

## New Approaches to Multilevel Analysis

John R. Beard

The paper by Yu-Sheng and Ying-Chih in this issue highlights some of the challenges confronting studies examining neighborhood-level influences on health. Studies of this sort often involve data on individuals nested within neighborhoods. A key issue that arises in analyses of these types of data structure is the potential for nonindependence of observations (i.e., the possibility of within-neighborhood correlations between individual-level outcomes). Ignoring this can result in invalid standard errors, incorrect (typically anticonservative) inferences, and inefficient estimates.<sup>1</sup> Multilevel models are particularly suited to analysis of these types of data structures and are being increasingly used in studies of neighborhood-level effects.<sup>2,3</sup>

Another issue that needs to be considered in studies of this type is the potential for colinearity between many of the environmental characteristics of interest. For example, neighborhoods with low education levels are likely to also be areas of high unemployment. To address this issue, a number of studies have used factor analysis to aggregate neighborhood-level characteristics into a smaller number of constructs that may better reflect the dimensions underlying these associations.<sup>4</sup>

Part of the analysis by Yu-Sheng and Ying-Chih uses factor analysis to identify three dimensions of neighborhood. They then test the influence of these using multilevel models. However, as the authors point out, while multilevel models address the potential for intraneighborhood clustering of individual-level characteristics, they treat neighborhood dimensions as independent characteristics and neither provide information on, nor take account of, any relationship between these dimensions. To address this issue, they applied cluster analysis to the neighborhood dimensions identified by factor analysis. This generated six types of neighborhoods with, for example, neighborhoods with a high concentration of inhabitants younger than 15 years, a moderate education level, and a moderate level of single-parent families being associated with poorer outcomes. This provides a more complex perspective on the mix of characteristics that may make a neighborhood supportive of good health.

A further important issue for studies investigating neighborhood effects is the spatial distribution of this data and the potential for *spatial* autocorrelation between neighborhood estimates. Thus, for example, a neighborhood immediately adjacent to a neighborhood with high unemployment may, itself, be more likely to have a higher rate of unemployment than would be expected by chance. This is exacerbated by the arbitrary (from a health perspective) boundaries of the administrative data often used in this type of research.

Spatial analyses can be useful in addressing these issues. These can be used firstly to assess whether areas sharing a common boundary have more similar area-level residuals than would be expected under spatial randomness.<sup>5</sup> If there is

---

Beard is with the Center for Urban Epidemiologic Studies, New York Academy of Medicine, New York, NY, USA and is with the University of Sydney, Sydney Australia and Southern Cross University, Lismore, Australia.

Correspondence: John R. Beard, Center for Urban Epidemiologic Studies, New York Academy of Medicine, New York, NY, USA. (E-mail: jbeard@nyam.org)

significant spatial autocorrelation in a particular analysis, spatial models can then be used to take account of spatial adjacency issues between neighborhoods.<sup>6</sup> The conditional autoregressive (CAR) model, for example, models area-specific intercepts as random effects.<sup>7</sup> This can be used to adjust relative risk estimates in neighborhoods toward the mean risk in neighboring areas, has the capacity to also deal with temporal adjacency (for example, if data has been collected over a number of years), and can include multivariable analyses. CAR models are also useful in situations where some expected cell counts are low as they draw information from neighboring cells in generating their estimates. By comparison, a rule of thumb for building multilevel models suggests these require >29 groups with >29 individuals per group.<sup>8</sup>

However, spatial models are Bayesian in their construction, which may present researchers who are trained in frequentist approaches some challenges in analysis. They also require extensive computing capacity and have limitations when dealing with large numbers of spatial units.

Perhaps we now need to explore how all these approaches can be brought together in a way that addresses interneighborhood and intraneighborhood autocorrelation, low cell numbers, colinearity of neighborhood measures, and clustering of these measures. One such approach may be to use spatial models to develop exposure estimates. This can help overcome low cell counts and take account of spatial and temporal adjacency. Factor analysis might then be applied to these neighborhood estimates in order to identify parsimonious dimensions of neighborhood, which could then be included in traditional multilevel models. If the approach suggested by Yu-Sheng and Ying-Chih becomes more widely accepted, clustering between these dimensions can also be explored.

## REFERENCES

1. Diez-Roux A. Multilevel analysis in public health research. *Annu Rev Public Health*. 2000;21:171–192.
2. Rice N, Jones A. Multilevel models and health economics. *Health Econ*. 1997;6:561–575.
3. Bryk A, Raudenbush S. *Hierarchical Linear Models*. Newbury Park: Sage; 1992.
4. Matheson FI, Moineddin R, Dunn JR, Creatore MI, Gozdyra P, Glazier RH. Urban neighborhoods, chronic stress, gender and depression. *Soc Sci Med*. 2006;63(10):2604–2616.
5. Burnett R, Ma R, Jerrett M. The spatial association between community air pollution and mortality: a new method of analyzing correlated geographic cohort data. *Environ Health Perspect*. 2001;109(suppl 3):375–380.
6. Bernardinelli L, Clayton D, Pascutto C, Montomoli C, Ghisladi M, Songini M. Bayesian analysis of space-time variation in disease risk. *Stat Med*. 1995;14:2433–2443.
7. Besag J, York J, Mollie A. Bayesian image restoration with application in spatial statistics. *Ann Inst Stat Math*. 1991;43:1–20.
8. Kreft I. *Are Multilevel Techniques Necessary? An Overview Including Simulation Studies*. Los Angeles: California State University Press; 1996.