

The Public Health Disparities Geocoding Project Monograph

Geocoding and Monitoring US Socioeconomic Inequalities in Health: An introduction to using area-based socioeconomic measures

MULTI-LEVEL MODELING

Modeling area-based socioeconomic measures: a multilevel perspective

It is well known that there are substantial area variations in mortality rates in the U.S. However, the presence of area differences in mortality does not necessarily mean that area matters. Area variations in mortality can be observed due to a number of reasons some of which may be due to characteristics that relate to areas and others that relate to the characteristics of the individuals who live in these areas. Disentangling the two sources of variation (e.g.: individual and area) in mortality is therefore vital to distinguishing area differences from the difference that area makes. Such an approach to examining area variations in mortality, consequently, entails describing the patterning and causes in mortality variations, which in turn, requires answering the following empirical questions preferably in a sequential manner.

Before we outline the questions, it is worth asking what role could places or areas play in influencing mortality (and indeed other health outcomes). Pure locational attributes of an area (e.g., altitude, proximity to coast) or environmental aspects of an area (e.g., levels of air pollution) or structural attributes of an area (e.g., residential segregation, labor markets, population density) or collective social aspects of an area (e.g., proportion of poor in an area, proportion population that has less than high school education) are some concrete elements along which area variations in mortality may get patterned. Indeed, the different examples mentioned above need not be mutually exclusive. Thus, an examination of area variations and area-based explanations to these variations could be addressed by answering the following questions:

First, how does the total variation in mortality get partitioned across the individual and area levels?

Second, how much of the variation in mortality that is attributable to areas is influenced by the characteristics of individual residents who live in these areas?

Third, does the magnitude of variation in mortality that is attributable to areas differ for different population groups? For instance, is the area-attributable variation in mortality

greater for blacks as compared to whites?

Fourth, to what extent do area-based characteristics account for the area-attributable variation in mortality, in whites and blacks, for example?

Fifth, what is the systematic relationship between area-based characteristics and mortality, and does this relationship systematically differ across different population sub-groups?

Answering these types of questions requires adopting a multilevel statistical modeling approach (also known as hierarchical, mixed and random-effects, covariance components or random-coefficient regression). These techniques have provided researchers one possible framework for incorporating and understanding the role of areas and context while studying mortality variations. The key advantage of this approach is, therefore, in analyzing, "why some areas are more likely to experience higher levels of mortality, while taking into account why some individuals (independent of which area they live in) are more likely to die".

The use of multilevel statistical techniques is especially pertinent under the following circumstances.

The first is when the individual health outcome measure (or group-specific prevalence) are anticipated to be clustered with the source of clustering being a geographic area, such as block-groups or/and census-tracts and the interest is in ascertaining the relative importance of the different levels for the outcome. This is particularly relevant for public health departments as they provide a clue about the level at which actions occur. The assessment of what level matters the most for the outcomes can be done unconditionally (not adjusting for covariates) and conditionally (adjusted for covariates).

The second situation that necessitates the use of multilevel methods is when the exposure is measured at multiple levels and the interest is in evaluating the relative importance of a same ABSM at different levels (e.g.: establishing whether the block-group poverty has a larger effect than the census-tract poverty).

Finally, multilevel methods offer a bridge between statistical modeling and descriptive map-based presentations. Since

the specific census-tracts and block-groups identifiers are intrinsic to the analytical design, it is possible to develop conditional statistical maps showing how different places are doing on a particular health outcome and importantly whether the “geography of health” differs for different population sub-groups. This provides a useful means to monitor health inequalities that is conditional on a range of important socioeconomic characteristics. Technical benefits also flow from utilizing this perspective. There of course are serious substantive issues (such as “naming” and “shaming” places) as well technical issues (such as instability in intrinsically small areas with less population; mismatch of outcome measure with the denominator information) that need to be considered given the immediate appeal of maps. While strategies drawing upon “empirically bayes” modeling (utilized widely within the multilevel models) or smoothing may bring certain technical solutions, issues of mapping for small areas in particular are complex and substantive.

While this approach is gaining usage in public health research, given the relative complexity of these modeling strategies it is yet to become a part of the mainstream public health surveillance and monitoring. At the same time, the reasons to empirically evaluate the above questions are compelling. For instance, patterns of all cause mortality are likely to be shaped by a complex constellation of compositional and contextual factors that may conceivably vary for different population subgroups, as suggested, for example, by different leading causes of death for different racial/ethnic groups. An investigation of the racial/ethnic heterogeneity in geographic variation in mortality can give insight into the relative importance of compositional and contextual effects to mortality experienced by different racial/ethnic populations. For example, if the geographic variation in mortality rates for a specific group is large, this suggests that geographically varying contextual factors may be of particular importance in shaping mortality risk for this population. Conversely, if the geographic variation in mortality rates is low for a particular group, it suggests that contextual factors are of relatively less importance in shaping overall mortality risk for that population.

The subject of modeling area-related effects – through measuring the area-attributable variation and through identifying area-based characteristics – is intrinsically multilevel and this note outlined the sort of questions and

motivations that could underlie investigations of variation in health and mortality.

Multilevel models may now be implemented using a variety of software packages including SAS, STATA, R and MLwiN. The Center for Multilevel Modeling website provides a list of these software packages at <http://multilevel.ioc.ac.uk/softrev/index.html>

For fundamental texts, see:

Goldstein H. Multilevel statistical models. 2nd ed. London: Arnold, 1995.

Longford N. Random coefficient models. Oxford: Clarendon Press, 1993.

Raudenbush S, Bryk A. Hierarchical linear models: applications and data analysis methods. Thousand Oaks: Sage, 2002.

For applied introductions to multilevel statistical models, see:

Hox J. Multilevel analysis: techniques and applications. Mahwah, NJ: Lawrence Erlbaum Associates, 2002.

Leyland AH, Goldstein H. Multilevel modelling of health statistics. Wiley Series in Probability and Statistics. Chichester: John Wiley & Sons Ltd., 2001.

Snijders T, Bosker R. Multilevel analysis: an introduction to basic and advanced multilevel modeling. London: Sage Publications, 1999.

Subramanian SV, Jones K, Duncan C, 2003, Multilevel methods for public health research, in Kawachi I, Berkman L. Eds. Neighborhoods and Health, New York: Oxford University Press, 65-111.

For hands-on tutorial, see:

Browne WJ. MCMC estimation in MLwiN. London: Centre for Multilevel Modelling, Institute of Education, 2002.

Rasbash J, Browne W, Goldstein H, Yang M, Plewis I, Healy M, Woodhouse G, Draper D, Langford I, Lewis T. A user's guide to MLwiN, Version 2.1. London: Multilevel Models Project, Institute of Education, University of London, 2000.

For issues related to mapping see:

Elliott P, Wakefield J, Best N, Briggs D (eds). Spatial Epidemiology:

Methods and Applications. Oxford: Oxford University Press, 2000.

Maantay J. Mapping environmental injustices: pitfalls and potential of geographic information systems in assessing environmental health and equity. *Environ Health Perspect* 2002; 110 (suppl 2):161-171.

Monmonier M. *How to Lie with Maps*. 2nd ed. Chicago: University of Chicago Press, 1996.

Monmonier M. *Cartographies of Danger: Mapping Hazards in America*. Chicago: University of Chicago Press, 1997.

Moore DA, Carpenter TE. Spatial analytical methods and geographic information systems: use in health research and epidemiology. *Epidemiol Rev* 1999 21:143-161.

Richards TB, Croner CM, Rushton G, Brown CK, Fowler L. Geographic information systems and public health: mapping the future. *Public Health Rep* 1999; 114:359-373.

This work was funded by the National Institutes of Health (1R01HD36865-01) via the National Institute of Child Health & Human Development (NICHD) and the Office of Behavioral & Social Science Research (OBSSR).

Copyright © 2004 by the President and Fellows of Harvard College - The Public Health Disparities Geocoding Project.