and summing the results. The Maine CDC plans to replicate this collaborative map process to strategically inform other chronic disease prevention programs.

Acknowledgments

Author Information

Corresponding Author: Caitlin Pizzonia, MPH, CPH, 286 Water St, Augusta, Maine. Telephone: 207-287-4102. E-mail: caitlin.pizzonia@maine.edu.

Author Affiliations: [1]Cutler Institute for Health and Social Policy, Muskie School of Public Service, University of Southern Maine, Portland, Maine. [2]Chronic Disease Program, Division of Disease Prevention, Maine Center for Disease Control and Prevention, Augusta, Maine.

References

1. Centers for Disease Control and Prevention. Behavioral Risk Factor Surveillance System, 2015. https://www.cdc.gov/brfss/annual_data/annual_2015.html. Accessed July 31, 2018.
2. Fryar CD, Ostchega Y, Hales CM, Zhang G, Kruszon-Moran D. Hypertension prevalence and control among adults, United States, 2015–2016. NCHS Data Brief 2017;(289):1–8.
3. Benjamin EJ, Muntner P, Alonso A, Bittencourt MS, Callaway CW, Carson AP, et al. American Heart Association Council on Epidemiology and Prevention Statistics Committee and Stroke Statistics Subcommittee. Heart disease and stroke statistics – 2019 update: a report from the American Heart Association. Circulation 2019;139(10):e56–528.
4. Wright JS, Wall HK, Ritchey MD. Million Hearts 2022: small steps are needed for cardiovascular disease prevention. JAMA 2018;320(18):1857–8.
5. US Department of Health and Human Services, Office of Disease Prevention and Health Promotion. Healthy people 2020: clinical preventive services. https://www.healthypeople.gov/2020/leading-health-indicators/2020-lhi-topics/Clinical-Preventive-Services. Accessed August 7, 2018.
6. Benjamin RM. The Million Hearts initiative: progress in preventing heart attacks and strokes. Public Health Rep 2012;127(6):558–60.
7. Philips BU Jr. 2017 Rural Health Report Guard. Rural Health Quarterly Fall 2017. http://ruralhealthquarterly.com/home/2017/12/15/u-s-rural-health-report-card-2017/. Accessed August 7, 2018.
8. Maine Department of Professional and Financial Regulation. Office of Professional and Occupational Regulation database. http://www.maine.gov/pfr/professionallicensing/. Accessed February 15, 2018.

9. US Census Bureau. 2012–2016 American Community Survey 5-year estimates. https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid = ACS_16_5YR_S0101&prodType = table. Accessed February 24, 2019.

10. US Census Bureau. 2010 Census population, housing units, area, and density. https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src = bkmk. Accessed February 24, 2019.

11. Maine Health Data Organization. All-payer claims database. https://mhdo.maine.gov/claims.htm. Accessed July 14, 2016.

12. Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Training and Technical Assistance Support Center. Calculating proportion of days covered for antihypertensive and antidiabetic medications: an evaluation guide for grantees. https://www.cdc.gov/dhdsp/docs/med-adherence-evaluation-tool.pdf. Accessed August 7, 2018.

Table

Table. Medication Adherence Among Adults Prescribed RASA Medications, by Demographic and Geographic Characteristics, Maine, 2015[a]

Characteristic	No. (%)[b]	% Adherent[c] (95% CI)	P Value[d]
Total	105,663 (100)	85.5 (85.3–85.7)	NA
Age, y			
18–64	45,544 (43.1)	83.8 (83.4–84.1)	<.001
65–85	60,119 (56.9)	86.9 (86.6–87.1)	
County, adults aged 18–64 y			
Androscoggin	3,944 (8.7)	82.8 (81.6–84.0)	
Franklin	849 (1.9)	80.9 (78.3–83.6)	
Hancock	1,581 (3.5)	82.2 (80.3–84.1)	
Kennebec	4,366 (9.6)	82.8 (81.7–83.9)	<.05[e]
Knox	1,142 (2.5)	81.6 (79.4–83.9)	
Penobscot	5,829 (12.8)	83.4 (82.5–84.4)	
Waldo	1,383 (3.0)	80.2 (78.1–82.3)	
York	6,573 (14.4)	85.2 (84.3–86.0)	—[f]

Abbreviations: CI, confidence interval; NA, not applicable; RASA, renin–angiotensin system antagonists.

[a] From Maine Health Data Organization, All Payer Claims Database (APCD) (11).

[b] Maine adults aged 18–85 recorded in APCD pharmacy claims who filled at least 2 prescriptions for RASA (renin–angiotensin system antagonists) medications in 2015 that totaled at least 90 continuous days' supply, with the first prescription filled on or before September 30, 2015. Study population numbers may not sum to group totals because of missing information on that demographic or geographic characteristic, and study population percentages may exceed 100% because of rounding.

[c] Percentage of Maine adults aged 18 to 85 adherent to medication regimens (had medication for 80% of days from the patient's prescription start date until the end of the calendar year) based on 2015 Maine Department of Health APCD pharmacy claims.

[d] P values were calculated based on Pearson χ^2 test at α = 0.05 between demographic groups.

[e] P values calculated by using the ANOVA test at α = 0.05 comparing York County to all Maine counties. Only significant counties were presented in this table.

[f] No P value was presented for York County because it is the county comparison rate.
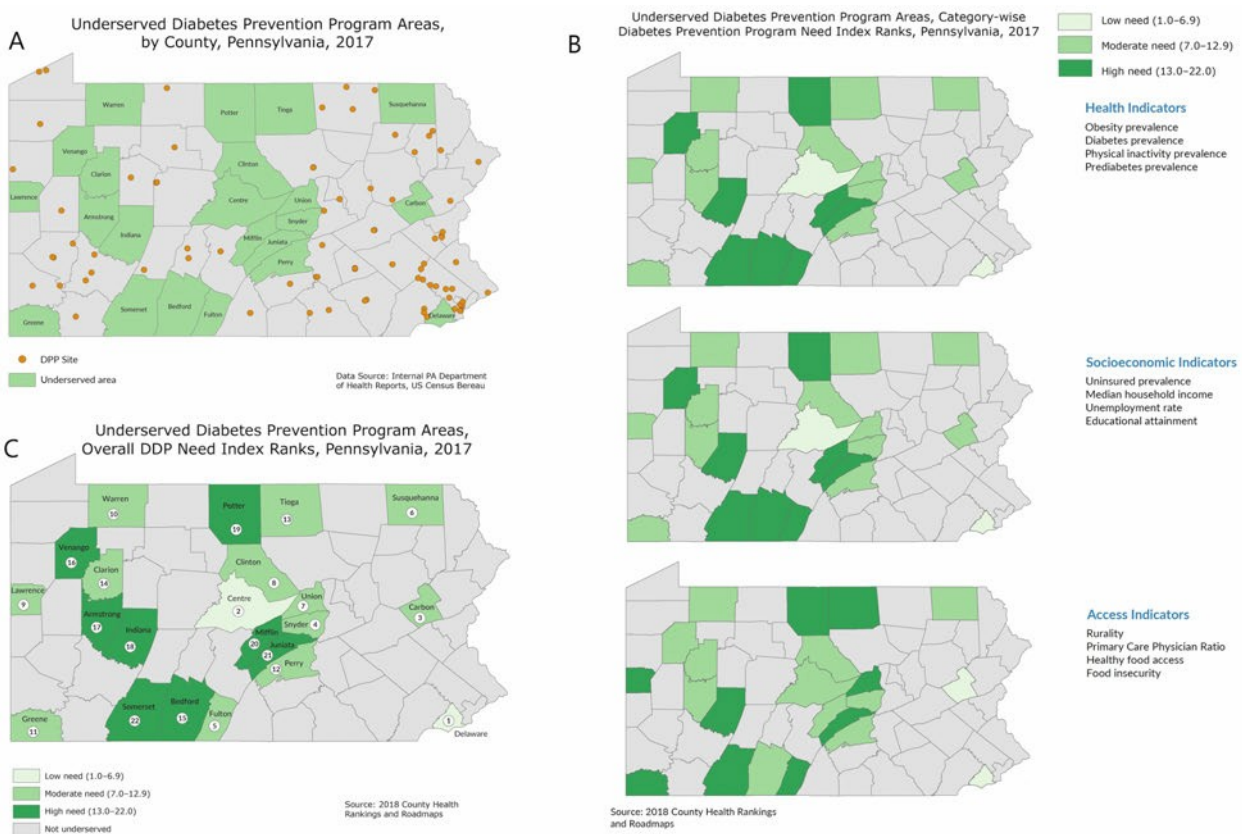
GIS SNAPSHOTS

Using Geographic Information Systems to Highlight Diabetes Prevention Program Expansion Areas in Pennsylvania

Brian Zepka, MPH[1]; Mohammad Anis, MSSP[1]; Jennifer D. Keith, MPH, CPH[1]; Duane Barksdale[1]; Camelia Rivera[2]

PEER REVIEWED

Map A shows underserved DPP areas, counties that do not have any CDC-recognized DPPs and have a population of 10,000 or more, in Pennsylvania. Map B shows the CDNIRs for each underserved county within Pennsylvania in 3 risk factor categories; health, socioeconomic, and access indicators. Numbers indicate ranking on 3 hierarchical tiers according to need for DPP: low (range, 1.0–6.9), moderate (range, 7.0–12.9), and high (range, 13.0–22.0) CDNIRs are an average of the county ranks for each indicator in the 3 categories. Map C shows the ODNIRs for the 22 underserved areas. ODNIR is a weighted average of 3 CDNIRs: health, socioeconomic, and access indicators. Abbreviations: CDC, Centers for Disease Control and Prevention; CDNIRs, Category-wise DPP Need Index Ranks; DPP, Diabetes Prevention Program; ONDIRs, Overall DPP Need Index Ranks; PADOH, Pennsylvania Department of Health.

Background

The Diabetes Prevention Program (DPP) is a lifestyle change program recognized by the Centers for Disease Control and Prevention that is intended to prevent patients diagnosed with prediabetes from developing type 2 diabetes. The Public Health Management Corporation's Research and Evaluation Group (R&E Group), the external evaluation partner for the Pennsylvania Department of Health's (PADOH's) DPP initiative, conducted an analysis to identify counties with no in-county access or limited access to sites offering DPP classes (underserved) and their relative need. R&E Group identified 22 underserved counties in Pennsylvania, a state in which diabetes is a leading cause of death. Thus, increasing access to evidence-based type 2 diabetes interventions is a priority for PADOH. To effectively prioritize DPP expansion efforts, it is important to examine resource allocation and program accessibility across the Commonwealth.

Methods

R&E Group produced an index to rank underserved counties on the basis of need to identify which of the 22 would benefit most from a new DPP site. This index is based on risk factors for developing type 2 diabetes and factors that influence the ability of populations to access health services. R&E Group identified 12 metrics and grouped them into 3 categories.

Indicators

Because DPP eligibility is based on a diagnosis of prediabetes, county-level rates of diabetes and prediabetes are included as indicators of need for DPP. Prevalence of conditions considered to be risk factors for developing type 2 diabetes, including obesity and low rates of physical activity, were also included in this analysis (1).

Extensive literature indicates that low socioeconomic status is a risk factor for developing type 2 diabetes. Two socioeconomic indicators, low household income and not having a college degree, are associated with high prevalence of type 2 diabetes (2) and were included in the index. Unemployment and not having health insurance were also included, because they are typically associated with low socioeconomic status (3).

The ability for target populations to access DPP classes influences the viability of DPP sites. This index accounts for the percentage of residents living in rural areas, because these areas often experience lower access to health services than nonrural areas (4). Food insecurity and the percentage of low income populations that do not live near a grocery store were also included, because lack of consistent access to healthy food is a risk factor for developing

type 2 diabetes (5). Finally, each county's ratio of population to primary care physicians was included in this analysis to identify counties with low capacity for delivering primary care. Access to primary care is crucial to preventing chronic diseases, including type 2 diabetes (6).

County DPP Need Index development

R&E Group used County Health Rankings data for the 22 underserved counties in each of the 12 indicators and used them to reassign each county a diabetes risk rank among the 22 counties. The new ranks were assigned based on a subset of the county's original County Health Rankings data. The new diabetes risk-focused ranking ranged from 1 for healthiest to 22 for unhealthiest.

On the basis of this framework, a DPP Need Index (DNI) was developed by using the counties' revised positions to determine risk for being a DPP underserved area. Counties located lower on the DNI were those that had lower levels of county-wide risks. Based on DNI rank, counties were divided into 3 hierarchical tiers according to need for DPP: low (range, 1.00–6.99), moderate (range, 7.00–12.99), and high (range, 13.00–22.00).

This methodology was applied to formulate preliminary category-wise and overall county DNI ranks. The Category-wise DPP Need Index rank (CDNIR) for a county was its average rank across all 4 metrics that make up the indicator group. The Overall DPP Need Index rank (ODNIR) was based on a weighted average ranking of the county across all 3 categories. Simple weights in each category were assigned on the basis of their direct relevance to prediabetes. Health category was assigned the highest weight of 50% in the calculation of the ODNIR on the basis of the 4 indicators (obesity, prediabetes, diabetes, and lack of physical activity) that directly affect diabetes risk in the county. High correlation among these measures indicates greater need for an intervention. Each of 2 remaining socioeconomic and access indicator categories were assigned weights of 25% in the final rank calculation as indirect indicators of diabetes risk. For example, Armstrong County, which had health, socioeconomic, and access CDNIRs of 17.00 (high), 11.25 (moderate), and 10.50 (moderate), respectively, had an unweighted ODNIR of 12.92 (moderate). However, once weights were assigned to the individual CDNIRs, the ODNIR for Armstrong County fell to 13.94 (high), because the county had a higher risk in the health category.

Data for producing this index were provided by County Health Rankings and Roadmaps (7). World Geocoder for ArcGIS (Esri) was used to locate DPP sites. Potential DPP sites were identified by using 2017 address data. Maps were produced in ArcMap version 10.4.

Main Findings

CDNIRs and ODNIRs were used to determine counties with the highest need for DPP. Three underserved counties, Juniata, Potter, and Somerset, had high CDNIRs across all 3 indicators. They also comprised 3 of the 4 counties with the highest ODNIRs. Five other counties, Bedford, Venango, Armstrong, Indiana, and Mifflin, showed a high need for DPP on the basis of their ODNIRs. Among the original 22 counties identified as having limited ac- cess to DPP, 8 were identified as having high need for DPP infrastructure.

Action

This series of maps highlights counties where PADOH can direct its DPP expansion efforts. Geographic visualization of DPP sites allows regional implementation partners to prioritize areas with limited program access and is a tool to engage partners in seeking expansion sites that can serve populations at high risk for developing type 2 diabetes. Our analysis focused on county data to maximize publically available data, support PADOH discussion, and align with DPP funding channels. Access across county lines is also important to explore in the future, because county boundaries are often an artificial barrier and within-county disparities in access may be missed.

Acknowledgments

Author Information

Corresponding Author: Brian Zepka, MPH, Public Health Management Corporation, Research and Evaluation Group, Centre Square East, 1500 Market St, Philadelphia, PA 19102. Telephone: 215-985-2500. Email: bzepka@phmc.org.

Author Affiliations: [1]Public Health Management Corporation, Research and Evaluation Group, Philadelphia, Pennsylvania. [2]Pennsylvania Department of Health, Bureau of Health Promotion and Risk Reduction, Harrisburg, Pennsylvania.

References

1. National Institute of Diabetes and Digestive and Kidney Diseases. Risk factors for type 2 diabetes. https://www.niddk.nih.gov/health-information/diabetes/overview/risk-factors-type-2-diabetes. Accessed August 22, 2018.
2. Saydah S, Lochner K. Socioeconomic status and risk of diabetes-related mortality in the U.S. Public Health Rep 2010; 125(3):377–88.
3. Winkleby MA, Jatulis DE, Frank E, Fortmann SP. Socioeconomic status and health: how education, income, and occupation contribute to risk factors for cardiovascular disease. Am J Public Health 1992;82(6):816–20.
4. Weisgrau S. Issues in rural health: access, hospitals, and reform. Health Care Financ Rev 1995;17(1):1–14.
5. Tait CA, L'Abbé MR, Smith PM, Rosella LC. The association between food insecurity and incident type 2 diabetes in Canada: a population-based cohort study. PLoS One 2018; 13(5):e0195962.
6. Saudek CD. The role of primary care professionals in managing diabetes. Clin Diabetes 2002;20(2):65–6.
7. County Health Rankings and Roadmaps. 2018. http://www.countyhealthrankings.org/app/pennsylvania/2018/downloads. Accessed August 22, 2018.
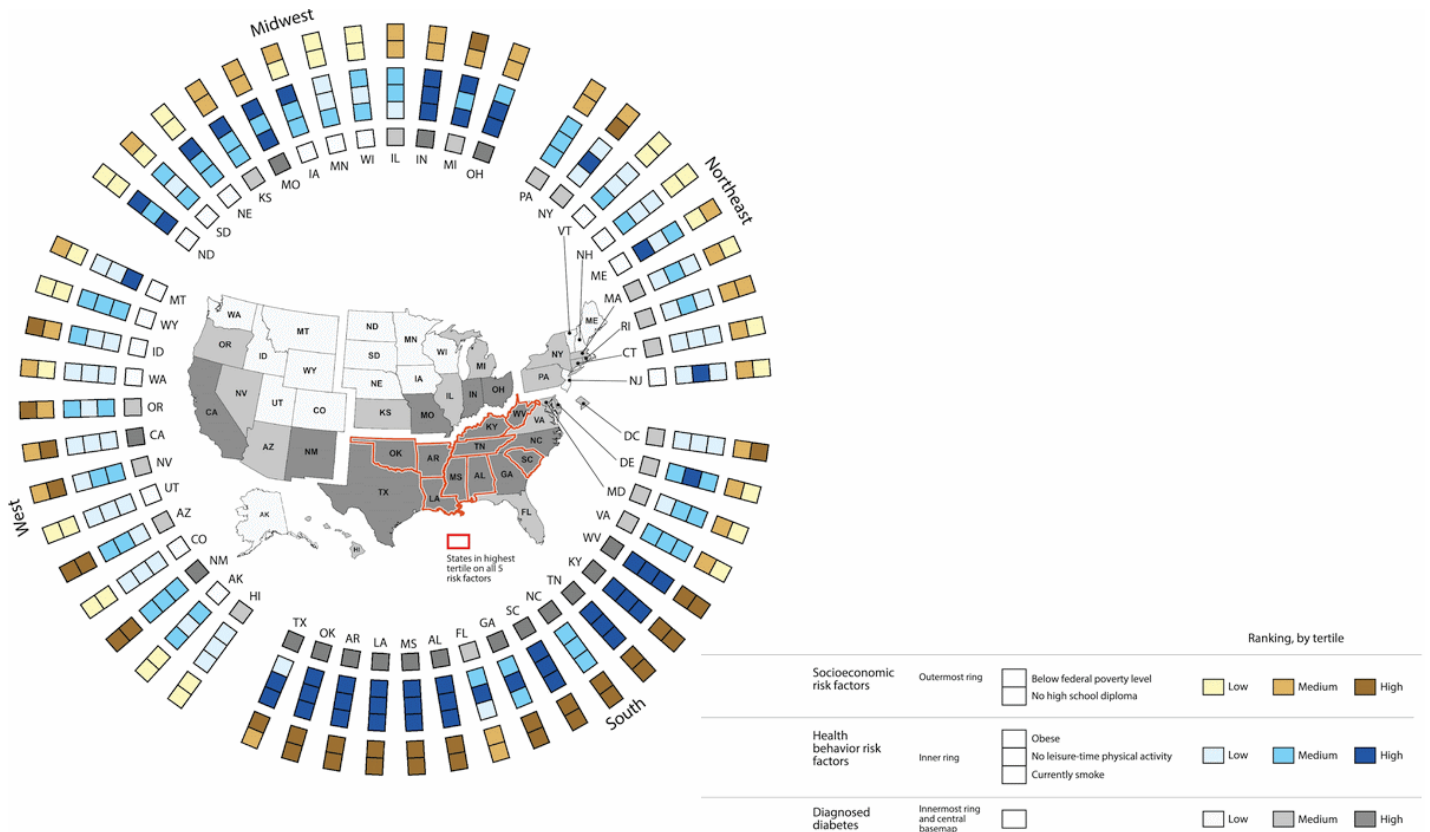
GIS SNAPSHOTS

Diagnosed Diabetes Prevalence and Risk Factor Rankings, by State, 2014–2016: A Ring Map Visualization

Ana Lòpez-DeFede, PhD[1]; John E. Stewart, PhD[1]

PEER REVIEWED

The ring map shows that states with a higher prevalence of risk factors generally have a higher prevalence of diagnosed diabetes. The 9 states in the highest tertile for all 5 risk factors also are in the highest tertile for diabetes prevalence. By integrating multiple spatial data elements in a single graphic, the ring map highlights possible state-level associations between diagnosed diabetes prevalence, socioeconomic disadvantage, and health behaviors. All mapped values represent data for adults aged ≥18, except the percentage with no high school diploma, which was measured for adults aged ≥25.

Background

In the United States, diabetes is a leading cause of adult-onset blindness, kidney failure, and death (1). Efforts to prevent and control diabetes must consider geographic variation in disease prevalence and risk factors such as obesity, sedentary lifestyle, and low educational attainment (2). Maps are essential to our understanding of geographic differences in population health and disease vulnerability. Comparing geographic patterns of disease and population risk across multiple maps, however, can be cumbersome. Ring mapping is an innovative geovisualization method that permits the display of multiple spatially referenced variables on a single map (3). We used a ring map to depict the prevalence of diagnosed diabetes and 5 associated risk factors (living below the federal poverty level, low educational attainment, obesity, no leisure-time physical activity, and current smoking) for adults in all 50 US states and the District of Columbia.

Data Sources and Map Logistics

We obtained data on the age-adjusted prevalence of diagnosed diabetes, obesity, physical activity, and current smoking among adults aged 18 or older from the Behavioral Risk Factor Surveillance System (4). For these measures, we calculated mean age-adjusted prevalence on the basis of the most recent 3 years of data available (2014–2016). We obtained data on poverty (percentage of adults aged ≥18 below the federal poverty level) and educational attainment (percentage of adults ≥25 with no high school diploma) from the US Census Bureau, American Community Survey, 2015 1-Year Estimates (5).

We constructed a ring map with 2 principal parts: a ring display and a central basemap. The ring display consists of 6 concentric rings, each comprising 51 symbolization units, 1 unit in each ring for each state and the District of Columbia. The 2 outermost rings represent the 2 socioeconomic risk factors; the 3 inner rings, the 3 health behavior risk factors; and the single innermost ring, the prevalence of diagnosed diabetes. The central basemap shows the geographic pattern of diagnosed diabetes prevalence across states; the shade used to depict the prevalence of diagnosed diabetes in each state on the basemap is the same shade used in the innermost ring. Diabetes and risk factor data are symbolized by using a tertile ranking scheme, with approximately equal numbers of observations in low, medium, and high classes. Tertiles were based on the distribution of values for all 50 states and the District of Columbia (Table). Intentional gaps in the rings and basemap indicate the 4 US Census regions, facilitating exploration of potential regional differences in diabetes prevalence and population risk.

A state-specific example (Montana) illustrates how to interpret the ring map. The ring display shows 6 symbolization units for Montana. Reading from the outermost rings to innermost ring, we see that Montana has a medium prevalence of poverty, a low prevalence of no high school diploma, a low prevalence of obesity, a low prevalence of no leisure-time physical activity, a high prevalence of current smoking, and low prevalence of diagnosed diabetes. The basemap shows the location of Montana and its low prevalence of diagnosed diabetes in relation to the rest of the United States.

The US basemap was created in ArcMap version 10.4 (Esri). A JavaScript was developed to draw the ring elements in Adobe Illustrator (Adobe, Inc). We assembled the basemap and rings and added text and legend elements in Adobe Illustrator.

Highlights

The ring map shows generally a higher prevalence of diagnosed diabetes in the South. This finding is consistent with the findings of previous research, which identified a "diabetes belt" of counties located predominantly in the South census region (2). The prevalence of socioeconomic and health behavior risk factors is also higher overall in the South. The 9 states in the highest tertile for all 5 risk factors (all located in the South) are also in the highest tertile for diagnosed diabetes. Conversely, of the 3 states in the lowest tertile for all 5 risk factors (all located in the West), 2 states (Colorado and Utah) are in the lowest tertile for diagnosed diabetes and 1 state (Hawaii) is in the medium class.

Some clear exceptions to the general spatial correspondence of diagnosed diabetes prevalence and population risk merit examination. In the midwestern states of Iowa, Nebraska, and North Dakota, for example, obesity prevalence is high, but diabetes prevalence is low. On the other hand, California, in the West, has a low prevalence of obesity, a low prevalence of no leisure-time physical activity, and a low prevalence of smoking but a high prevalence of diagnosed diabetes. Thus, although the ring map highlights possible associations between diagnosed diabetes prevalence, socioeconomic disadvantage, and health behaviors at the state level, it also suggests potential regional differences in risk (6).

This ring map has several limitations. The geovisualization does not indicate the significance of potential associations between the selected risk factors and diabetes prevalence, nor does it convey statistical information about spatial autocorrelation of risk factors and diabetes. Based on state-level data, the ring map does not permit visual assessment of small-area geographic variation in diabetes and population risk within states. Finally, graphic space and

legibility constraints limit the number of rings displayed and thus the number of potential risk factors mapped.

Action

This novel geovisualization can help raise public awareness about spatial variability in diabetes prevalence and vulnerability. The striking visual association between the prevalence of diagnosed diabetes and population risk, especially in the South, can inform and motivate state initiatives to address such modifiable risk factors as poverty, obesity, sedentary lifestyle, and smoking. The ring map also might encourage further exploration of additional area-level factors that alone, or in combination, influence diabetes morbidity and mortality, including, racial/ethnic composition (1,7), characteristics of the built environment (3), and state decisions to expand Medicaid (8).

Acknowledgments

Author Information

Corresponding Author: Ana Lòpez-DeFede, PhD, Institute for Families in Society, University of South Carolina, 1600 Hampton St, Ste 507, Columbia, SC 29208. Telephone: 803-777-5789. Email: adefede@mpr.sc.edu.

Author Affiliations: [1]Institute for Families in Society, University of South Carolina, Columbia, South Carolina.

References

1. Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion. Diabetes at a glance 2016: working to reverse the US epidemic; 2018. https://www.cdc.gov/chronicdisease/resources/publications/aag/pdf/2016/diabetes-aag.pdf. Accessed August 2, 2018.
2. Barker LE, Kirtland KA, Gregg EW, Geiss LS, Thompson TJ. Geographic distribution of diagnosed diabetes in the U.S.: a diabetes belt. Am J Prev Med 2011;40(4):434–9.
3. Stewart JE, Battersby SE, Lopez-De Fede A, Remington KC, Hardin JW, Mayfield-Smith K. Diabetes and the socioeconomic and built environment: geovisualization of disease prevalence and potential contextual associations using ring maps. Int J Health Geogr 2011;10(1):18.
4. Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion. Chronic disease indicators (2014–2016). https://www.cdc.gov/cdi/index.html. Accessed July 3, 2018.
5. US Census Bureau. American Community Survey (20151-year estimates). https://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml?refresh=t. Accessed July 3, 2018.
6. Myers CA, Slack T, Broyles ST, Heymsfield SB, Church TS, Martin CK. Diabetes prevalence is associated with different community factors in the diabetes belt versus the rest of the United States. Obesity (Silver Spring) 2017;25(2):452–9.
7. Gaskin DJ, Thorpe RJ Jr, McGinty EE, Bower K, Rohde C, Young JH, et al. Disparities in diabetes: the nexus of race, poverty, and place. Am J Public Health 2014; 104(11):2147–55.
8. Kaufman HW, Chen Z, Fonseca VA, McPhaul MJ. Surge in newly identified diabetes among Medicaid patients in 2014 within Medicaid expansion states under the Affordable Care Act. Diabetes Care 2015;38(5):833–7.

Table

Table. Ranges for Low, Medium, and High Tertiles for Prevalence of Diagnosed Diabetes and Selected Associated Risk Factors, Based on Distribution of Values Among Adults Aged ≥18 in 50 States and the District of Columbia[a]

Measure	Low	Medium	High
Diagnosed diabetes, %	6.5–8.2	8.3–9.7	9.8–12.8
Socioeconomic risk factors			
Below federal poverty level, %	7.6–11.1	11.2–13.8	13.9–18.9
No high school diploma, %[b]	6.4–9.2	9.3–12.4	12.5–17.8
Health behavior risk factors			
Obese, %	21.2–27.3	27.4–30.9	31.0–36.8
No leisure-time physical activity, %	16.6–21.3	21.4–24.9	25.0–32.4
Currently smoke, %	9.1–16.0	16.1–19.7	19.8–27.3

[a] Data sources: Behavioral Risk Factor Surveillance System (4), US Census Bureau, American Community Survey, 2015 1-Year Estimates (5).
[b] Among adults aged ≥25.

TOOLS FOR PUBLIC HEALTH PRACTICE

An Online Geographic Data Visualization Tool to Relate Preterm Births to Environmental Factors

Marta M. Jankowska, PhD[1,2]; Jiue-An Yang, PhD[1]; Jessica Block, MS[1]; Rebecca J. Baer, MPH[2,3];

Laura L. Jelliffe-Pawlowski, PhD[2,4]; Sandra Flores[5]; Tania Pacheco-Warner, PhD[6]; Amber Costantino, MA[5];

Jonathan Fuchs, MD, MPH[2,7]; Christina D. Chambers, PhD[3]; Gail Newel, MD, MPH[5,8]

PEER REVIEWED

Summary

What is already known about this topic?

Preterm birth is a complex health problem with numerous risk factors. Data visualization and mapping of preterm birth and related data are valuable methods of exploring data and engaging the public and stakeholders.

What is added by this report?

This project details the process of designing, gathering user feedback, and implementing an online and open source data visualization and exploration tool for preterm birth and related data in Fresno County, California.

What are the implications for public health practice?

By giving researchers, stakeholders, and the public free and open source data exploration tools, more informed discussions for reducing preterm birth can occur, and new avenues of research can be explored.

Abstract

Preterm birth (<37 weeks gestation) continues to be a significant cause of disease and death in the United States. Its complex causes are associated with several genetic, biological, environmental, and sociodemographic factors. Organizing and visualizing various data that may be related to preterm birth is an essential step for pattern exploration and hypothesis generation and presents an opportunity to increase public and stakeholder involvement. In this article, we describe a collaborative effort to create an online geographic data visualization tool using open software to explore preterm birth in Fresno County, where rates are the highest in California. The tool incorporates information on births, environmental expos-ures, sociodemographic characteristics, the built environment, and access to care. We describe data sets used to build the tool, the data-hosting platform, and the process used to engage stakeholders in its creation. We highlight an important example of how collaboration can increase the utility of geographic data visualization to improve public health and address health equity in birth outcomes.

Preterm Birth and the Need for Data Visualization Tools

Preterm birth (<37 weeks of gestation) contributes significantly to disease and death in the United States, both in the short term and long term. It is associated with higher death rates through infancy and childhood, decreased reproduction, increased risk of having preterm offspring, increased risk of high blood pressure, and symptoms of metabolic syndrome (1–5). The causes of preterm birth are complex and vary for early gestation (20–31 weeks) and late gestation (32–36 weeks) as well as spontaneous (eg, sudden or unplanned preterm birth) and medically indicated (eg, planned and induced preterm birth to minimize other health risks of the baby or mother) subtypes (6). Understanding the causes of preterm birth is vital to informing overarching risk reduction strategies and to developing early detection methods and interventions and can lead to new discoveries in subtype and population-specific risks. However, exploring the myriad of risks for preterm birth — from a woman's health history, to biomarker data, to behaviors — is challenging for researchers, clinicians, and community health organizations seeking to understand preterm birth and work with women to reduce their risks. These challenges increase as the importance of environment and context become increasingly relevant in preterm birth research and clinical care. Factors such as air pollution, neighborhood environment, and socioeconomic status introduce new data and analytic challenges derived from geographic data formats, which must be integrated with traditional clinical data (7–9). Furthermore, new sources of open-source data are becoming increasingly available, leading to new research and

data integration opportunities for better understanding preterm birth (10).

At a population level, health data can be linked geographically through address, census tract, and zip code information. Exploratory spatial data analysis (ESDA) uses geographic linkages to explore patterns, compare nearby geographic regions, and analyze spatial clusters (11), which can indicate underlying place-based variables important to a health outcome, such as crime, pollution, or unexplored environmental factors (12). Visualization of such data is essential for pattern exploration and hypothesis generation. Data visualization offers a field of research and developed tools for exploring patterns, identifying relationships, and synthesizing information in large, multiscale, and multivariate data sets (13). Being able to explore and visualize multilevel and multifactor risks for preterm birth may lead to new mechanistic hypotheses (14), and allow researchers, clinicians, and community health organizations to work with patients in the context of population-level patterns (15). For such a visualization tool to be useful it must be easy to use and immediately accessible, preferably through an online platform; it must leverage ESDA and geographic data science exploratory tools (16) and must have an integrated data structure that includes medical, behavioral, social, and environmental factors.

In this article we discuss the collaborative efforts of several organizations in Fresno County and the state of California to create an online data visualization and exploration tool by using open software (https://delphidata.ucsd.edu/ptbi) to describe preterm birth in Fresno County. We describe data sources and data collection, data features, and mapping functions of the tool. We highlight the utility of online geographic data visualization to explore possible causes of preterm birth and interventions to address it. We also detail how such tools can be built in collaboration with on-the-ground organizations and stakeholders.

Setting and Partners

Fresno County, California, exemplifies the many challenges presented by the complex and multiple-pathway mechanisms of preterm birth. According to vital statistics for 2007 through 2012 of the California Department of Public Health, Fresno County had the highest overall preterm birth rate in the state, 9.9%, representing 3.7% of California's preterm births. At a finer geographic scale, 41.4% of all preterm births in that time frame occurred in the south and west-central areas of the city of Fresno. These are the most populated areas of the county, and more than 70% of pregnant women residing there receive Medi-Cal health insurance for prenatal care or delivery on the basis of low-income status. For these reasons, Fresno County is one location of focus for the Uni-

versity of California–San Francisco (UCSF) California Preterm Birth Initiative, a multiyear interdisciplinary research effort with the goal of reducing the prevalence of preterm birth. The Fresno County part of the initiative is a Collective Impact effort that brings together strategic partners from different sectors to focus on the prevention of preterm birth and to address racial and ethnic disparities in its prevalence. Members include local leaders representing public institutions (Fresno County Office of Education, Fresno Housing Authority, Fresno Police Department), public health (Fresno County Department of Public Health; Special Supplemental Nutrition Program for Women, Infants, and Children; First 5 of Fresno), health systems and hospitals (Cal Viva Health, Valley Children's Healthcare, Community Medical), higher education (Fresno State University, UCSF–Fresno), health clinics (eg, Clinica Sierra Vista), community benefit organizations (eg, AMOR Foundation, Every Neighborhood Partnership), and mothers who experienced preterm birth. Collectively, these groups have committed to address and reduce preterm birth.

A research team from the University of California–San Diego and UCSF composed of geographers, computer scientists, research experts on preterm birth, and a practicing obstetrician was formed in 2016. The team partnered with the Collective Impact effort in Fresno County to begin planning an online geographic data visualization tool that could aid in the assessment, exploration, and discovery of patterns in preterm birth and other social, environmental, and hazard factors. The team met several times with members of the Fresno County Preterm Birth Initiative to obtain feedback about data to include in the tool and the tool's usability and design.

Data Inputs and Collection

Finding, formatting, and amalgamating data inputs is one of the largest tasks of any data visualization project. A core goal of this project was to collect data resources that may not be traditionally associated with preterm birth to aid in research discovery and relationship exploration. As a start to the project, various preterm birth researchers were invited to brainstorm variables and data of interest ranging from birth outcomes to environmental factors to supportive resources. From those lists, the core team worked to collect as many data sources as possible (Table). Data were formatted at 2 geographic levels: census tracts, when possible, and Medical Service Study Areas (MSSAs). MSSAs are geographical analysis units defined by the California Office of Statewide Health Planning and Development (OSHPD) and are based on census tracts. Numerous health-related state data sets are provided only at the MSSA unit. Three main types of software were used to amalgamate and format data: ArcMap, version 10.5 (Esri), Microsoft Excel (Microsoft Corp), and SPSS (IBM Corp). All data were

manually entered into master CSV (comma-separated values) spreadsheets by geographic unit with data dictionaries. The process included manual data curation decisions, for example, deciding what variables to keep or discard because of redundancy, or methods of aggregation such as counts versus averages.

Data about births were obtained from a birth cohort database maintained by OSHPD. These data are shared through data use agreements and were obtained through collaboration with the California Preterm Birth Initiative under an institutional review board–approved protocol from the California Committee for the Protection of Human Subjects, protocol no. 12–09-0702, ongoing since 2009. The birth cohort file contains detailed information on maternal and infant characteristics derived from linked hospital discharge, birth certificate, and infant death records. Included in the file were all singleton births in Fresno County from 2007 through 2012 with an obstetrician's best estimate of a gestation at delivery of 20 to 44 weeks and with no known chromosomal abnormalities or major structural birth defects (17). Births were categorized by weeks of gestation based on best obstetric estimates (early preterm birth, gestational age <32 wk; late preterm birth, gestational age 32 to ≤36 wk; term birth, gestational age ≥37 wk) and spontaneous and medically indicated subtypes of birth. Additional data derived from the linked birth certificate and hospital discharge data were race/ethnicity, age, parity, country of birth, pre-pregnancy weight, height, insurance status, smoking, diabetes, hypertension, anemia, maternal education, reported drug use, diagnosed infection, mental illness, and for multiparous women, known number of previous caesarean sections and interpregnancy interval. Data files provided diagnoses and procedure codes based on the International Classification of Diseases, 9th Revision, Clinical Modification (18). The final data set included 81,021 women. Data were aggregated to MSSAs and census tracts with a minimum of 16 women per geographic unit to preserve privacy. In addition to the OHSPD data set of women giving birth, several other indicators were collected from a variety of data sources including the US Census, the Environmental Protection Agency, California Health and Human Services, Fresno County, Esri (Environmental Systems Research Institute), Google, and other California State agencies (Table). Almost all data sets included in the tool are open source and available online, except for data about crime. Not all the indicators align in date with the birth data (2007–2012) because of availability of online data sets. This is a limitation of the tool, which will be improved when newer birth data can be incorporated and as more data are made available online to include past years. Data ingestion is ongoing, and new data sets will be added as they are made available and processed.

Online Infrastructure and Key Features

The online platform has 2 sides with a user-friendly front end featuring simple click and view options for preterm birth–related topics, and a password protected back end for more advanced users with more data and complex visualization options. Software to develop the tool are all open source framework environments and libraries built with shareability and reproducibility in mind. Using the open source frameworks can allow the tool to be repurposed not only for health but also for other data applications. Code developed for the front-end infrastructures is available online through our GitHub repository (https://github.com/hdscalecollab-ucsd/PTBi-Viz), with back-end code to be added in the near future. We encourage users to use these codes for their own health data visualization projects and to contribute new visualization features back to the repository as they are developed. The front-end development was a direct result of input from stakeholders who, early in the process, voiced a need for a user-friendly and guided data experience for people unfamiliar with data exploration and visualization techniques. Thus, the 2 applications were designed differently to provide their targeted audiences with the data visualization and analytic tools that support knowledge discovery for preterm birth.

The topic-driven front end is intended to create a guided experience to browse data by predefined topics (eg, demographics, environmental pollution) with curated data selected by the team. Each topic has no more than 15 variables per screen, thus limiting the scope of data exploration and making it more manageable for lay users. The data-driven back end application allows users to examine preterm birth–related variables by region and by their relationships with other variables with no limit to the quantity or type of variables included in an analysis. For example, a researcher can explore associations between very preterm birth (between 28 and 32 wk), environmental pollution, socioeconomic disadvantage, and access to care simultaneously.

When logging onto the public site for the first time (https://delphidata.ucsd.edu/ptbi), the user is welcomed with a video briefly explaining the California Preterm Birth Initiative. The 8 topical areas (birth data, health care, demographics, environmental pollution, socioeconomics, pregnancy-related factors, built environment, and health risk) encourage users to begin their exploration of preterm birth and selected geographically organized data. These topics are built with a module tab design pattern where content for each topic is constructed in a separate tab panel and only 1 topic is viewable at a time. When a topic is clicked, the first window to appear in the new page is information about the sources of data being displayed. Within each topic tab, data visualization components are laid out in a side-by-side grid system. These data

visualization components are modularized map or graph items that support interactive data visualization by itself or with intercomponent highlighting (across map and graph) and synchronization (between maps). We selected the combination of data visualization components for each topic on the basis of the data type and the spatial resolution of data. For example, pollution data are presented solely with map representations (Figure 1a), whereas birth data include graphical representations of indicators (Figure 1b).

cluded in the back-end site and are organized by topic area. This organization does require the user to filter through and select variables of interest, but is an important aspect of the data exploration process. In numerous pages the option to save data selections is included so that users can go back and reload previous visualizations. Users can also export various results and outputs into Excel, CSV, and PDF. In future iterations we plan to add ability to download spatial data formats as well.
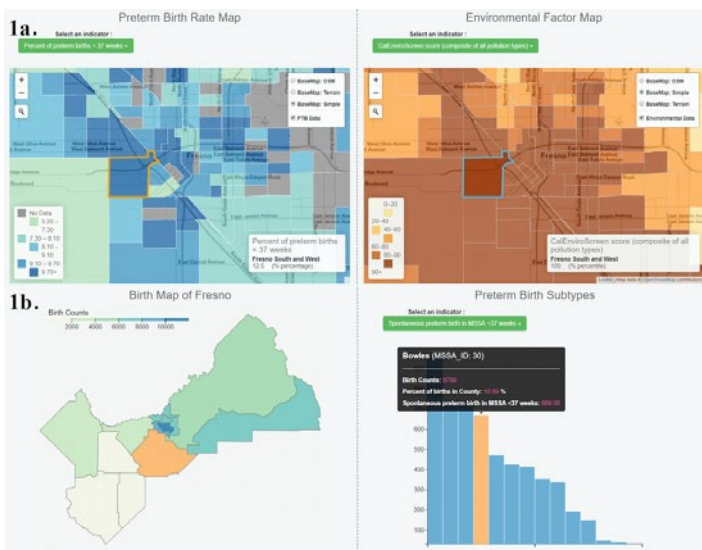

Figure 1. A. Demonstration of the environmental pollution topic of an online geographic data visualization tool for exploring preterm birth in Fresno County, California, using linked maps. Different indicators can be selected for each map. B. The birth topic with overall preterm birth shown on the map and spontaneous preterm birth shown on the indicator histogram. Data elements are linked so that as selection of 1 data value occurs, the location of that data value is displayed on the map.
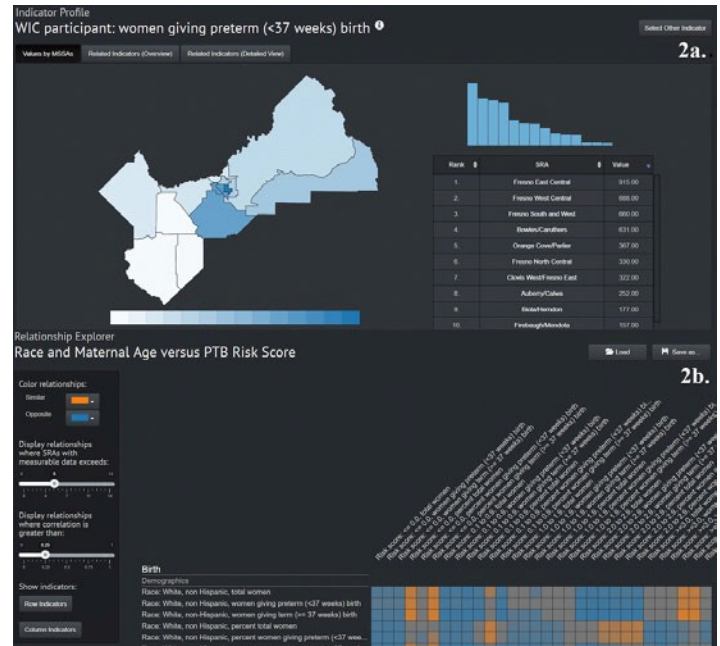

Figure 2. A. Demonstration of the indicator explorer feature of an online geographic data visualization tool to explore preterm birth in Fresno County, California. The indicator explorer feature allows mapping and side-by-side histogram evaluation of any of the hundreds of variables included in the back-end site. As one variable is selected, the relationship between that indicator and others can be viewed in alternate tabs. B. The relationship explorer tab, which builds correlation matrixes for selected indicators. Users can change color, specify the minimum number of geographical units that must be included, and specify the minimum correlation value.

Underlying the back end of the site is the National Science Foundation–supported DELPHI (Data E-Platform Leveraged for Patient Empowerment and Population Health Improvement) developed at UC San Diego, which was implemented to design an asthma management system in San Diego County (19) and can support data-driven public health discoveries from multiple approaches. The site is organized by 4 main visualization functionalities to allow for data exploration by geographic region, by indicator (Figure 2a), by indicator relationships, and by correlation matrixes (Figure 2b). These different functions give the user multiple options to explore data relationships through dynamic pie chart visuals, histograms, ranked associations with other indicators, and heat-map matrix visuals. All functions are linked and highlighted when hovering over them so that as 1 data element or geographical unit is selected, the corresponding units are highlighted. All data are in-

We used several popular programming framework environments and libraries to develop the applications. Both the front-end and back-end site were developed with the Bootstrap framework (Bootstrap) and the Node.js (Joyent, Inc) environment. The visualization components are built around the D3.js (Mike Bostock) library in the back-end site, and the front-end site mixes up D3.js and its extensions for the chart items and the open-source Leaflet.js (Vladimir Agafonkin) for some mapping features. The data framework transfers processed data between the PostgreSQL (PostgreSQL Global Development Group) database and the visualization platforms through Node.js routes and server-side SQL functions.

Stakeholder Involvement

The design team followed a user-centered design model for facilitating stakeholder involvement and designing the tool for optimal interface success, which proved successful for designing web-mapping, visualization, and data exploration projects (20–22). We used the user–utility–usability loop developed by Roth, Ross, and MacEachren, in which we collected input and feedback on needs and designs from preterm birth researchers and stakeholders (user), prompting revisions to the conceptualization and functional requirements of the tool (utility), leading to new versions of the data visualization tool (usability) for additional evaluation by our target users, thus restarting the loop (23).

An initial draft of the visualization tool using the DELPHI platform was presented to the Fresno County Preterm Birth Initiative in early 2017. Feedback obtained from stakeholders included comments about variables to be used, geographic resolution, and the need for a more user friendly and simple site that would allow less advanced users to explore and visualize key data sets. The public-facing side of the site was presented again to the Fresno County Preterm Birth Initiative in early 2018 to obtain further input and to narrow down the key topics of interest. The design team then worked with the Fresno County Preterm Birth Initiative Shared Measures Committee, a subgroup of the Fresno initiative that focuses on data and measurement issues, over several meetings to come up with 8 topics to feature on the public-facing side of the tool. The Shared Measures Committee comprises cross-sector leaders and experts in measurement and evaluation and a mother who experienced preterm birth. This committee helps set goals, inform strategies, and establish or develop measures of progress for the Fresno initiative. The iterative feedback between the design team and the Shared Measures Committee was critical for the design of the tool and for determining how the 8 topics should be populated and visualized. Having the design team attend multiple meetings of the Fresno initiative gave additional context to challenges surrounding preterm birth, such as social disadvantage and health care access, and the need to represent such phenomena in maps in a transparent way. Further discussions highlighted how the tool needed to be easy to use for nonprofit organizations to create figures for grant applications and accessible to the public and elected officials. In addition, the tool needed to be bilingual, for both English and Spanish speakers.

The tool was presented to the public and other Fresno initiative stakeholders in July 2018 at the Fresno County Preterm Birth Initiative Forum. The purpose of the event was to convene community members and stakeholders to raise awareness about preterm birth and communicate Preterm Birth Initiative strategies, successes, and challenges. As an example, the team walked through an assessment of environmental pollution as related to preterm birth from the viewpoint of a concerned community member and then from the viewpoint of a public health researcher. Beginning with the front-end, demographics around a neighborhood of interest as related to preterm birth were examined using pie charts and histograms with linked maps (eg, Figure 1b). Moving to the environmental pollution tab, rates of ozone, PM2.5 (atmospheric particulate matter with a diameter of less than 2.5 micrometers), traffic density, and toxic release were examined in relation to 3 neighborhoods in downtown Fresno (eg, Figure 1a). Resulting figures could be used in public presentations, community discussions, or public grant applications. The demonstration then turned to the back-end site where a researcher might want to examine all possible pollution factors in association with not only preterm birth rates, but also mothers' rates of hypertension, community asthma rates, and poverty indicators by using a correlation matrix (eg, Figure 2b). The tool was presented alongside a poster outlining "Unequal Neighborhoods: Fresno" research on historical zoning and land use policies influencing health disparities. At the event, attendees were encouraged to interact with the tool and with the design team to ask questions and provide feedback. They were given a link to submit further feedback through an online survey and to stay engaged with the project. Feedback is considered and incorporated into the tool design.

Challenges, Next Steps, and Final Thoughts

The project faced numerous challenges. Data collection, formatting, and updating from various agencies and resources will continue to be a significant task moving forward. The issue of updating data in future years as funding for the project expires is a current topic of discussion for the Fresno County initiative. Our current funding from the California Preterm Birth Initiative supports travel to Fresno and meeting support with stakeholders, a part-time developer, and a part-time–equivalent data-focused researcher. At the very least, a sustainable data updating plan will need to be enacted so that the tool can become a staple feature of the Fresno County Preterm Birth Initiative. As new and updated data are pulled into the tool, the team will also have to face the challenge of demonstrating change from time-based data. The collaborative nature of the project, while an essential feature of making the tool a success, is also not without challenges. When soliciting feedback from stakeholders, the design team had to be cognizant of what was possible in terms of data limitations, the design work involved, and timeline management. Balancing what is possible and what is feasible within a project scope and budget is a challenge for any collaborative project.

The next step in the process will be further dissemination of the tool, because use of the tool by stakeholders thus far has been limited. To this end, the authors are planning a series of workshops focused on local policy makers and staff members, nonprofit organizations, and academic researchers. The tool will also be presented to obstetrics and gynecology residents and students during a UCSF–Fresno grand rounds. The team is also creating a set of how-to videos with examples and trainings to explain how to use tool components and featuring specific use cases such as generating visuals for a grant application or for presentation to a community group, drawing on feedback and previous research showing the need for training across health and geospatial visualization tools (24). We are working to create an Esri story map as an example of how a narrative concerning preterm birth might be created using data on the tool site, and we will also explore open source visualization publishing tools to allow advanced users to create their own story-like narratives from the data visualizations. In the future, a major goal will be to allow users to securely upload and analyze their own collected data sets against the backdrop of public data amalgamated in the tool. We are also in the process of developing similar tools for the San Francisco and San Diego regions.

The Fresno Preterm Birth Initiative's data visualization tool provides information on births, environmental exposures, sociodemographic characteristics, the built environment, and access to care in a format that makes the information more accessible than it has ever been before. The front-end site design was customized to suit each topic and made as intuitive as possible, allowing community users to engage with preterm birth data, and allowing stakeholders from nonprofit organizations to use figures and data in the tool for grant writing, communication, and policy discussions about preterm birth. The back end offers significantly more data for exploration, along with more powerful tools for data visualization and relationship discovery. The design process, with significant input from the Fresno initiative, preterm birth researchers, and other stakeholders demonstrates the power of working with health and community experts who work daily to improve public health. The original concept of the tool focused solely on using the DELPHI platform and would have been inaccessible to many potential users. Through iterative feedback the design team was able to create a user-friendly front-end site and had significant assistance in identifying and obtaining data sets for inclusion in the tool. Furthermore, because of the collaboration, dissemination of the tool is increased by public events such as the Fresno County Preterm Birth Initiative Forum and publicity of the tool through the initiative's platforms. Lastly, our feedback to date is unanimously positive. Users of the tool report powerful value for the community at large and express hope to see it grow.

Acknowledgments

Author Information

Corresponding Author: Marta Jankowska, 9500 Gilman Drive, MC 0811, La Jolla, CA 92093-0811. Telephone: 858-246-1826. Email: majankowska@ucsd.edu.

Author Affiliations: [1]Calit2/Qualcomm Institute, University of California San Diego, La Jolla, California. [2]California Preterm Birth Initiative, University of California San Francisco, San Francisco, California. [3]Department of Pediatrics, University of California San Diego, La Jolla, California. [4]Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, California. [5]Fresno County Preterm Birth Initiative, California State University, Fresno. [6]Central Valley Health Policy Institute, California State University, Fresno. [7]San Francisco Department of Public Health, San Francisco, California. [8]Department of Obstetrics and Gynecology, University of California San Francisco, Fresno Center for Medical Education and Research, Fresno, California.

References

1. Tomashek KM, Shapiro-Mendoza CK, Davidoff MJ, Petrini JR. Differences in mortality between late-preterm and term singleton infants in the United States, 1995-2002. J Pediatr 2007;151(5):450–6, 456.e1.
2. Swamy GK, Ostbye T, Skjaerven R. Association of preterm birth with long-term survival, reproduction, and next-generation preterm birth. JAMA 2008;299(12):1429–36.
3. Shapiro-Mendoza CK, Lackritz EM. Epidemiology of late and moderate preterm birth. Semin Fetal Neonatal Med 2012; 17(3):120–5.
4. Parkinson JRC, Hyde MJ, Gale C, Santhakumaran S, Modi N. Preterm birth and the metabolic syndrome in adult life: a systematic review and meta-analysis. Pediatrics 2013; 131(4):e1240–63.
5. Tinnion R, Gillone J, Cheetham T, Embleton N. Preterm birth and subsequent insulin sensitivity: a systematic review. Arch Dis Child 2014;99(4):362–8.

6. Goldenberg RL, Culhane JF, Iams JD, Romero R. Epidemiology and causes of preterm birth. Lancet 2008; 371(9606):75–84.

7. Auger N, Gamache P, Adam-Smith J, Harper S. Relative and absolute disparities in preterm birth related to neighborhood education. Ann Epidemiol 2011;21(7):481–8.

8. O'Campo P, Burke JG, Culhane J, Elo IT, Eyster J, Holzman C, et al. Neighborhood deprivation and preterm birth among non-Hispanic Black and White women in eight geographic areas in the United States. Am J Epidemiol 2008; 167(2):155–63.

9. Gehring U, Wijga AH, Fischer P, de Jongste JC, Kerkhof M, Koppelman GH, et al. Traffic-related air pollution, preterm birth and term birth weight in the PIAMA birth cohort study. Environ Res 2011;111(1):125–35.

10. Stevens KB, Pfeiffer DU. Sources of spatial animal and human health data: Casting the net wide to deal more effectively with increasingly complex disease problems. Spat Spatio-Temporal Epidemiol 2015;13:15–29.

11. Anselin L. Interactive techniques and exploratory spatial data analysis. In: Longley P, Goodchild M, Maguire D, and Rhind D, editors. Geographical information systems: principles, techniques, management and applications, second edition. New York (NY): Wiley; 1999, p 251–64.

12. Anselin L, Sridharan S, Gholston S. Using exploratory spatial data analysis to leverage social indicator databases: the discovery of interesting patterns. Soc Indic Res 2007; 82(2):287–309.

13. MacEachren AM, Kraak M-J. Research challenges in geovisualization. Cartogr Geogr Inf Sci 2001;28(1):3–12.

14. Vadillo-Ortega F, Osornio-Vargas A, Buxton MA, Sánchez BN, Rojas-Bracho L, Viveros-Alcaráz M, et al. Air pollution, inflammation and preterm birth: a potential mechanistic link. Med Hypotheses 2014;82(2):219–24.

15. Katsis Y, Balac N, Chapman D, Kapoor M, Block J, Griswold WG, et al. Big data techniques for public health: a case study. In: 2017 IEEE/ACM International Conference on Connected Health: Applications, Systems, and Engineering Technologies. 2017 July 17–19; Philadelphia, Pennsylvania. Piscataway (NJ): IEEE Press; 2017, pp 222–231.

16. Arribas-Bel D, Reades J. Geography and computers: past, present, and future. Geogr Compass 2018;12(10):e12403.

17. Baer RJ, Lessard L, Jankowska MM, Anderson JG, Block J, Chambers CD, et al. Comparison of risk and protective factors for preterm birth in rural, suburban, and urban Fresno County, California. J Epidemiol Res 2018;4(2):50.

18. American Medical Association. International Classification of Diseases: ICD-9-CM 2008. 2008 ed. Chicago, IL: American Medical Association; 2007.

19. Katsis Y, Baru C, Chan T, Dasgupta S, Farcas C, Griswold W, et al. DELPHI: Data-E-platform for Personalized Population Health. In: IEEE 15th International Conference on e-Health Networking, Applications and Services (Healthcom). 2013 October 9–12; Lisbon, Portugal. Piscataway (NJ): IEEE Press; 2013.

20. Koh LC, Slingsby A, Dykes J, Kam TS. Developing and applying a user-centered model for the design and implementation of information visualization tools. In: Proceedings of the International Conference on Information Visualisation, IV. 2011 July 13–15; London, UK. Piscataway (NJ): IEEE Press; 2011, pp 90–95.

21. McGuire M, Gangopadhyay A, Komlodi A, Swan C. A user-centered design for a spatial data warehouse for data exploration in environmental research. Ecol Inform 2008;3(4-5):273–85.

22. Bernard J, Daberkow D, Fellner D, Fischer K, Koepler O, Kohlhammer J, et al. VisInfo: a digital library system for time series research data based on exploratory search — a user-centered design approach. Int J Digit Libr 2015;16(1):37–59.

23. Roth R, Ross K, MacEachren A. User-centered design for interactive maps: a case study in crime analysis. ISPRS Int J Geoinf 2015;4(1):262–301.

24. Robinson AC, MacEachren AM, Roth RE. Designing a web-based learning portal for geographic visualization and analysis in public health. Health Informatics J 2011;17(3):191–208.

25. US Census. American Community Survey. https://www.census.gov/programs-surveys/acs/data.html. Accessed June 27, 2019.

Table

Table. Data Topics, Tool for Visualization of Preterm Birth and Environmental Factors: Variables, Source, and Date Ranges, Fresno County, California[a]

Births	Birth counts, preterm birth by week, spontaneous / indicated	OSHPD, data agreement	2007–2012
Pregnancy indicators	Nulliparity, previous cesarean, previous preterm birth, interpregnancy intervals, prenatal care, using WIC during pregnancy	OSHPD, data agreement	2007–2012
Demographics of women	Race/ethnicity, age, education, place of birth	OSHPD, data agreement	2007–2012
Health of women	BMI, diabetes, hypertension, infection, anemia, mental illness, smoked, drug/alcohol use	OSHPD, data agreement	2007–2012
Preterm birth risk	Risk scores for preterm birth calculated from several variables	OSHPD, data agreement	2007–2012
Care access	Primary care physicians, dental care, psychiatric care	OSHPD, https://oshpd.ca.gov/data-and-reports	2014
	Hospitals, clinics, WIC locations, federally qualified health centers	CHHS, https://data.chhs.ca.gov	2015–2016
Environmental pollution	Ozone, PM2.5, diesel, drinking water contaminations, pesticides, traffic density, clean-up sites, asthma ER visits, overall pollution score	CalEnviroScreen 3.0, https://oehha.ca.gov/calenviroscreen/report/calenviroscreen-30	2010–2014
Demographics of region	Age, race/ethnicity, education, population density, marital status	ACS,[b] US Census, https://factfinder.census.gov	2007–2012 (ACS), 2010 (Census)
Socioeconomic indicators	Poverty, income, unemployment, housing, crowding, female headed households, car ownership, average house value	ACS,[b] Census, https://factfinder.census.gov	2007–2012 (ACS), 2010 (Census)
	Crime	ESRI, Business Analyst Data (paid subscription)	2016
	Liquor stores	California Alcohol and Beverage Control, https://www.abc.ca.gov	2012
Cultural indicators	Place of birth, living in same location 1 year ago, diversity index, language spoken	ACS,[b] https://factfinder.census.gov	2007–2012

Abbreviations: ACS, American Community Survey; CHHS, California Health and Human Services; ER, emergency room; ESRI, Environmental Systems Research Institute; OSHPD, Office of Statewide Health Planning and Development; PM2.5, particulate matter 2.5 (atmospheric particulate matter with a diameter of less than 2.5 micrometers); WIC, Special Supplemental Nutrition Program for Women, Infants, and Children.
[a] Topics and indicators from the OSHPD data agreement set include the 81,021 women from the data set. All other indicators are general environmental or population data sets.
[b] American Community Survey (25).

(continued)

Table. Data Topics, Tool for Visualization of Preterm Birth and Environmental Factors: Variables, Source, and Date Ranges, Fresno County, California[a]

Topic			
Built environment	Bicycle and pedestrian traffic collision	Statewide Integrated Traffic Records System, http://iswitrs.chp.ca.gov	2010–2012
	Walkability	Environmental Protection Agency, https://www.epa.gov/smartgrowth/smart-location-mapping	2010–2012
	Vegetation index, urbanization index, water index	Landsat composite satellite imagery (30m), downloaded and calculated in Google Earth Engine	2010
	Parks	ESRI, http://www.esri.com/data/data-maps	2010–2018
	Transit stops	Fresno County Rural Transit Agency, Fresno Area Express, www.transitwiki.org	2016
	Food access and deserts	United States Department of Agriculture, https://www.ers.usda.gov/data-products/food-access-research-atlas	2015
Health indicators	California Adolescent Sexual Health Needs Index	California Department of Public Health, https://www.cdph.ca.gov	2014
	Child care facilities, elder care facilities, counseling services	CHHS, https://data.chhs.ca.gov	2014–2017
	Immunization rates, newborn deaths, asthma hospitalizations	CHHS, https://data.chhs.ca.gov	2014–2016
	Healthy Priorities Index	Fresno County, http://gis.co.fresno.ca.us/HealthPriorityNDX/	2010–2014

Abbreviations: ACS, American Community Survey; CHHS, California Health and Human Services; ER, emergency room; ESRI, Environmental Systems Research Institute; OSHPD, Office of Statewide Health Planning and Development; PM2.5, particulate matter 2.5 (atmospheric particulate matter with a diameter of less than 2.5 micrometers); WIC, Special Supplemental Nutrition Program for Women, Infants, and Children.

[a] Topics and indicators from the OSHPD data agreement set include the 81,021 women from the data set. All other indicators are general environmental or population data sets.

[b] American Community Survey (25).

ORIGINAL RESEARCH

Geographic and Social Factors Associated With Chronic Disease Self-Management Program Participation: Going the "Extra-Mile" for Disease Prevention

Julie Bobitt, PhD[1]; Liliana Aguayo, PhD[2,3]; Laura Payne, PhD[1]; Taylor Jansen[4]; Andiara Schwingel, PhD[1]

PEER REVIEWED

Summary

What is already known on this topic?

Two widely accepted chronic disease self-management education programs, the Chronic Disease Self-Management Program (CDSMP) and the Diabetes Self-Management Program (DSMP), give participants knowledge and skills to manage chronic diseases. However, program attendance and completion are a challenge to many program providers, and little is known about their barriers and facilitators.

What is added by this report?

Participation in CDSMP and DSMP was associated with distance traveled from home to workshop site, program type, class size, and education.

What are the implications for public health practice?

Our findings underscore the need to develop strategies to improve attendance in CDSMP and DSMP among adults aged 50 or older.

Abstract

Introduction

We examined geographic and social factors associated with participation in the Chronic Disease Self-Management Program (CDSMP) and the Diabetes Self-Management Program (DSMP) implemented at 144 sites in Illinois.

Methods

Programs were delivered by trained facilitators, once per week, during 6 weeks to 1,638 participants aged 50 or older. Of the 1,638 participants, we included in our analysis 1,295 participants with complete geographic information and baseline data on demographic characteristics, health history, and health behaviors. We assessed the following program data: program type (CDSMP or DSMP), workshop location, class size, and number of sessions attended by participants. We geocoded each participant's home address, classified the home address as rural or urban, and calculated the distance traveled from the home address to a workshop. We used linear and logistic regression analyses to examine the associations between participant and program factors with number of sessions attended and odds of program completion by whether participants lived in an urban or rural county.

Results

Average program attendance was 4.2 sessions; 71.1% (1,106 of 1,556) completed 4 or more sessions. Most participants enrolled in CDSMP (59.6% [954 of 1,600]), but DSMP had greater completion rates. Less than 7% (85 of 1,295) of our sample lived in a rural county; these participants had better completion rates than those living in urban counties (89.4% [76 of 85] vs 75.6% [890 of 1,178]). Traveling shorter distances to attend a workshop was significantly associated with better attendance and program completion rates among urban but not rural participants. The number of sessions attended was significantly higher when class size exceeded 16 participants. Not having a high school diploma was significantly associated with lower levels of attendance and program completion.

Conclusion

Participation in CDSMP and DSMP was associated with distance traveled, program type, class size, and education. Increasing participation in self-management programs is critical to ensure participants' goals are met.

Introduction

Management of chronic health conditions such as diabetes, hypertension, and arthritis is a public health concern among the growing older population (1). Currently, 68% of older adults have at least 2 chronic diseases (2), and people who are a racial/ethnic minority, live in a rural area, or have lower socioeconomic status are disproportionally affected (3,4). People with chronic diseases have higher risks of disability, loss of independence, and reduced quality of life, and these higher risks can lead to decreases in productivity and increases in health care costs and the burden of caregivers (5). Furthermore, chronic diseases account for 95% of health care costs in the United States (1). However, chronic diseases can be prevented and managed through healthy lifestyles and self-management education (6,7).

Two widely accepted chronic disease self-management education programs, the Chronic Disease Self-Management Program (CDSMP) and the Diabetes Self-Management Program (DSMP), give participants knowledge and skills to manage chronic diseases (7). These programs are endorsed by the National Council on Aging, and they have strong evidence to support their implementation (8). However, program attendance and completion are a challenge to many program providers, and little is known about their barriers and facilitators. Factors that influence participation are particularly important in rural areas, which often have limited access to health care services and chronic disease management programs (9). The limited evidence available about program participation in chronic disease management programs suggests that participation is greater among healthier people and urban dwellers (10,11). The objective of our study was to describe geographic and social factors associated with participation of adults aged 50 or older in chronic disease self-management programs in rural and urban areas.

Methods

As part of a state-wide effort to disseminate and implement chronic disease self-management programs, many organizations (senior centers, Cooperative Extension offices, assisted living facilities, Area Agencies on Aging, and local hospitals) implemented the CDSMP and DSMP at 144 sites in Illinois during 2016–2017. Workshop sessions were offered once per week for 6 weeks by trained facilitators in the community's language of preference.

Participants aged 50 or older completed questionnaires at the beginning of the first workshop. The questionnaires collected information about demographic characteristics (age, sex, race, and education), health history, physical activity level, whether the respondent cared for someone with a long-term health problem or disability, and health care practices (eg, health confidence, self-efficacy in communicating with health providers). We also collected data on each participant's home address and each workshop location.

Measures

Program attendance and completion. We obtained information on attendance (number of workshops attended by participants), class size, and type of program (CDSMP or DSMP) from the workshop facilitators. The number of workshop sessions attended ranged from 1 to 6. According to program developers, attendance in 4 or more sessions is considered program completion (7). We measured program attendance as a continuous variable and completion as a dichotomous variable. We classified participants who completed 4 or more sessions as completers.

The Stanford guideline for class size requires their self-management workshops (including CDSMP and DSMP) to have a minimum of 10 and a maximum of 16 participants (12). However, adherence was low in evaluations of Stanford's CDSMP (12). Thus, we examined adherence to class size criteria by comparing attendance and completion rates among participants in workshops that satisfied the class-size criteria and workshops that did not. We further compared attendance and completion rates for 3 workshop sizes: small (<10 participants), medium (10–16 participants), and large (>16 participants). In all regression analyses, we evaluated the effects of class size as a continuous measure.

Geographic factors. Geographic factors associated with program attendance and completion were whether a participant lived in a rural or an urban county and the distance traveled from a participant's home address to the workshop site. We geocoded the address for each workshop site and participant residence into latitude and longitude by using Google Earth Pro. We used a participant's home address to identify county of residence; we then classified these counties as rural or urban according to classifications of the US Census Bureau's Office of Management and Budget (13,14), which defines a metropolitan (ie, urban) area as having a population of 50,000 or more and a micropolitan (ie, rural) area as having a population of 10,000 to 49,999 (13,14). We

used rural as the reference category. After mapping workshop sites and participant residences by using ArcMap 10.5.1 (Esri), we calculated the distance in miles traveled by participants from home to a workshop site by using the Network Analysis tool Origin-Destination Cost Matrix. We used the dot-density function to indicate the correct number of participants per county while protecting information on participants' exact residential locations.

To examine the effect of proximity (including on-site delivery) on attendance and completion, we dichotomized data on distance traveled by participants from home to a workshop into 2 categories: participants who traveled less than 0.1 miles (considered living in proximity) and participants who traveled 0.1 miles or more (considered not living in proximity). Participants who traveled less than 0.1 miles included participants who resided at a workshop location (residents of community housing programs, nursing homes, assisted living facilities, retirement communities, or other organizations that hosted workshops). We adopted 0.1 miles as a cut point on the basis of previous research. One study on mathematical modeling of proximity relations suggested 0.1 miles as the minimum distance for linguistic proximity analyses (15); another study used distances shorter than 200 m (approximately 0.12 miles) to set accessibility benchmarks for walking and public transportation journeys among older adults (16); and a US survey on walking for transportation that found that older adults and those with chronic diseases were more likely to favor walking short distances (17).

Social factors. We collected self-reported data on age, sex, race/ethnicity, level of education, and care for someone with a long-term health problem or disability. We coded race/ethnicity as a dummy variable, with non-Hispanic white as the reference category. Participants reported their education level, and we dichotomized responses into participants who did not receive a high school diploma and those who received a high school diploma or more. Participants were asked if they cared for someone with a long-term health problem or disability, and responses were dichotomized into yes or no. We assessed weekly physical activity by asking about time spent in physical activities such as walking, bicycling, and gardening. We classified answers into 2 categories to examine differences in attendance between respondents who met the national physical activity guidelines (≥150 min per week [18]) and those who did not. Finally, we collected data on class size for each workshop.

Statistical analyses

Overall, 1,638 adults aged 50 or older participated in CDSMP or DSMP; for 38 of these participants, we did not have information on which program they attended. We excluded 343 participants from analysis because we did not have their geographic information; our analytic sample consisted of 1,295 participants. We calculated descriptive statistics for the overall sample (those with geographic information and those without) and for our analytic sample, stratified by rural and urban residence. Not all 1,295 participants answered all questions on the questionnaire; we calculated percentages according to the number of participants who answered each question. For all available data, we conducted t tests to compare participants' outcomes and identify potential cofounders. Linear regression examined factors associated with program attendance. Logistic regression examined how these factors influenced the odds of program completion. Logistic regression tested the odds of completion by using the dependent dichotomous variable to examine completion of 4 or more sessions. Independent variables were added by using the enter method, which enters variables into the model simultaneously. Both linear and logistic regression tested the influence of meeting the physical activity recommendations, living in proximity to programs (<0.1 miles), class size, type of program attended (CDSMP or DSMP), and the main effects and interaction of the miles traveled from participant's home to a workshop, by residence in an urban or rural county. We found that attendance and completion were not influenced by caregiving, and thus we did not include this variable in our analyses. Statistical models controlled for the effect of sex, race/ethnicity, education, and program type. All analyses were computed in SPSS Statistics 24 (IBM Corporation).

Results

Overall, 59.6% (954 of 1,600) participants attended CDSMP, and 40.4% (646 of 1,600) attended DSMP. Of the 1,295 participants in our analytic sample, 93.4% (n = 1,210) lived in urban counties, and 6.6% (n = 85) participants lived in rural counties (Table 1 and Figure 1).
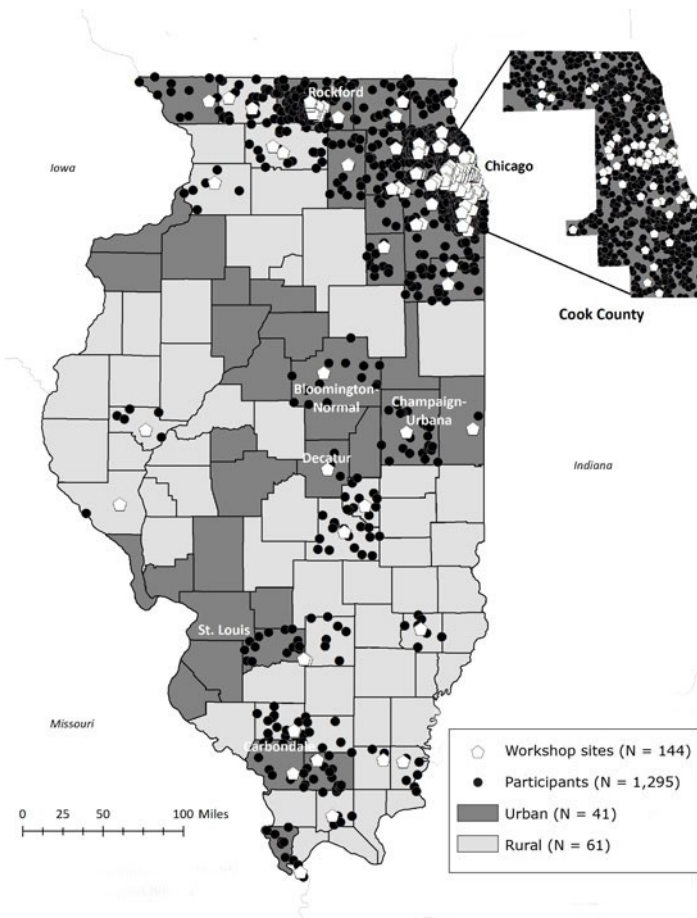
Figure 1. Location of workshop sites for the Chronic Disease Self-Management Program and the Diabetes Self-Management Program and distribution of participants' home addresses by rural and urban counties in Illinois during 2016–2017. We used the dot-density function to indicate the correct number of participants per county while protecting information on participants' exact residential locations. Data sources: Illinois Pathways to Health (19), US Census Bureau (20).

Of the 1,556 participants for whom we had attendance data, 1,106 (71.1%) completed 4 or more sessions (Table 1), and mean (standard deviation [SD]) attendance was 4.2 (1.9) sessions. Participants living in rural counties had a higher completion rate (76 of 85 [89.4%]) than participants living in urban counties (890 of 1,178 [75.6%]). Program completion and mean number of sessions attended in CDSMP and DSMP varied by workshop site. Overall, mean (SD) attendance in DSMP (4.4 [1.8] sessions) was significantly greater than attendance in CDSMP (4.1 [1.9] sessions; $t_{1,550}$ = −2.9, P = .004). Overall, estimated median (interquartile range [IQR]) distance traveled to a workshop site was 2.0 (0.4–5.2) miles. Participants in rural counties traveled a median (IQR) distance of 1.3 (0.5–9.2) miles to a workshop, and participants in urban counties traveled a median (IQR) of 2.1 (0.4–5.0) miles. Most

(88.7%; 71 of 80) participants in rural counties were non-Hispanic white; the 1,064 participants in urban counties were more racially and ethnically diverse, with 52.4% (n = 558) non-Hispanic white, 30.9% (n = 329) non-Hispanic black or African American, and 8.6% (n = 91) Hispanic. Of the 1,295 participants in our analytic sample, 28.6% (n = 370) reported caring for someone with a long-term health problem or disability.

Workshop sizes ranged from 4 to 39 participants (mean [SD] = 16.0 [7.3] participants). Overall, 41.6% (653 of 1,569) of participants attended workshops that satisfied the Stanford guideline of 10 to 16 participants. Of the 58.4% (916 of 1,569) who attended workshops that did not satisfy the Stanford guideline, 19.9% (313 of 1,569) of participants attended workshops with fewer than 10 participants, and 38.4% (603 of 1,569) attended programs with more than 16 participants. Mean [SD] attendance in small workshops (4.3 [1.8] sessions) was greater than attendance in workshops that satisfied the Stanford guideline (4.1 [1.9] sessions) ($t_{1,519}$ = 2.1; P = .04). We found no significant differences in mean attendance between large (>16 participants) and small (<10 participants) workshops. However, mean (SD) attendance per class was higher for large workshops (4.4 [1.8] sessions) than for workshops that satisfied class size requirements (4.1 [1.9] sessions) ($t_{1,215}$ = −2.7; P = .008).

The overall model explained 4.8% of the variance in the number of sessions attended (R = 0.24, adjusted R^2 = 0.048) (Table 2). Not having a high school diploma (β = −0.09, P = .01) and class size (β = 0.11, P = .01) were significantly associated with fewer sessions attended. Fewer miles traveled from home to a workshop (β = −0.12, P = .001) was significantly associated with a greater number of sessions completed.

Moderation between distance and attendance among participants who resided in urban counties

We found a significant interaction between miles traveled from home to a workshop and whether participants lived in an urban or a rural county (β = 0.09, P = .049). The simple slopes showed that distance traveled from home to a workshop significantly influenced the number of sessions completed by participants living in an urban county (b = −0.29, P < .001) (Figure 2). In contrast, distance traveled had no significant effect on the number of sessions completed by participants living in a rural county (b = 0.04, P = .17).
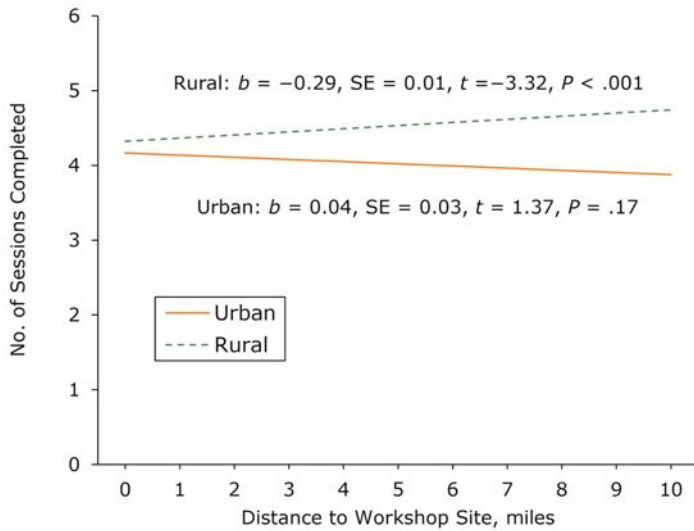
Figure 2. Simple slopes describing the association between the distance in miles between a participant's home address to a workshop site and the number of program sessions completed among from adults aged 50 or older in the Chronic Disease Self-Management Program or the Diabetes Self-Management Program in Illinois during 2016–2017. We used a participant's home address and criteria from the US Census Bureau's Office of Management and Budget (13,14) to determine whether the participant lived in a rural county or an urban county. Abbreviation: SE, standard error.

Although we did not find an interaction effect, the main effects of several geographic and social factors were significantly associated with the odds of program completion (Table 3). The overall logistic regression model explained 9% of the variance in odds of completing the program (R = 0.06, Nagelkerke-adjusted R^2 = .090). Participants with a high school diploma or more were nearly 2 times as likely as participants who did not have a high school diploma to complete at least 4 workshop sessions (odds ratio [OR] = 0.54, P = .02). Participants who traveled less than 0.1 miles to attend a workshop session were 1.69 times as likely as participants who traveled 0.1 miles or more to complete the program (OR = 1.69, P = .04). Odds of program completion decreased as distance between a participant's home and workshop site increased (OR = 0.96, P = .008). The interaction between distance traveled by whether participants lived in an urban or a rural county was not significant (OR = 1.20, P = .07).

Discussion

Our findings indicate that travel distance was a barrier for attendance among participants who lived in urban counties but not rural counties. A study in 2014 also found that distance was a barrier to accessing health care resources more often among urban dwellers than rural dwellers (21). That study asserted that rural residents are used to navigating distances and therefore may negotiate them better, whereas urban dwellers may have difficulties finding transportation for even a short distance if they have no vehicle or have mobility issues (21). Therefore, when working with older adults, especially those in urban communities, program planners should pay attention the distance people must travel to get to program sites. Sites should be situated in neighborhoods where the target population lives. On-site delivery (delivery in senior housing programs or community housing sites where adults aged 50 or older reside) may be an option.

Rural dwellers in our study had higher rates of completion and attendance than urban dwellers. This finding is consistent with findings of a nationwide study of chronic disease management dissemination that examined data on more than 300,000 participants in rural and urban areas (9). One explanation is that health education programs may compete with other activities to which urban dwellers have access locally. In contrast, rural communities tend to offer fewer "distractions" and residents may make such health education programs their priority.

Although program coordinators reached out to rural areas through senior centers, Extension offices, AAAs, and local hospitals, rural areas were underserved. Rural areas are home to 18.6% of adults aged 65 or older in Illinois (22), yet they represented only 6.6% of our sample. Several factors, such as access to appropriate meeting facilities, affect rural service delivery (23). The availability of program facilitators and partnerships in the southern, less populated areas of Illinois was a challenge for program delivery. The workshops in rural areas were attended mostly by non-Hispanic white people. However, rural areas are more homogenous than urban areas in Illinois: only 9.5% of the population is black and 11.7% Hispanic (22). Rural areas can benefit from more culturally tailored recruitment strategies.

Attendance rates in DSMP were significantly greater than in CDSMP. The better attendance in DSMP could have been due to a more focused workshop content. Erdem and Korda reported higher completion rates for people with diabetes who participated in DSMP than in CDSMP, also attributing that outcome to the focus on diabetes in DSMP (24). The study suggested that higher completion rates could be due to factors such as type of recruiting methods, program site (ie, senior center vs health care facility), and type of program offered near participants' home. Future research should consider administering both the CDSMP with DSMP in the same location at the same time to determine differences in attendance or completion outcomes and whether any differences can be explained by geographic factors.

Our results indicated that class size was associated with attendance. This finding suggests that the experience of participating in a chronic disease program goes beyond its content. Participation in such programs likely promotes social interactions essential to motivating and encouraging attendance. Gallant reported on the valuable role of friends in chronic disease self-management and emphasized the importance of self-management educational programs to incorporate skills and strategies that enhance social interactions (25). Meek and colleagues found that older adults with chronic disease restrict their social engagement with family, friends, and community (25). They suggested that behavioral interventions could help older adults better manage their chronic conditions and maintain active social lives (25). These findings underscore the importance of recruitment that leads to larger classes, and ultimately, increases attendance. Although some studies found that programs with fewer participants had higher attendance rates, the settings for these programs varied, and therefore the results were inconclusive (12,24). Interestingly, these studies included sites that delivered programs to classes that were larger than the class size specified in the Stanford guideline. These findings warrant further study because adapting chronic disease self-management programs to fit a particular facility could affect program fidelity, which may also affect program efficacy (27). However, Smith and colleagues suggested that some flexibility can be allowed in program implementation to fit the needs of a particular facility as long as the core elements of the program are maintained and program outcomes are not compromised (28). Both Smith and colleagues (28) and Carvahlo and colleagues (27) suggested further research to determine whether significant outcomes can be achieved when interventions are adapted to a particular environment.

Low educational attainment was associated with lower attendance in rural and urban areas. Much literature exists on the association between education and health literacy, defined as "the degree to which individuals have the capacity to obtain, process, and understand basic health information and services needed to make appropriate health decisions" (29). Older adults have a double burden, with a disproportionally high prevalence of chronic diseases and a greater risk of poor health literacy (29). In the context of chronic disease self-management, education affects a person's ability to read health information, process oral communication, and conceptualize activities (30). Mackey and colleagues found that health literacy-sensitive interventions resulted in significant improvements in self-care practices (30). Educational barriers may be associated with a limited sense of purpose and belonging in health programs, which can create feelings of frustration and affect attendance and completion. Zoellner and colleagues underscored the importance of integrating recruitment strategies that attend to the needs of audiences with a low level of education (31). Although CDSMP

and DSMP materials address low literacy levels, our findings reiterate the importance of focusing on how chronic disease self-management programs are designed and marketed. Best practices include the creation of easy-to-read marketing and communication materials with appropriate language, font style, and font size (32). Graphic illustrations and experiential activities are useful strategies in mitigating some education limitations and could res- ult in increased attendance (32).

Our study had several limitations. First, we had a small sample of rural residents. Although we assessed distance objectively through geographic information, our assessment did not account for the time burden associated with various modes (eg, car, bus, walking) of transportation. Additional limitation was the use of only 2 measures of rurality/urbanicity (ie, rural and urban), which may not have accounted for racial/ethnic diversity among the rural population. Also, information about physical activity collected in the questionnaire did not include information on the intensity of physical activity. Therefore, we were unable to account for the difference between light and moderate or vigorous activity. Further examination is needed to better understand the role of travel time, modes of transportation, diversity among the rural population, and physical activity levels. Future studies should also consider using survey instruments to assess health literacy levels.

Our findings underscore the need to develop strategies to improve attendance in CDSMP and DSMP among adults aged 50 or older. Ideally, workshop sites should be located near to participants' homes to promote completion and increase attendance among urban dwellers. Classes with a larger number of participants should be a goal, keeping in mind program fidelity. Recruitment and program materials should be developed to appeal to people who may not have a high school diploma.

Acknowledgments

Author Information

Corresponding Author: Julie Bobitt, PhD, Interdisciplinary Health Sciences, College of Applied Health Sciences, University of Illinois at Urbana Champaign, 228 Huff Hall, MC-586, Champaign, IL 61820. Telephone: 217-333-6657. E-mail: jbobitt@illinois.edu.

Author Affiliations: [1]University of Illinois at Urbana Champaign, Champaign, Illinois. [2]Northwestern University Feinberg School of Medicine, Chicago, Illinois. [3]Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, Illinois. [4]University of Massachusetts Boston, Boston, Massachusetts.

References

1. Centers for Disease Control and Prevention. The state of aging and health in America 2013. Atlanta (GA): Centers for Disease Control and Prevention, US Department of Health and Human Services; 2013.
2. Lochner KA, Shoff CM. County-level variation in prevalence of multiple chronic conditions among Medicare beneficiaries, 2012. Prev Chronic Dis 2015;12:E07.
3. Louie GH, Ward MM. Socioeconomic and ethnic differences in disease burden and disparities in physical function in older adults. Am J Public Health 2011;101(7):1322–9.
4. Rural Health Information Hub. Rural health disparities 2017. November 2017. https://www.ruralhealthinfo.org/topics/rural-health-disparities. Accessed April 1, 2018.
5. Bryant LL, Altpeter M, Whitelaw NA. Evaluation of health promotion programs for older adults: an introduction. J Appl Gerontol 2006;25(3):197–213.
6. Lock SL, Belza B. Promoting healthy aging: a presidential imperative. Journal of the American Society on Aging. Winter 2016–2017. Vol. 40, no. 4. https://depts.washington.edu/mbwc/content/news-files/Basia.pdf. Accessed January 16, 2019.
7. Lorig KR, Sobel DS, Stewart AL, Brown BW Jr, Bandura A, Ritter P, et al. Evidence suggesting that a chronic disease self-management program can improve health status while reducing hospitalization: a randomized trial. Med Care 1999;37(1):5–14.
8. Siantz E, Aranda MP. Chronic disease self-management interventions for adults with serious mental illness: a systematic review of the literature. Gen Hosp Psychiatry 2014;36(3):233–44.
9. Smith ML, Towne SD, Herrera-Venson A, Cameron K, Kulinski KP, Lorig K, et al. Dissemination of chronic disease self-management education (CDSME) programs in the United States: intervention delivery by rurality. Int J Environ Res Public Health 2017;14(6):638.
10. Hanson S, Jones A. Is there evidence that walking groups have health benefits? A systematic review and meta-analysis. Br J Sports Med 2015;49(11):710–5.
11. White M, Adams J, Heywood P. How and why do interventions that increase health overall widen inequalities within populations? In: Babones S, editors. Health, inequality and society. Bristol (UK): Policy Press; 2009. p. 65–82.
12. IMPAQ International. Administration on Aging Chronic Disease Self-Management Program (CDSMP) process evaluation final report (2013). Report no. HHSP233201100492G. https://www.acl.gov/sites/default/files/programs/2017-02/CDSMPProcessEvaluationReportFINAL062713.pdf. Accessed July 12, 2018.
13. Office of Rural Health Policy. List of rural counties and designated eligible census tracts in metropolitan counties. Rockville (MD): US Department of Health and Human Services, Health Resources and Service Administration; 2010. https://www.hrsa.gov/sites/default/files/ruralhealth/resources/forhpeligibleareas.pdf. Accessed December 21, 2017.
14. Hall SA, Kaufman JS, Ricketts TC. Defining urban and rural areas in U.S. epidemiologic studies. J Urban Health 2006;83(2):162–75.
15. Yao X, Thill J-C. How far is too far? — A statistical approach to context-contingent proximity modeling. Trans GIS 2005;9(2):157–78.
16. Titheridge H, Achuthan K, Mackett R, Solomon J. Assessing the extent of transport social exclusion among the elderly. J Transp Land Use 2009;2(2):31–48.
17. Watson KB, Carlson SA, Humbert-Rico T, Carroll DD, Fulton JE. Walking for transportation: what do U.S. adults think is a reasonable distance and time? J Phys Act Health 2015;12(6 Suppl 1):S53–61.
18. Physical Activity Guidelines Advisory Committee. 2018 Physical Activity Guidelines Advisory Committee scientific report. Washington (DC): US Department of Health and Human Services; 2018.
19. Illinois Pathways to Health. https://www.ilpathwaystohealth.org. Accessed March 7, 2018.
20. US Census Bureau. TIGER/Line Shapefiles and TIGER/Line Files. https://www.census.gov/geo/maps-data/data/tiger-line.html. Updated October 4, 2018. Accessed March 1, 2018.
21. Haggerty JL, Roberge D, Lévesque JF, Gauthier J, Loignon C. An exploration of rural-urban differences in healthcare-seeking trajectories: implications for measures of accessibility. Health Place 2014;28:92–8.

22. US Census Bureau. American FactFinder. Race and Hispanic or Latino: state–urban/rural and inside/outside metropolitan and micropolitan area. 2010Census Summary File 1. https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_17_5YR_GCT0103.ST26&prodType=table. Accessed January 29, 2019.

23. Reddy P, Hernan AL, Vanderwood KK, Arave D, Niebylski ML, Harwell TS, et al. Implementation of diabetes prevention programs in rural areas: Montana and south-eastern Australia compared. Aust J Rural Health 2011;19(3):125–34.

24. Erdem E, Korda H. Self-management program participation by older adults with diabetes: Chronic Disease Self-Management Program and Diabetes Self-Management Program. Fam Community Health 2014;37(2):134–46.

25. Gallant MP. The influence of social support on chronic illness self-management: a review and directions for research. Health Educ Behav 2003;30(2):170–95.

26. Meek KP, Bergeron CD, Towne SD, Ahn S, Ory MG, Smith ML. Restricted social engagement among adults living with chronic conditions. Int J Environ Res Public Health 2018; 15(1):E158.

27. Carvalho ML, Honeycutt S, Escoffery C, Glanz K, Sabbs D, Kegler MC. Balancing fidelity and adaptation: implementing evidence-based chronic disease prevention programs. J Public Health Manag Pract 2013;19(4):348–56.

28. Smith ML, Hochhalter AK, Cheng Y, Wang S, Ory MG. Programmatic influences on outcomes of an evidence-based fall prevention program for older adults: a translational assessment. Transl Behav Med 2011;1(3):384–93.

29. Institute of Medicine Committee on Health Literacy. Nielsen-Bohlman L, Panzer AM, Kindig DA, editors. Health literacy: a prescription to end confusion. Washington (DC): National Academies Press; 2004.

30. Mackey LM, Doody C, Werner EL, Fullen B. Self-management skills in chronic disease management: what role does health literacy have? Med Decis Making 2016; 36(6):741–59.

31. Zoellner J, You W, Almeida F, Blackman KC, Harden S, Glasgow RE, et al. The influence of health literacy on reach, retention, and success in a worksite weight loss program. Am J Health Promot 2016;30(4):279–82.

32. Weiss BD. Removing barriers to better, safer care. Health literacy and patient safety: help patients understand. Manual for clinicians. 2nd ed. Chicago (IL): American Medical Association Foundation and American Medical Association; 2007.

Tables

Table 1. Characteristics of Adults Aged ≥50 Who Participated in the CDSMP or DSMP, Overall and in the Analytic Sample, Categorized as Living in a Rural County or an Urban County,[a] in Illinois, 2016–2017

Characteristic	Analytic Sample, by County of Residence		Overall[b] (N = 1,638)[c]
	Urban (n = 1,210)[c]	Rural (n = 85)[c]	
No. in analytic sample	1,210 (93.4)	85 (6.6)	1,295 (79.1)
Sex, no. (%)			
Male	244 (21.4)	12 (14.5)	282 (21.1)
Female	894 (78.6)	71 (85.5)	1,056 (78.9)
Age, mean (SD), y	70.7 (10.6)	74.7 (7.4)	71.0 (10.5)
Race, no. (%)			
Non-Hispanic white	558 (52.4)	71 (88.7)	679 (54.1)
Non-Hispanic black or African American	329 (30.9)	7 (8.8)	377 (30.0)
Hispanic	91 (8.6)	1 (1.2)	107 (8.5)
Other	86 (8.1)	1 (1.2)	92 (7.3)
Education			
<High school diploma	122 (11.0)	5 (6.0)	148 (11.3)
High school diploma or GED	279 (25.0)	35 (42.2)	332 (25.4)
Some college or technical school	392 (35.2)	28 (33.7)	459 (35.1)
≥College graduate	321 (28.8)	15 (18.1)	370 (28.3)
Time spent in physical activity per week, no. (%)			
<30 min	260 (26.4)	14 (20.6)	348 (26.6)
30 min to 2.5 h	460 (46.7)	32 (47.1)	619 (47.3)
>2.5 h[d]	266 (27.0)	22 (32.4)	340 (26.0)
Program participation			
Enrolled in CDSMP, no. (%)	702 (58.6)	58 (68.2)	954 (59.6)
Enrolled in DSMP, no. (%)	496 (41.4)	27 (31.8)	646 (40.4)
Mean no. (SD) of sessions attended	4.4 (1.7)	4.9 (1.1)	4.2 (1.9)
Attended ≥4 sessions, no. (%)	890 (75.6)	76 (89.4)	1,106 (71.1)
Distance traveled from participant's residence to workshop site, median (IQR), mile	2.1 (0.4–5.0)	1.3 (0.5–9.2)	2.0 (0.4–5.2)
Distance traveled from participant's residence to workshop site, mean (SD), mile	4.1 (8.1)	4.9 (5.7)	4.9 (6.1)
Class size, mean (SD)	16.5 (7.8)	11.5 (3.6)	16.0 (7.3)
Provides care to someone with a long-term health problem or disability, no. (%)	319 (28.9)	16 (19.5)	370 (28.6)

Abbreviations: CDSMP, Chronic Disease Self-Management Program; DSMP, Diabetes Self-Management Program; GED, general educational development certificate; IQR, interquartile range; SD, standard deviation.

[a] Geographic information was available for 1,295 of 1,638 participants; only these 1,295 participants were classified as living in a rural or an urban county and comprised our analytic sample. Urban and rural classifications were determined by using participants' home address and criteria from the US Census Bureau's Office of Management and Budget (13,14).

[b] "All" participants refers to all participants in CDSMP or DSMP: the 1,295 for whom geographic information was available (the analytic sample), plus the 343 participants for whom geographic information was not available.

[c] Not all numbers in categories add to number in column head because not all participants answered all questions. Percentages in each category sum to 100% (unless because of rounding they do not) and are based on number of participants who answered the question.

[d] Satisfies current physical activity recommendations per US guidelines (≥150 min/wk [18]).

Table 2. Linear Regression Coefficients of Variables Associated With Number of Sessions Attended Among Adults Aged ≥50 in the CDSMP and DSMP, Illinois, 2016–2017[a]

Variable	B (Standard Error)	β (95% Confidence Interval) [P Value]
Sex		
Male	−0.13 (0.14)	−0.03 (−0.40 to 0.13) [.32]
Female		Reference
Race		
Non-Hispanic white	−0.27 (0.14)	−0.09 (−0.55 to 0.01) [.055]
Non-Hispanic black, Hispanic, or other		Reference
Education		
<High school diploma	−0.49 (0.19)	−0.09 (−0.86 to −0.11) [.01]
≥High school diploma		Reference
Physical activity recommendations[b]		
Satisfies	0.07 (0.12)	0.02 (−0.17 to 0.31) [.57]
Does not satisfy		Reference
Distance traveled from participant's residence to workshop site		
<0.1 mile	0.32 (0.16)	0.07 (0 to 0.64) [.05]
≥0.1 mile		Reference
Class size	0.02 (0.01)	0.11 (0.01 to 0.04) [.01]
Type of program		
DSMP	0.02 (0.13)	0.01 (−0.24 to 0.29) [.87]
CDSMP		Reference
Classification of participant's county of residence		
Urban	0.16 (0.30)	0.03 (−0.43 to 0.74) [.60]
Rural		Reference
No. of miles traveled from participant's home to workshop site	−0.03 (0.01)	−0.12 (−0.05 to −0.01) [.001]
Distance traveled × urban or rural county of residence[c]	0.07 (0.04)	0.09 (0 to 0.14) [.049]

Abbreviations: CDSMP, Chronic Disease Self-Management Program; DSMP, Diabetes Self-Management Program.
[a] Attendance information for each participant in the program was provided by facilitators of the CDSMP and DSMP programs. The number of attended sessions ranged from 1 to 6. The overall model explained 4.8% of the variance in the number of sessions attended (R = 0.24; Adjusted R^2 = 0.048; ΔR^2 = 0.004; P = .049).
[b] Per US guidelines (≥150 min per week [18]).
[c] Urban and rural classifications were determined by using participants' home address and criteria from the US Census Bureau's Office of Management and Budget (13,14).

Table 3. Logistic Regression Coefficients of the Odds of Program Completion of the CDSMP and DSMP Among Rural and Urban Adults Aged ≥50 in Illinois, 2016–2017[a]

Variable	Odds Ratio (95% Confidence Interval) [P Value]
Sex	
Male	0.81 (0.54–1.19) [.28]
Female	1 [Reference]
Race	
Non-Hispanic white	0.75 (0.49–1.16) [.20]
Non-Hispanic black, Hispanic, or other	1 [Reference]
Education	
<High school diploma	0.54 (0.31–0.92) [.02]
≥High school diploma	1 [Reference]
Physical activity recommendations[b]	
Satisfies	1.26 (0.86–1.85) [.24]
Does not satisfy	1 [Reference]
Distance traveled from participant's residence to workshop site	
<0.1 mile	1.69 (1.02–2.81) [.04]
≥0.1 mile	1 [Reference]
Class size	1.04 (1.01–1.08) [.007]
Type of program	
DSMP	1.27 (0.84–1.94) [.26]
CDSMP	1 [Reference]
Classification of participant's county of residence[c]	
Urban	0.86 (0.31–2.40) [.78]
Rural	1 [Reference]
No. of miles traveled from participant's home to workshop site	0.96 (0.92–0.99) [.008]
Distance traveled × urban or rural county of residence[c]	1.20 (0.99–1.47) [.07]

Abbreviations: CDSMP, Chronic Disease Self-Management Program; DSMP, Diabetes Self-Management Program.

[a] The number of attended sessions ranged from 1 to 6. Attendance at ≥4 of 6 sessions is considered program completion by the program developers (7). The overall logistic regression model explained 9% of the variance in odds of completing the program (R = 0.06, Nagelkerke-adjusted R^2 = .090).

[b] Per US guidelines (≥150 min per week [18]).

[c] Urban and rural classifications were determined by using participants' home address and criteria from the US Census Bureau's Office of Management and Budget (13,14).

GIS SNAPSHOTS

Ensuring the Safety of Chronically Ill Veterans Enrolled in Home-Based Primary Care

Judith Katzburg, PhD, MPH, RN[1]; Debra Wilson, RN, BSN[2]; Jacqueline Fickel, PhD[3]; Jason D. Lind, PhD, MPH[4];
Diane Cowper-Ripley, PhD[5]; Marguerite Fleming, MPA[6]; Michael K. Ong, MD, PhD[3,7,8]; Alicia A Bergman, PhD[3];
Sarah E. Bradley, MA[4]; Sarah A. Tubbesing, MD, MSc[7,9]

PEER REVIEWED
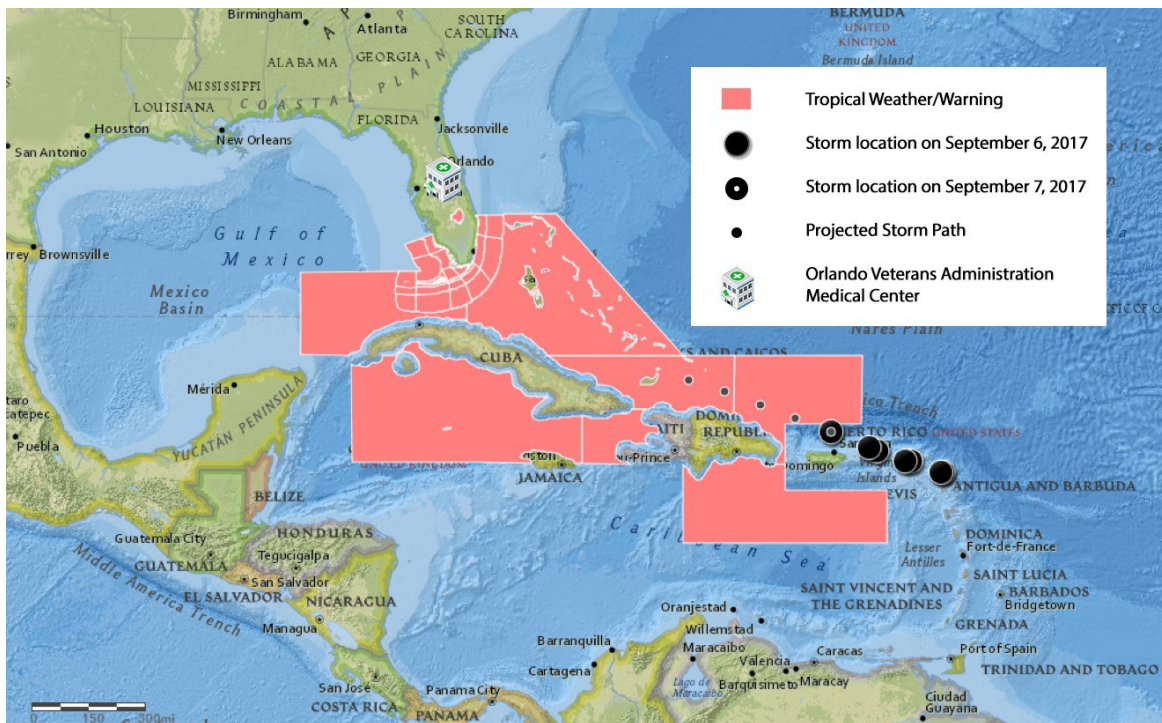
Figure 1. Hurricane Irma approaching Florida coastline, September 7, 2017, based on data collected September 6 and September 7, 2017. The registered nurse, a patient care manager, also served as geographic information system mapmaker (RNCM/mapmaker) for the Orlando Veterans Health Administration Home Based Primary Care program (OVAMC-HBPC), tracking the path of Hurricane Irma. Irma made landfall in the Florida Keys as a Category 4 hurricane with 132 mph winds. This powerful image of the looming threat helped inform the nurse manager, who supervised the OVAMC–HBPC nursing staff, of the severity of the storm. The RNCM/mapmaker also used the maps, in combination with patient information and other data, to educate and manage her patients. Map source: Portal for ArcGIS version 10.5 (2017), created for the Veterans Health Administration by Environmental Systems Research Institute (Esri). Additional sources: National Geographic, Environmental Systems Research Institute, Garmin, HERE Technologies, United Nations Environment World Conservation Monitoring Center, United States Geological Survey, National Aeronautics and Space Administration, European Space Agency, Micro Engineering Tech Inc., Natural Resources Canada, General Bathymetric Chart of the Oceans, National Oceanic and Atmospheric Administration, Increment PCorporation.

Figure 2. Oxygen-dependent and ventilator-dependent patients in home-based primary care, September 7, 2017. In preparation for Hurricane Irma, the nurse care manager, serving as the geographic information system mapmaker for the Orlando Veterans Health Administration Home Based Primary Care (OVAMC–HBPC) program, made maps for program leadership, including this map of oxygen-dependent and ventilator-dependent veterans. Leadership used these types of maps together with other clinical and care manager information in a dynamic process to make decisions regarding patient management in preparation for the storm. Map source: Portal for ArcGIS version 10.5 (2017) created for the Veterans Health Administration by Environmental Systems Research Institute. Additional Sources: Earthstar Geographics LLC, Environmental Systems Research Institute, HERE Technologies, Garmin.

Background

Geographic information system (GIS) maps can be used effectively for emergency planning and response (1). Vulnerable populations, especially chronically ill older people and those dependent on medical equipment for survival, might be at particular risk during disasters (2). The use of GIS maps to plan for and respond to emergencies is becoming an important strategy for ensuring the safety of chronically ill patients (1,3,4). The Veterans Health Administration Home Based Primary Care program (VHA-HBPC) has been demonstrating the innovative use of GIS mapping for practice and patient care management through a quality improvement project, the HBPC-GIS mapping project, which is currently disseminated to 30 geographically diverse VHA-HBPC sites nationwide.

The VHA-HBPC program was designed to serve veterans with complex chronic disease (5). Home-based primary care consists of an interdisciplinary team of clinicians who provide ongoing primary care in the patient's home (6). Veterans enrolled in VHA-HBPC are a vulnerable population, averaging more than 8 chronic conditions per patient (5). Currently, approximately 140 VHA-HBPC programs nationwide serve almost 38,000 veterans (personal communication, D. Davis, July 5, 2018).

To enhance practice management, the mapping project trains staff members at 30 VHA-HBPC programs to use VHA's Portal for ArcGIS mapping software, version 10.5 (Esri). Self-paced, online computer-based training modules usually require several hours, with ongoing training thereafter. This novel project was designed so that any member of the VHA-HBPC staff, including frontline staff members providing direct patient care, could make maps tailored to their local program's needs. As the mapping project expanded, evaluations indicated increasing use of GIS mapping for both emergency preparedness and response.

In 2017, some mapping project sites were adversely affected by disasters that inflicted historic costs in terms of human suffering and fiscal impact (7). For example, following Hurricane Irma, excessive heat and power outages accounted for a sizable percentage of deaths in the general population, including many elderly chronically ill patients (8). Below, we describe a case study that illustrates the innovative use of GIS maps by the Orlando Veterans Administration Medical Center HBPC Program (OVAMC-HBPC) leadership and a frontline clinical care provider to support the emergency management of patients.

Data Sources and Map Logistics

OVAMC-HBPC joined the mapping project in 2015; a nurse care manager trained as the mapmaker (RNCM/mapmaker). Maps were created by using Portal for ArcGIS software, version 10.5 (Esri), and the RNCM/mapmaker supplied patient information. The RNCM/mapmaker incorporated several types of patient data in the map (Box), which is viewable in a popup box on the map when the cursor is moved over patient locations. Layers were added to the map indicating location of emergency services (eg, hospital). Environmental threats could be identified by additional layers (eg, hurricane path, storm surges). Event-related map layers were obtained from open sources such as the National Oceanic and Atmospheric Administration.

Box. Patient Data Available for Incorporation into GIS Maps

Veteran's name	Emergency priority rating
Last 4 digits of Social Security number	Medical foster home
Physical address	On dialysis
Zip code	On ventilator
County	Mental health provider (name)
Geocode address	Occupational or physical therapist (name)
Phone number	Dietician (name)
RN case manager (name)	Social worker (name)
Provider (Name)	Registered for special needs shelter

Patient acuity score	Lives alone
Uses oxygen	Electricity dependent

Highlights

The map showing Hurricane Irma's path as it approached the tip of Florida crystallized the enormity of the impending threat for the nurse manager who supervised OVAMC-HBPC nursing staff. As the storm approached, the nurse manager requested a map that identified 2 groups who might be at particular risk during a power outage: oxygen-dependent and ventilator-dependent patients.

Action

OVAMC–HBPC had 364 veterans enrolled in September 2017. In preparation for Hurricane Irma's landfall in Florida, the RNCM/mapmaker frequently checked the path of the storm in Portal. The map of the oncoming storm was a powerful tool. As the nurse manager reported, "The map made me realize that it was real and it was going to come." The RNCM/mapmaker provided requested maps to the OVAMC-HBPC program director and nurse manager, including maps showing the locations of vulnerable patients, such as oxygen-dependent and ventilator-dependent patients and patients near the coast. Maps facilitated clear and secure communication between the mapmaker and program leaders. Maps of patient locations, the storm path, and other clinical and care manager information were used by leadership in a dynamic process to make decisions regarding patient management in preparation for the storm.

As the hurricane approached, the RNCM/mapmaker used Portal to improve the quality of the care management she provided to her patients. The RNCM/mapmaker synthesized information from the GIS maps and other sources regarding the storm's path, wind force, patient location and level of vulnerability, and areas with high likelihood of power outages. For example, with this knowledge, she effectively facilitated the sheltering-in-place of a patient with brittle diabetes by educating the patient's daughter of the impending risk. The daughter purchased a generator to power an air conditioning unit for the patient's room and a small refrigerator to keep his insulin cool. The RNCM/mapmaker also worked with the family of a patient diagnosed with chronic obstructive pulmonary disease and congestive heart failure who required oxygen. She convinced the family of the need for evacuation to the OVAMC hospital on the basis of the patient's vulnerabilities identified in information from the GIS maps and from other sources.

OVAMC facilitated the transport of 23 VHA-HBPC patients to its hospital, including 2 who required admission to the intensive care unit. Because of the advanced planning of the OVAMC-HBPC and, in part, their use of GIS to integrate and analyze environmental and clinical information, fewer than 7% of their patients (23 of 364) needed to be sheltered at the hospital. No OVAMC-HBPC patient deaths or injuries were attributed to the hurricane.

In review, GIS maps in conjunction with other data informed OVAMC-HBPC leaders and facilitated care management of patients with multiple chronic diseases who possibly required emergency management before the Florida landfall of Hurricane Irma. In post-disaster assessment, the nurse manager found value in using the GIS maps and believes that maps might assist VHA employees tasked with transporting patients in future disasters (eg, by locating patients who could be evacuated together). This case study demonstrates how the use of GIS maps in emergency planning had significant benefits for patients with complex chronic conditions who receive primary medical care at home. The feasibility of having local public health departments and other home care programs provide GIS training for frontline staff in emergency management of patients is worthy of consideration.

Acknowledgments

Author Information

Corresponding Author: Judith R. Katzburg, PhD, MPH, RN, Health Services Researcher, VA Greater Los Angeles Healthcare System, 16111 Plummer St, Mail code 152, North Hills, CA 91343. Telephone: 805-777-7856. Email: Judith.Katzburg@va.gov or jkatzburg@gmail.com.

Author Affiliations: [1]VA Health Services Research and Development, VA Greater Los Angeles Healthcare System, Los Angeles, California. [2]Orlando VA Medical Center Home Based Primary Care Program, Orlando, Florida. [3]VA Health Services Research and Development Center for the Study of Healthcare

Innovation, Implementation, and Policy, VA Greater Los Angeles Healthcare System, Los Angeles, California. [4]James A. Haley Veterans' Hospital and Clinics, Research Section, Tampa, Florida. [5]VA North Florida/South Georgia Veterans Health System, Health Services Research and Development; VA Office of Rural Health, GeoSpatial Outcomes Division, Gainesville, Florida. [6]Veterans Health Administration, VA Office of Reporting, Analytics, Performance, Improvement, and Deployment, Washington, District of Columbia. [7]David Geffen School of Medicine at UCLA, Department of Medicine, Los Angeles, California. [8]Department of Health Policy and Management, UCLA Fielding School of Public Health, Los Angeles, California. [9]VA Greater Los Angeles Healthcare System, Geriatrics and Extended Care, Los Angeles, California.

References

1. Aldrich N, Benson WF. Disaster preparedness and the chronic disease needs of vulnerable older adults. Prev Chronic Dis 2008;5(1):A27.

2. Dostal PJ. Vulnerability of urban homebound older adults in disasters: a survey of evacuation preparedness. Disaster Med Public Health Prep 2015;9(3):301–6.

3. DeSalvo K, Lurie N, Finne K, Worrall C, Bogdanov A, Dinkler A, et al. Using Medicare data to identify individuals who are electricity dependent to improve disaster preparedness and response. Am J Public Health 2014;7(104):1160–4.

4. Douglas R, Kocatepe A, Barrett AE, Ozguyen EE, Gumber C. Evacuating people and their pets: older Floridians' need for and proximity to pet-friendly shelters. J Gerontol B Psychol Sci Soc Sci 2017;00(00):1–9.

5. Beales JL, Edes T. Veteran's Affairs Home Based Primary Care. Clin Geriatr Med 2009;25(1):149–54, viii–ix.

6. Edes T, Kinosian B, Vuckovic NH, Nichols LO, Becker MM, Hossain M. Better access, quality, and cost for clinically complex veterans with home-based primary care. J Am Geriatr Soc 2014;62(10):1954–61.

7. Smith AB. 2017 U.S. billion-dollar weather and climate disasters: a historic year in context. National Oceanic and Atmospheric Administration (NOAA); 2018Jan; https://www.climate.gov/news-features/blogs/beyond-data/2017-us-billion-dollar-weather-and-climate-disasters-historic-year. Accessed July 31, 2018.

8. Issa A, Ramadugu K, Mulay P, Hamilton J, Siegel V, Harrison C, et al. Deaths related to Hurricane Irma — Florida, Georgia, and North Carolina, September 4–October 10, 2017. MMWR Morb Mortal Wkly Rep 2018;67(30):829–32.

ESSAY

Application of Geographic Information Systems to Address Chronic Disease Priorities: Experiences in State and Local Health Departments

Ian Brissette, PhD[1]; Michele Casper, PhD[2]; Sara L. Huston, PhD[3]; Melita Jordan, CNM, MSN, APRN C, CPM[4]; Becky Karns, MPH[5]; Christopher Kippes, MS[5]; Michael R. Kramer, PhD[6]; James M. Peacock, PhD, MPH[7]; Adam S. Vaughan, PhD, MPH, MS[2]

Summary

What is already known about this topic?

Health departments are keenly aware of the importance of local-level data to effectively and efficiently reduce the burden of chronic disease. We asked 4 state and local health departments about their experiences using GIS to address chronic disease priorities.

What is added by this report?

These responses reveal the extent to which maps and spatial analyses help to 1) document the geographic patterns of chronic disease, 2) inform resource allocation and policy, 3) develop culturally competent programs, and 4) assist with program planning, monitoring, and evaluation.

What are the implications for public health practice?

The continued and enhanced application of GIS to chronic disease surveillance, prevention, and treatment priorities can provide valuable benefits to both health departments and the communities they serve.

Introduction

Health department staffs are keenly aware of the importance of local-level data to effectively and efficiently reduce the burden of chronic disease (1). These local data — including data on disease burden, demographic factors, socio-environmental conditions, risk factors, and health care facilities — may be generated, analyzed, and mapped using geographic information systems (GIS), providing health department leadership and staff members with valuable

information. This application of GIS (and the underlying capacity of health departments to perform GIS work) allows health departments to better incorporate chronic disease prevention activities into the places where people live, work, and play (1–3).

Given the value of applied GIS, we invited staff members from 1 local and 3 state health departments to describe their use of GIS to address chronic disease priorities. These health departments previously participated in the Building GIS Capacity for Chronic Disease Surveillance training, which is provided through a collaboration of the Centers for Disease Control and Prevention (CDC), the National Association of Chronic Disease Directors, and the Children's Environmental Health Initiative (CEHI) (currently at Rice University) (2). The prompts given to the health departments were:

1. Please provide a brief description of how your health department currently uses GIS to address chronic disease priorities.
2. Please describe the benefits and challenges of using GIS to address the chronic disease priorities in your health department.
3. Please discuss how the use of GIS has enhanced your health department's ability to perform one or more of the Public Health Foundation's Core Competencies for Public Health Professionals (4) or the CDC/Council of State and Territorial Epidemiologists Applied Epidemiology Competencies (5). These competencies provide health departments with a definition of applied epidemiology and provide a framework within which to develop a public health workforce and to enhance the health of their communities.

The health departments' answers to these prompts reveal the extent to which "Where" has become a key component in their chronic disease work. These health departments have extended basic GIS skills and maps into a more robust GIS infrastructure that supports the integration of GIS into many facets of their chronic disease programs. Critically, these health departments describe the value of using maps and spatial analyses to communicate the bur-

den of disease; to inform decisions about resource allocation and policy, and to determine priority communities for intervention efforts; to develop culturally competent programs; and to assist with program planning, monitoring, and evaluation. Furthermore, health departments have experienced the benefits of using maps to communicate with diverse audiences, both within the health department and to partners, policy makers, and the public. Communicating disparities in the geographic patterns of chronic disease to the public enhances and improves community engagement. Finally, the health departments' responses reveal that GIS is an important tool to support internal cross-disciplinary workgroups within an agency, to create a stronger workforce for public health programming, and to further each agency's core mission for chronic disease prevention and management. As demonstrated in the following examples, GIS facilitates additional intersections across many of the public health and applied epidemiology core competencies, multiplying the benefits to the health departments and the communities they serve.

Cuyahoga County, Ohio, Board of Health

Question 1: How is your health department using GIS to address chronic disease priorities?

The Cuyahoga County Board of Health (CCBH) participated in CDC's Building GIS Capacity for Chronic Disease Surveillance training in 2014. Since that time, CCBH continues to expand and improve its GIS capacity (6). Our agency works in collaboration with the Cuyahoga County Planning Commission on a CDC-funded grant called Creating Healthy Communities (CHC) (7), administered by the Ohio Department of Health. CHC aims to reduce leading causes of death by increasing access to healthy foods, active living, and healthy eating; reducing smoking; and reducing childhood obesity. One outcome of this partnership was the creation of a map depicting both supermarket access and chronic disease health indicators, which was used to highlight food deserts and to examine the potential impact of future stores on neighborhood health.

CCBH uses GIS in grant applications and reporting. CCBH has been funded for CDC's Racial and Ethnic Approaches to Community Health program (REACH) (8). In response to the project's focus on enhancing access to healthy foods, physical activity, and chronic disease management in communities with the greatest need, we produced a map showing locations of community facilities with shared-use agreements for chronic disease self-management workshops and active living activities (Figure 1). The maps also indicated populations living near those facilities, defined by using half-mile buffer rings around those facilities. This map then allowed us to identify nearby populations that could potentially gain access to these facilities and programs. Agreements with

these facilities increased access to resources for an estimated 140,838 people who previously did not have access.
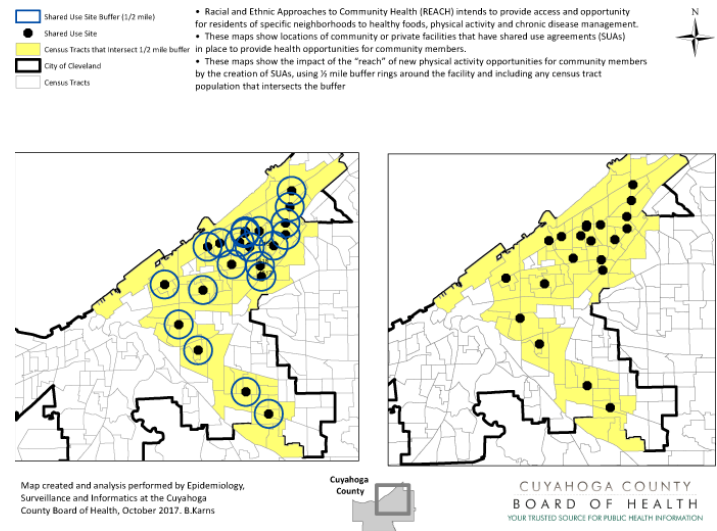

Figure 1. GIS map generated by Cuyahoga County Health Department to map locations of community facilities with shared-use agreements for chronic disease self-management workshops and active living activities and populations within a half mile of those facilities.

Our agency also produces cancer surveillance reports. In these reports, we analyze data for 24 types of cancer at the community level. Documenting the geographic distribution and burden of cancer helps prevention and control specialists make informed decisions and provides community-specific resources and education (9).

The Ohio Department of Health's Healthy Homes Lead Poisoning Prevention Program provides data to CCBH to monitor blood lead levels in children. We publish annual maps of blood lead surveillance data by neighborhood and municipality (10).

CCBH provides data to a local open-data platform, Health Data Matters, where users can create custom maps of health indicators (such as chronic disease and childhood lead poisoning) (11). In addition, local communities in our jurisdiction use our GIS products in their master plans to explore the health of their community, including chronic disease death rates, life expectancy, and access to grocery stores (12).

Question 2: What are the benefits and challenges of using GIS to address the chronic disease priorities in your health department?

CCBH confronts some challenges when using GIS to address chronic disease: creating standardized geographic databases for consistency throughout the agency, maintaining the best and most current geographic boundary data, the expense of GIS software, time and resource allocation for staff training, and finding best practices around GIS and chronic disease mapping. Despite these challenges, our agency has formally considered GIS-based activities to be an agency-wide asset and priority by identifying GIS in the agency strategic plan, forming an agency GIS workgroup, creating a map library, and implementing a formal electronic request process. Maps are also an invaluable way to help tell a story quickly to many audiences and to advocate for specific interventions.

Question 3: How is your health department using GIS to enhance the Core Competencies for Public Health?

Analytical/assessment skills: CCBH recognizes the importance of applying GIS analytical skills, with staff members continuing to sharpen their skills. CCBH now has staff members, including data analysts, field staff, information specialists, and supervisors, who are part of an internal workgroup providing GIS services to the agency. The goal of the workgroup is to establish an agency-wide, location-based approach to data collection and reporting. Members of the workgroup share knowledge and skills to deepen current GIS capacity and encourage open participation from all staff members interested in broadening their GIS work.

Policy development/program planning skills: CCBH staff members use GIS for program planning and communication to advocate for interventions to community partners, stakeholders, and health care systems. The CCBH staff is often invited to speak on areas of need and uses GIS to support how communities could benefit from various interventions (eg, a new labor and delivery hospital).

Cultural competency skills: Our agency uses GIS to document geographic patterns of social determinants of health including redlining, poverty, and education, along with their geographic overlap with health outcomes, particularly on chronic diseases and infant mortality. For example, GIS was used to identify target census tracts for the REACH grant. CCBH then helped identify residents in those neighborhoods to become community health ambassadors to provide context and develop messaging for their own unique neighborhoods (13).

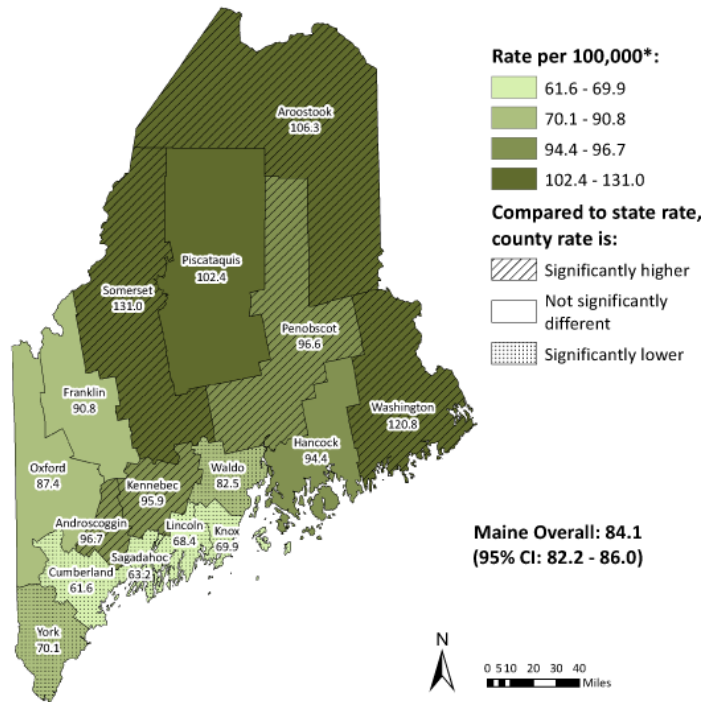As in many places, in Cuyahoga County and the City of Cleveland poor health outcomes and social determinants of health exhibit similar geographic patterns. The connections between patterns of poverty, race, and child deaths have been illustrated in our annual Child Fatality Review report (14).

Maine Center for Disease Control and Prevention

Question 1: How is your health department using GIS to address chronic disease priorities?

Since participating in CDC's Building GIS Capacity for Chronic Disease Surveillance training in 2011 (15), chronic disease epidemiologists and program staff members of the Maine Center for Disease Control and Prevention have incorporated GIS into routine business across all chronic disease programs. This process began slowly with a few basic maps. Now, as a standard part of epidemiologic work in our chronic disease programs, we have systematized map-making for local chronic disease measures and surveillance indicators. Maps can also show intervention sites, local policies, health care resources, or other related information. These maps are created using standard ArcGIS templates (Esri), and are incorporated into routine public and internal products such as the Maine Cancer Registry's Annual Report (16) and the State Public Health Actions (1305 cooperative agreement) program's standard epidemiology figures (Figure 2) (17). Our standard map templates have evolved and now include the state rates and 95% confidence intervals and identify counties with significantly higher or lower rates than the state overall. These maps are used to identify areas of the state with high burden, need, and opportunities for intervention.

Coronary Heart Disease Death Rate, by County, Maine, 2012-2016

Rate per 100,000*:
- 61.6 - 69.9
- 70.1 - 90.8
- 94.4 - 96.7
- 102.4 - 131.0

Compared to state rate, county rate is:
- Significantly higher
- Not significantly different
- Significantly lower

Maine Overall: 84.1 (95% CI: 82.2 - 86.0)

Coronary heart disease deaths are defined as deaths by county of residence in which the underlying cause of death was coded as ICD-10 codes I20-I25.
Data Source: Maine Mortality Database; Data, Research, and Vital Statistics Program, Maine CDC.
*Age-adjusted rates are deaths per 100,000 population age-adjusted to the U.S. 2000 standard population.
95% CI: 95% confidence interval of the rate.
Statistical significance was determined by comparing confidence intervals and non-overlapping confidence intervals were considered to be statistically different.
Prevalence estimates mapped using the quantiles method with four categories.
Map created by Caitlin Pizzonia, May 2018.

Figure 2. GIS map generated by the Maine Center for Disease Control and Prevention to show distribution of coronary heart disease death rates by county. Abbreviation: CI, confidence interval.

We have also begun creating interactive mapping tools using Instant Atlas (Esri) and, more recently, ArcGIS Online. We have 3 Instant Atlas products publicly available for our Maine Integrated Youth Health Survey (18), the State Public Health Actions (1305 cooperative agreement) program (17), and our Division of Disease Prevention Key Surveillance Indicators (covering chronic disease, maternal and child health, and injury) (19). These products allow users to choose the indicators and years of data, toggle between county and public health district, manipulate map characteristics, and more. We are also developing an ArcGIS Online WebApp for programmatic decision-making around blood pressure medication adherence. This product includes county-level blood pressure medication adherence and blood pressure control rates, locations of pharmacies and health care practices (including

those already involved in health department–led interventions), and demographic data, such as population density. The WebApp allows the program staff to use customized maps to make programmatic decisions, such as which pharmacies or practices may be high priorities for programmatic involvement.

Question 2: What are the benefits and challenges of using GIS to address the chronic disease priorities in your health department?

The biggest benefit of GIS is that almost everyone loves maps, and maps are a great communication tool. Maps make very clear the areas of the state with greatest burden and need. Well-designed maps translate complex data into understandable information for public health action in a way that few other tools do. GIS products also promote discussion and collaboration among any group of people and lead to more questions and curiosity about public health issues. GIS products also help communicate to health department leadership the value of surveillance and epidemiology.

Our major challenge has been developing and retaining staff GIS expertise. Even today, not all newly minted MPH epidemiologists have GIS skills. The CEHI trainings and materials have helped us enormously with training new epidemiologists (2). Our approach has been to train all our chronic disease epidemiologists in basic GIS skills to create quality maps using our standard templates. We also have several epidemiologists with advanced GIS skills who lead development of more innovative GIS products. Developing a relationship with our Maine Office of GIS has also helped us a great deal, particularly as we move into the world of ArcGIS Online.

Another challenge is the limitation of static maps. Our Maine asthma disparities map, for instance, shows the most recent county-level asthma emergency department and hospitalization rates along with asthma program intervention sites, showing alignment between burden and intervention coverage areas. However, rates and programs change over time and these changes are extremely difficult to show clearly on a static map. As described above, we are developing interactive maps to overcome this limitation.

Question 3: How is your health department using GIS to enhance the Core Competencies for Public Health?

Analytical/assessment skills: Incorporating GIS into our routine work has enhanced our surveillance and epidemiology work by ensuring we are always asking and answering the critical question of "Where?" Mapping and spatial analysis have been key skills of epidemiology and public health since John Snow's investigation of cholera in 19th century London (20). Developing our GIS ex-

pertise has given us 21st-century tools for these key skills. All epidemiologists should have GIS in their toolbox.

Communication skills: A picture is worth a thousand words. A well-crafted map communicates patterns of disease and risk factors far more quickly and clearly than a data table can. Well-crafted maps lead to increased understanding among public health program staff members, health department leadership, and the public.

Leadership and systems thinking skills: Our work in GIS demonstrates our commitment to workforce development and to developing and enacting chronic disease epidemiology efforts that support our health agency's mission. When hiring new epidemiologists we look for GIS skills, and we have worked to develop GIS skills in our current team of chronic disease epidemiologists. Health department leadership buy-in is needed to ensure staff time for GIS work and investment in training. Creating even a small number of basic maps on key issues can help develop that buy-in.

New Jersey Department of Health

Question 1: How is your health department using GIS to address chronic disease priorities?

GIS and maps have allowed the New Jersey Department of Health (NJDOH) programs to make informed decisions on financial planning, policy development, program planning, and resource allocation regarding chronic disease. Using GIS to produce maps provides a mechanism by which public health professionals can visualize disease trends and the conditions that could be affecting trends in specific geographic areas — where we live, work, and play.

Financial planning: For program year 2019, NJDOH used maps to make funding decisions. We created maps showing disease burden, provider locations, demographic information (race, sex, education level), and causes of death. These maps have been especially helpful to the Heart Disease and Stroke Prevention Program where models were developed to assist New Jersey–based health care organizations in meeting nationally recognized best practices and standards to prevent and treat heart disease and stroke. The Office of Tobacco Control, Nutrition, and Fitness (OTCNF) used maps to identify members for the Tobacco Control Network, and to view utilization patterns for the New Jersey QuitLine. The New Jersey Cancer Education and Early Detection program and the Office of Cancer Control and Prevention (OCCP) have historically allocated state and federal funds to address the needs in all 21 New Jersey counties. However, as resources dwindle, we recognized the need to be more strategic in funding programs. Maps were instrumental in obtaining the support of leadership to reallocate and redirect funding for certain geographic regions.

Asset mapping: NJDOH uses GIS to identify assets that either impede or facilitate residents' abilities to take advantage of opportunities for screening for breast, cervical, prostate, and colorectal cancer or to complete follow-up care. We mapped resources such as transportation, medical specialists, and other services that coincide with or are required for complete patient care. Asset mapping has also been very helpful in identifying potential partnerships at the community level. Those partnerships include faith-based organizations that were instrumental in delivering health messages to African American men, health care providers who deliver screening services, worksite wellness sites, and employer groups with at least 50 employees where screening, education, and promotion of wellness policies by the employer are needed.

Funding application development: NJDOH uses GIS to support funding applications by demonstrating the need and diversity of the state's residents. These maps inform the development of culturally appropriate programming and services for population groups that are affected by specific health concerns. For instance, the NJDOH application for CDC's WISEWOMAN grant proposal used maps to target specific populations and geographic locations. If funded, the maps will be instrumental in the implementation of the goals and objectives; if not funded, these maps have use across programs and will be a source of support and information for current and future projects.

Question 2: What are the benefits and challenges of using GIS to address the chronic disease priorities in your health department?

Benefits: Mapping chronic disease incidence, prevalence, and mortality rates, including cancer, has allowed NJDOH to develop an integrated, collaborative, and multidisciplinary plan for addressing these health concerns through strategic partnerships at the community level and within NJDOH. For example, links between cancer and obesity, poor eating habits, lack of exercise, and smoking have led to collaborative education and awareness campaigns between the OTCNF and the OCCP. The maps reflect where high prevalence of these often comorbid conditions overlap. Maps of point-of-sale audits and vendor failures have allowed community partners to see where their efforts were most needed for education and policy development. Importantly, the identification and collaboration of these partners prevented duplication of efforts.

Through the OCCP's partnership with the New Jersey State Cancer Registry (21), all available cancer data (by cancer site, sex, race, and ethnicity) is easily accessible and available to NJDOH staff, community partners, and other stakeholders. With these data, the registry develops county-level maps of cancer incidence and mortality data for the state of New Jersey.

Challenges: Availability of GIS training has been a challenge for NJDOH staff. Recently, the Integrated Health Services Branch, Division of Community Health Services, Community Health and Wellness Unit sponsored training for 16 employees across the Agency through CEHI on the development and application of maps using GIS. The Integrated Health Services Branch leadership is committed to continuing support for future GIS training of existing staff members and making this training a desired skill for new hires.

Recommendations: Senior staff members could identify staff members for GIS training and provide scheduled time to practice and advance this skill set. Where possible, allocate funds or seek funding for this purpose. Also, use free opportunities to learn and practice GIS mapping skills, such as the Health Resources and Services Administration's Uniform Data System mapper (22) and the Healthy City tool, which provides a community research laboratory toolkit on participatory asset mapping (23). Free online GIS training for both new and advanced users is available through the GIS Exchange (24).

Question 3: How is your health department using GIS to enhance the Core Competencies for Public Health?

Assessment: NJDOH and the OCCP use maps to identify needs and gaps and to monitor the health of residents. Assessment is an ongoing, continual process that must be undertaken to show effectiveness and the need for new policies and programs.

New York State Division of Chronic Disease Prevention

Question 1: How is your health department using GIS to address chronic disease priorities?

Over the past 8 years, the New York State Division of Chronic Disease Prevention (DCDP) has incorporated GIS into all phases of programs to address chronic disease. In program planning, maps are used to understand how the burdens of chronic disease and key risk factors, including social determinants of health, vary across communities. GIS analysis has aided in identifying specific communities in need of focused public health action (25,26). The resulting maps enabled us to communicate effectively with decision makers and engage communities. GIS has helped to visualize community assets, including certified mammography facilities, locations offering lifestyle change programs (26,27), and retail outlets offering healthy food (28), and to identify gaps and plan local action. In program monitoring, local partners provide information about community locations where evidence-based interventions are implemented (26,28). Using GIS to analyze and display these data enables stakeholders to understand the schools, worksites, childcare centers, hospitals, corner stores, and other community locations affected by these interventions. Maps of data

collected by DCDP-funded grantees demonstrate accountability to stakeholders and support performance management. In program evaluation, maps are used to assess whether policy, system, and environmental changes established by grantees occur in high-need communities and have the potential to affect a sizable population (26,28). In ongoing public health surveillance, maps of key chronic disease health and risk factor indicators based on time series data illustrate the potential impact of public health action and are used to justify and advocate for additional resources.

Question 2: What are the benefits and challenges of using GIS to address the chronic disease priorities in your health department?

One benefit of GIS is that maps have proven to be an effective way to communicate with many audiences about geographic disparities in chronic disease burden. Public health action is local, and many of our partners identify with a specific geographic area, be it a county, town, school district, or region. When health indicator data are displayed in maps, partners see their communities in the data reports. Maps make it easier to make the case for action at specific locations and to increase receptivity for evidence-based interventions that DCDP and our partners are promoting.

One challenge with GIS is that developing a map requires more resources than making a graph or data table. In terms of staffing resources, it has taken years to develop sufficient capacity to meet the GIS needs of the major programs within DCDP. Maintaining a staff with GIS experience has become an ongoing staff resource priority in our division. With regard to time resources, it has taken time to establish realistic expectations with partners about how long it will take for GIS projects to be completed. Partners unfamiliar with GIS often lack perspective on the skill and time involved in making maps and what appear visually to be simple changes.

A second challenge is that the effectiveness of a GIS project depends on the availability of quality, geo-referenced data. Mapping data that are inherently unreliable, come from unknown or nonvalidated data sources, or are collected at incongruent levels of geography often do not produce meaningful information. It has been a challenge to communicate with partners that their enthusiasm and good intentions will not compensate for the absence of reliable data. Fortunately, DCDCP's internal and external partners have developed an understanding of the data requirements for GIS projects.

Question 3: How is your health department using GIS to enhance the Core Competencies for Public Health?

Financial planning and management skills: A universal challenge public health practitioners face is determining how best to allocate limited resources in a jurisdiction while most effectively address-

ing the health needs of the population. In New York, we have used GIS to help address this challenge and support financial planning for funding programs in which DCDP issues requests for proposals. To distribute resources that support preventive services for breast and cervical cancer in New York State, service delivery data from each contractor were subject to GIS analysis. These res- ults helped support decision making on the total number of grantees needed to achieve statewide coverage and efficiently meet the demands for preventive cancer screening. To focus fund- ing for community interventions that promote healthy eating, physical activity, and breastfeeding, we used GIS to identify and display the high-need communities where eligible applicants were able to propose work.

Analytical/assessment skills and community dimensions of practice skills: GIS enables our division to incorporate geographic and environmental data into community assessment processes, and to display our findings in a manner that promotes community engagement. Therefore, maps have enhanced our analytical and assessment process by adding place as a key dimension and contributed to development of Community Dimensions of Practice Skills by increasing our capacity to involve community stakeholders and effectively advocate for public health action.

Conclusion

GIS has become a critical tool for state and local health departments and has allowed them to extend the concept of geography into many aspects of their chronic disease programs. These responses reveal the extent to which health departments are using maps and spatial analyses to 1) communicate the burden of disease; 2) inform decisions about resource allocation, policy, and priority communities for intervention efforts; 3) develop culturally competent programs; and 4) assist with program planning, monitoring, and evaluation. The continued and enhanced application of GIS to chronic disease surveillance, prevention, and treatment priorities can provide valuable benefits both to health departments and to the communities they serve.

Acknowledgments

Author Information

Corresponding Author: Michele Casper, PhD, Division for Heart Disease and Stroke Prevention, Centers for Disease Control and Prevention, 4770 Buford Hwy, Atlanta, GA, 30341. Telephone: 770-488-2571. Email: mcasper@cdc.gov.

Author Affiliations: [1]Bureau of Chronic Disease Evaluation and Research, New York State Department of Health, Albany, New York. [2]Division for Heart Disease and Stroke Prevention, Centers for Disease Control and Prevention, Atlanta, Georgia. [3]Maine Center for Disease Control and Prevention, Augusta, Maine, and University of Southern Maine, Portland, Maine. [4]Division of Community Health Services, New Jersey Department of Health, Trenton, New Jersey. [5]Epidemiology, Surveillance, and Informatics Services, Cuyahoga County Board of Health, Parma, Ohio. [6]Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, Georgia. [7]Cardiovascular Health Unit, Minnesota Department of Health, St. Paul, Minnesota.

References

1. National Committee on Vital and Health Statistics. Measuring health at the community level: data gaps and opportunities. Washington (DC): US Department of Health and Human Services; 2017.
2. Miranda ML, Casper M, Tootoo J, Schieb L. Putting chronic disease on the map: building GIS capacity in state and local health departments. Prev Chronic Dis 2013;10:E100.
3. Auchincloss AH, Gebreab SY, Mair C, Diez Roux AV. A review of spatial methods in epidemiology, 2000–2010. Annu Rev Public Health 2012;33(1):107–22.
4. Council on Linkages Between Academia and Public Health Practice. Core competencies for public health professionals. 2014. http://www.phf.org/resourcestools/Documents/Core_Competencies_for_Public_Health_Professionals_2014June.pdf. Accessed September 3, 2018.
5. Council of State and Territorial Epidemiologists. Applied epidemiology competencies: competencies for applied epidemiologists in governmental public health agencies. Vol. 1.2. 2008. http://www.cste2.org/webpdfs/AppliedEpiCompwcover.pdf. Accessed September 3, 2018

6. Children's Environmental Health Initiative. Mapping heart disease, stroke and other chronic diseases: a program to enhance GIS capacity within state and local health departments. Map highlights from California; Kansas; New Mexico; South Dakota;Vermont; Cuyahoga County, Ohio; Cleveland, Ohio; and Denver, Colorado. https://www.cdc.gov/dhdsp/docs/GIS_Training_highlights_v6.pdf. Accessed September 3, 2018.

7. Ohio Department of Health. Cuyahoga County supermarket assessment. http://www.ccbh.net/cuyahoga-county-supermarket-assessment/. Accessed September 3, 2018.

8. Cuyahoga County Board of Health. Health improvement partnership — Cuyahoga. http://hipcuyahoga.org/. Accessed September 3, 2018.

9. Cuyahoga County Board of Health. Cancer surveillance http://www.ccbh.net/cancer/. Accessed September 3, 2018.

10. Cuyahoga County Board of Health. Lead poisoning prevention. http://www.ccbh.net/lead-poisoning/. Accessed September 3, 2018.

11. Cleveland and Cuyahoga Health Resource. Health Data Matters https://www.healthdatamatters.org/. Accessed September 3, 2018.

12. Cuyahoga County Planning Commission. The City of Brecksville Master Plan: Current Conditions. http://www.countyplanning.us/projects/brecksville-master-plan/. Accessed September 3, 2018.

13. Prevention Research Center at Case Western Reserve University. Community health ambassadors. http://www.prchn.org/CommunityHealthAmbassadors.aspx. Accessed September 3, 2018.

14. The Cuyahoga County Child Fatality Review Board. Protecting our future: the Cuyahoga County child fatality report, twentieth edition, child fatalities 2016. Cleveland (OH): Cuyahoga County Planning Commission; 2016.

15. Children's Environmental Health Initiative. Mapping heart disease, stroke and other chronic diseases: a program to enhance GIS capacity within state and local health departments. Highlights from Idaho; Indiana; Louisiana; Maine; New York; Delta & Menominee, Michigan; and RiverStone, Montana. Ann Arbor (MI): University of Michigan; 2012.

16. Maine Center for Disease Control and Prevention. Maine annual cancer report 2015: 2012 cancer incidence, 2012 cancer mortality. https://www.maine.gov/dhhs/mecdc/public-health-systems/data-research/vital-records/mcr/reports/documents/Maine_Annual_Cancer_Report_2015_07112016.pdf. Accessed September 3, 2018.

17. Maine Center for Disease Control and Prevention. Prevention and control of diabetes, heart disease, and obesity in Maine. https://www.maine.gov/dhhs/mecdc/population-health/data/prevention-diabetes/atlas.html. Accessed September 3, 2018.

18. Maine Integrated Youth Health Survey. Maine's youth health data atlas. https://data.mainepublichealth.gov/miyhs/files/atlas/atlas.html. Accessed September 3, 2018.

19. Maine Center for Disease Control and Prevention. Maine's disease prevention dashboard. www.maine.gov/dhhs/mecdc/population-health/data/disease-prevention-dashboard/atlas.html. Accessed September 3, 2018.

20. Brody H, Rip MR, Vinten-Johansen P, Paneth N, Rachman S. Map-making and myth-making in Broad Street: the London cholera epidemic, 1854. Lancet 2000;356(9223):64–8.

21. New Jersey Department of Health. NJ State Cancer Registry https://www.state.nj.us/health/ces/reporting-entities/njscr/. Accessed September 3, 2018.

22. Health Resources and Services Administration. How to create a service area map and data table using the UDS mapper. Rockville (MD): Health Resources and Services Administration; 2018.

23. Healthy City. Community research http://www.healthycity.org/cbpar-toolbox/. Accessed September 3, 2018.

24. Centers for Disease Control and Prevention. Chronic disease GIS exchange. https://www.cdc.gov/dhdsp/maps/gisx/index.html. Accessed March 1, 2018.

25. New York State Department of Health Board of Tobacco Control. Prevalence of current smoking among adults in New York by county: NYS BRFSS 2016. StatShot 2018;11(4). https://www.health.ny.gov/prevention/tobacco_control/reports/statshots/volume9/n3_county_adult_smoking_prevalence.pdf. Accessed September 3, 2018.

26. Health Data NY. Creating healthy places child focused interventions and student weight status county map. https://health.data.ny.gov/Health/Creating-Healthy-Places-Child-Focused-Intervention/qshu-47dt. Accessed September 3,2018.

27. Ruberto RA, Brissette IF. Geographic access to diabetes prevention program sites: New York State Department of Health. Prev Chronic Dis 2014;11(1):E27.

28. New York State Department of Health. Creating healthy places to live, work, and play. Healthy schools New York. 2015. https://www.health.ny.gov/prevention/obesity/statistics_and_impact/docs/creating_health_places_to_live_work_play.pdf. Accessed September 3, 2018.