

Understanding County-Level, Cause-Specific Mortality

The Great Value—and Limitations—of Small Area Data

Cheryl R. Clark, MD, ScD; David R. Williams, PhD, MPH

In this issue of JAMA, Dwyer-Lindgren and colleagues¹ present advanced methods and applications of small area estimation techniques to produce county-level summary measures of cause-specific mortality rates across the United States and estimates of temporal trends in these rates.



Related article [page 2385](#)

The study used validated redistribution methods to recapture mortality data that would have been lost to so-called garbage coding, the practice of assigning potentially noninformative mechanisms of mortality (eg, cardiopulmonary arrest) rather than underlying disease codes (eg, congestive heart failure) to death certificates. The authors used enhanced generalized linear mixed-effects regression models to incorporate information on geographic spatial patterns, time and age associations, and relevant population-level covariates, to achieve valid cause-specific mortality rate estimates from the National Vital Statistics System (NVSS) without pooling data across years for counties with small sample sizes.

The scale and scope of the estimates presented in this study are novel. Among more than 80 million deaths analyzed from 1980 to 2014, the study found spatial and temporal patterns in county-level mortality rates that differed by cause of death. For example, cardiovascular disease rates declined slowly over the study period among counties in south-central US states between Oklahoma, Alabama, and Kentucky compared with other counties. Mortality from self-harm and interpersonal violence was patterned differently, with the highest county-level rates observed in Alaska, Native American reservations in North Dakota and South Dakota, and southwestern states.

The authors highlight the novel methods and suggest potential uses of these estimates, including by state and county health departments to develop specific policies and programs; by physicians to better understand the health concerns of the populations they serve; and by researchers to “identify counties that have done unexpectedly well or poorly with regard to a particular cause of death and that warrant additional study to identify factors driving these trends.”¹

But what information can these data and maps reveal about local needs or where to prioritize efforts for deeper study and intervention? For example, what insights could be drawn from data on county-level variation in cardiovascular disease mortality rates or similarly designed data on death rates from self-harm and interpersonal violence? How should these data inform considerations about priorities and strategies for resolving geographic health disparities? Before answering these ques-

tions, it is worth reviewing some of the limitations of small area analysis that should be considered when attempting to interpret these data.

The Modifiable Areal Unit Problem

Geographic analyses are subject to a technical consideration called the “modifiable areal unit problem.”² That is, population-level rates are a direct result of how boundaries are drawn around the populations of interest. The problem is not that a wrong or right geographic boundary exists for examining mortality rates. However, the boundaries drawn around area-level data shape the perceptions of risks, etiology, and potential policy solutions. Additionally, the geographic boundaries selected are often unintentionally—or preferably intentionally—shaped by the way hypotheses are framed about how environments contribute to rates of disease. Well-designed boundaries help in testing underlying etiologic hypotheses explicitly. Purely administrative boundaries may lead to missed opportunities to examine important environmental risk factors.

Currently, much area-level research in the United States examines county-level administrative boundaries as a manageable small area that exists across the United States where data are readily available and where local public health is thought to be delivered in many places in the country. But limitations in this approach must be considered when drawing conclusions about epidemiologic patterns and the policy importance of clusters or hot spots, or when concluding that ostensibly low-risk areas have fewer needs for intervention.

Consider the example of the county-level self-harm and interpersonal violence rates presented in Figure 5 in the article by Dwyer-Lindgren et al.¹ The high county-level rates in the southeast, in areas of Native American reservations within North Dakota and South Dakota, and in boroughs in Alaska argue convincingly for the need for additional research and intervention efforts in these specific areas. However, these data do not highlight the interpersonal violence in urban cities in the Midwest and east, which do not appear as hot spots in these data.³ If these findings are viewed uncritically, clinicians, public health officials, or national policy makers using these data may not identify or target these Midwestern and eastern urban areas—which might appear to be at low risk using the proposed methods applied at the county level—as having needs for intervention against interpersonal violence.

Framed differently, NVSS data have been quantified in a *Morbidity and Mortality Weekly Report (MMWR)* study that

sought specifically to identify the burden of firearm-related deaths and suicides in urban metropolitan statistical areas (MSAs).³ Through a method that disaggregated data on firearm-related homicide rates from suicide rates at the MSA level, these data identified several areas including Philadelphia, Pennsylvania; Newark, New Jersey; Detroit, Michigan; Baltimore, Maryland; Chicago, Illinois; and Oakland, California, as well as other areas with high firearm-related homicide rates. The *MMWR* data have limitations in terms of geographic coverage. Taken together with the county-level rates presented by Dwyer-Lindgren and colleagues, the known picture of urban and rural differences in cause-specific mortality from self-harm vs interpersonal violence mortality begins to emerge, and together, these data may be used more effectively to shape debates on needs for populations in different geographic boundaries along urban and rural dimensions.

Strategies to overcome the modifiable areal unit problem are used in public health practice. For example, analysts in King County, Washington, have shown that focusing on overall life expectancy only at the county level, which is higher than the national average, obscures the 30-year gap in life expectancy that is evident at the census tract level within the county.⁴ County officials analyze data at multiple levels of geography to guide health equity interventions.⁴ From a research policy perspective, the modifiable areal unit problem is an argument to make the individual-level data from NVSS available to designated researchers, with appropriate safeguards for confidentiality, to facilitate scientific investigations of the role of geography in the etiology of specific risks.

Place Heterogeneity

A second and related issue is that population-level mortality rates vary along demographic lines within a shared geography, and summary statistics at the county-level may mask inequalities experienced by smaller populations. In this way, a single county-level rate may give an inaccurate estimate of geographic risks and years of life lost for some groups. The issue of inequality in shared spaces is particularly salient for groups that experience structural stigma, defined as the exposure to societal-level conditions, cultural norms, and institutional practices that limit resources, threaten psychological or physical safety, and reduce well-being for affected groups.⁵

An example is found in the work on structural stigma directed against lesbian, gay, bisexual, and transgender populations. Hatzenbuehler et al⁶ examined data from the General Social Survey (GSS) linked to the National Death Index and found that after controlling for individual- and community-level risk factors, structural stigma was strongly associated with premature mortality among sexual minorities, with an average life expectancy loss of 12 years for sexual minorities who lived in geographic areas that rated highest on GSS measures of structural stigma. Several cause-specific mortality rates also were patterned by area-level structural stigma, including cardiovascular disease mortality and deaths due to homicide and suicide, for which sexual minorities in geographic areas with high stigma had an 18-year difference in the age at suicide compared with sexual minorities in geographic areas of low stigma (37.5 vs 55.7 years, respectively).⁶

A type II place heterogeneity arises when the mechanism of structural stigma includes residential segregation, such that minority and majority groups do not share the same spaces or health-promoting resources at a micro level. The lack of comparable exposure distributions complicates the interpretation of geographic patterns at the county level by obscuring the experience of disadvantaged groups at smaller levels of aggregation. This point is particularly important when assessing inequalities along the lines of race/ethnicity and geography in the United States. For example, Acevedo-Garcia et al⁷ examined US Census 2000 census tract data for 100 metropolitan areas covering 45 million children and found that, on average, the best census tract environments in which black or Latino children lived had higher average poverty levels and higher unemployment rates than the worst neighborhoods in which non-Hispanic white children lived.

A substantial body of literature documents this important source of heterogeneity by place for mortality outcomes and other risk factors.⁸⁻¹¹ To examine place heterogeneity, data should be presented as stratified by age (within similar groupings across the life course), race/ethnicity, sexual orientation and gender identity, socioeconomic status, or other population characteristics to identify important dimensions along which geographic inequalities may be different for diverse groups. Alternatively, Murray et al⁹ have developed creative strategies to explore heterogeneity, including constructing race-county units, which have generated important insights on racial and ethnic mortality differences according to multiple dimensions of race, socioeconomic position, and geography.

Need for Participatory Scientific Data Collection

Existing administrative data are rarely available to perform the comprehensive examination of area-level data needed to drive policy change. Thus, stakeholder engagement will be critical to refining the interpretation of cause-specific county-level mortality rates and suggesting additional data collection needed to deepen current understanding of these inequalities.

For example, the County Health Rankings & Roadmaps initiative of the Robert Wood Johnson Foundation and University of Wisconsin Population Health Institute leverages county-level data on health status and environmental conditions to invite community participation in investigating patterns and suggesting additional data sources needed to understand drivers of geographic health inequities.¹² A number of national and federal agencies, including the Patient-Centered Outcomes Institute, the Agency for Healthcare Research and Quality, the National Institute of Minority Health and Health Disparities, and the US Environmental Protection Agency, have dedicated funding, developed models of collaborative research, and supported infrastructure to encourage academic-community partnerships and other forms of stakeholder engagement in research, through which the innovative methods and county-level trends presented by Dwyer-Lindgren and colleagues may be additionally leveraged.

A potentially important next use of the data presented in this study may involve working with stakeholder groups to suggest data that may be linked to further investigate small-area mortality trends, including links with disease

registry data,¹³ social media data, survey data, and increasingly available data from biorepositories that may yield insights on social and biological mechanisms of environmental exposures.

In summary, Dwyer-Lindgren and colleagues have provided powerful tools with which to examine geographic in-

equalities in health. The careful analyst and policy maker should observe the limitations of these data, use participatory science to interpret these patterns, and find additional data to understand geographic trends relevant to diverse groups for the purpose of assigning priorities for further etiologic investigations and interventions.

ARTICLE INFORMATION

Author Affiliations: Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts (Clark); Department of Social and Behavioral Sciences, Harvard T.H. Chan School of Public Health, Boston, Massachusetts (Williams).

Corresponding Author: David R. Williams, PhD, MPH, Department of Social and Behavioral Sciences, Harvard T.H. Chan School of Public Health, 677 Huntington Ave, Kresge Bldg, Room 615, Boston, MA 02115 (dwilliam@hsph.harvard.edu).

Conflict of Interest Disclosures: Both authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest and none were reported.

REFERENCES

- Dwyer-Lindgren L, Bertozzi-Villa A, Stubbs RW, et al. US county-level trends in mortality rates for major causes of death, 1980-2014. *JAMA*. doi:10.1001/jama.2016.13645
- Cromley EK, McLafferty SL. *GIS and Public Health*. 2nd ed. New York, NY: Guilford Press; 2011.
- Centers for Disease Control and Prevention (CDC). Violence-related firearm deaths among residents of metropolitan areas and cities—United States, 2006-2007. *MMWR Morb Mortal Wkly Rep*. 2011;60(18):573-578.
- Assessment, Policy Development and Evaluation Unit, Public Health—Seattle and King County. Guide for Sub-County Assessment of Life Expectancy (SCALE). http://c.ymcdn.com/sites/www.cste.org/resource/resmgr/pdfs/pdfs2/Guide_for_Sub-County_Assessm.pdf. Accessed November 3, 2016.
- Link BG, Phelan JC. Conceptualizing stigma. *Annu Rev Sociol*. 2001;27:363-385. doi:10.1146/annurev.soc.27.1.363
- Hatzenbuehler ML, Bellatorre A, Lee Y, Finch BK, Muennig P, Fiscella K. Structural stigma and all-cause mortality in sexual minority populations. *Soc Sci Med*. 2014;103:33-41.
- Acevedo-Garcia D, Osypuk TL, McArdle N, Williams DR. Toward a policy-relevant analysis of geographic and racial/ethnic disparities in child health. *Health Aff (Millwood)*. 2008;27(2):321-333.
- Williams DR, Collins C. Racial residential segregation: a fundamental cause of racial disparities in health. *Public Health Rep*. 2001;116(5):404-416.
- Murray CJ, Kulkarni SC, Michaud C, et al. Eight Americas: investigating mortality disparities across races, counties, and race-counties in the United States. *PLoS Med*. 2006;3(9):e260.
- 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.
- Clark CR, Ommerborn MJ, Coull BA, Pham DQ, Haas JS. Income inequities and Medicaid expansion are related to racial and ethnic disparities in delayed or forgone care due to cost. *Med Care*. 2016;54(6):555-561.
- Robert Wood Johnson Foundation; University of Wisconsin Population Health Institute. County Health Rankings and Roadmaps. <http://www.countyhealthrankings.org/using-the-rankings-data>. Accessed August 31, 2016.
- Masoudi FA, Ponirakis A, Yeh RW, et al. Cardiovascular care facts: a report from the National Cardiovascular Data Registry: 2011. *J Am Coll Cardiol*. 2013;62(21):1931-1947.