Int. J. Epidemiol. Advance Access published March 31, 2005

Published by Oxford University Press on behalf of the International Epidemiological Association

International Journal of Epidemiology
doi:10.1093/ije/dyi020

Commentary: Estimating and understanding area health effects

Ana V Diez Roux

In this issue of the *Journal*, Breeze *et al*. report that living in a deprived area is associated with poor quality of life in a large population-based sample of older adults living in the UK.[1] Their paper adds to a large body of work reporting associations between area socio-economic characteristics or area deprivation and a variety of health outcomes.[2] The focus on the elderly population is especially interesting because, as Breeze *et al*. note, there are reasons to believe that area characteristics may be especially relevant to the health and well-being of elderly people who are likely to spend more time in their local areas and rely on their local areas for services and social interactions.

Like other researchers, Breeze *et al*. analyse data from an observational study to estimate 'area effects' after controlling for individual-level social class. The need to control for differences in the socio-economic position of people living in different areas has been a key challenge in the field, especially because the forces shaping residential location generate associations between individual-level social class and area deprivation. The most common approach in the literature is to use regression methods to adjust for individual-level characteristics in the estimation of area effects. An important assumption in the use of regression methods is that the overlap in the distribution of measured individual-level confounders across categories of area deprivation is sufficient for the regression adjustment to yield valid estimates of the 'independent' effect of area. In the absence of sufficient overlap, adjusted estimates necessarily imply extrapolations beyond the information available in the data, the validity of which cannot be tested.[3] There is no unique answer as to how much overlap is sufficient to allow meaningful estimates. A certain amount of extrapolation is inherent in all scientific inquiry. In fact, the process of adjustment always involves varying amounts of extrapolation from the data at hand to what would have been observed if certain features of the data were different (e.g. if the age distributions of two groups being compared were not as different as they are). Nevertheless, it is important to be explicit about how far we are straying from the data in our extrapolated conclusions, so that readers can judge for themselves whether the assumptions implicit in the extrapolation are likely to be valid.

Breeze *et al*. address the issue of overlap between area deprivation and social class in their sample by reporting the distribution of individual-level confounders by categories of area characteristics in Table 2 of their paper. They show that although, as expected, social class and area deprivation are associated, there is still substantial variability in deprivation of the area of residence within social class categories. A similar pattern has been reported in other contexts,[4] suggesting that lack of overlap is unlikely to be as important a problem in estimates of area effects from observational studies as is sometimes implied. In addition, the analytical approach followed by Breeze *et al*. is slightly different from the usual approach in the literature in that they estimate the combined (as opposed to the 'independent') effects of social class and area deprivation. This is accomplished very simply by estimating prevalence ratios for cross-classified cells of social class and area deprivation. This approach has the advantage that estimates are based on the people in each cross-classified cell, and is closer to reality than adjusted estimates which smooth (and extrapolate) over cells and artificially separate out effects which are inextricably linked in reality. A disadvantage is of course that the cells can get very small, as they sometimes do in the analyses reported by Breeze *et al*., although they are somewhat protected from this problem by relatively large overall sample size. Another disadvantage is that presenting results for each cell can become very cumbersome. Breeze *et al*. avoid this by reporting results only for the most extreme social class and area deprivation categories, but this means that we do not see all the data.

An important assumption in the use of adjustment (including stratification) strategies to estimate causal area effects from observational data is that the adjusted comparison approximates the counterfactual contrast of interest. In the results reported by Breeze *et al*., the prevalence ratio of 1.49 reported in Table 3 for home management in people of social class I/II living in the most deprived Carstairs quartile compared with people of similar social class living in the least deprived Carstairs quartile can be interpreted as meaning that if people living in the least deprived area quartile lived instead in the most deprived quartile, their probability of having a mobility impairment would be 49% higher. The assumption is that the people living in the least deprived areas are a good proxy for what the people living in the most deprived areas would be like if they did not live in the most deprived areas. Usually in epidemiological jargon, this implies no residual confounding by measured variables and no unmeasured confounders. The inability to undeniably confirm the validity of this assumption is the crucial limitation of observational analyses, like those reported by Breeze *et al*. However, there are ways to test the sensitivity of results due to this assumption, for example, by examining how much results change when covariates are modelled in different ways or by estimating how strongly an omitted confounder would have to be associated with area deprivation and with the outcome to create the associations observed. These types of sensitivity analyses have only recently

Department of Epidemiology and Center for Social Epidemiology and Population Health, University of Michigan, Ann Arbor, MI, USA.

Correspondence: 1214 South University 2nd Floor, Ann Arbor, MI 48104, USA. E-mail: adiezrou@umich.edu

begun to appear in epidemiological reports and have not been systematically used in studies of area effects to date.

Aside from the possibility of confounding by social class, cross-sectional analyses like those reported by Breeze *et al*. are vulnerable to the critique that the associations they report result from the selection of unhealthy people in the poorer areas. This type of explanation is akin to the downward drift hypothesis often alluded to, in individual-level studies of social inequalities in health, which has been shown to be only a partial explanation for differences in health by social class. Limited longitudinal evidence suggests that place of residence is related to changes in functional status over time in elderly people,[5] and hence that selection factors are unlikely to fully account even for cross-sectional findings. However, because they rely exclusively on cross-sectional data, Breeze *et al*. cannot categorically rule out selection effects, although they do show that effects are present after statistically controlling the health status information available. One could also argue that health status is at least partly the result of past exposure to area deprivation, and hence is a mediator rather than a confounder of area effects on quality of life. This conundrum (unadjusted estimates may be affected by residual confounding but adjusted estimates may underestimate true effects) is typical of these types of analyses and can only be resolved with longitudinal data and experimental or quasi-experimental designs. Complexity is increased by the possibility that some variables may be simultaneously mediators and confounders of area effects over time, which poses particular problems for estimation.[6]

Area socio-economic characteristics have been found to be associated with many different health-related outcomes in a variety of contexts. Although it is plausible that interrelated features of areas are related to multiple health outcomes through common processes as well as through different but interrelated processes, the absence of a clearly articulated theory on what these processes might be leaves open the possibility that the associations reported are simply the repeated manifestation of confounding by individual-level attributes. Similarly to research on life course influences on adult health, where it has been suggested that the search for specificity of associations may help shed light on the processes involved,[7] identifying outcomes unrelated to area deprivation may help understand the mechanisms and processes linking areas to health. Breeze *et al*. found that area deprivation was much more weakly related to mobility than to home management, self-care, and social interaction. Whether this reflects a real pattern or is only a chance finding remains to be determined. If real, investigating the reasons for the difference could suggest hypotheses about specific mechanisms through which area effects might operate. Breeze *et al*. suggest that selection effects may be stronger for mobility, but unfortunately their data do not allow them to test this.

The need for greater specificity implies that we need to be more specific not only in defining the outcomes for which we expect (or do not expect) area effects but also in moving beyond general measures of area deprivation to the examination of specific area attributes: the testing of hypotheses relating specific features of areas to specific health outcomes. The need for specific explanations does not mean that common processes are not involved (for example, access to public recreational spaces may have effects on a variety of health outcomes through common or interrelated mechanisms), but these general processes need to be articulated in their specific details. This is necessary not only from the perspective of drawing causal inferences but also in order to understand what specific policies and interventions on areas hold the greatest promise for improving health.

References

[1] Breeze E, Jones DA, Wilkinson P *et al*. Area deprivation, social class and quality of life among people aged 75 years and over in Britain. *Int J Epidemiol* doi:10.1093/ije/dyh328.

[2] Pickett KE, Pearl M. Multilevel analyses of neighbourhood socioeconomic context and health outcomes: a critical review. *J Epidemiol Community Health* 2001;**55:**111–22.

[3] Oakes JM. The (mis)estimation of neighborhood effects: causal inference for a practicable social epidemiology. *Soc Sci Med* 2004; **58:**1929–52.

[4] Diez Roux AV, Kiefe CI, Jacobs DR Jr *et al*. Area characteristics and individual-level socioeconomic position indicators in three population-based epidemiologic studies. *Ann Epidemiol* 2001;**11:**395–405.

[5] Balfour JL, Kaplan GA. Neighborhood environment and loss of physical function in older adults: evidence from the Alameda County Study. *Am J Epidemiol* 2002;**155:**507–15.

[6] Cole SR, Hernan MA. Fallibility in estimating direct effects. *Int J Epidemiol* 2002;**31:**163–65.

[7] Davey Smith G, Hart CL, Blane D *et al*. Lifetime socioeconomic position and mortality: prospective observational study. *BMJ* 1997; **314:**547–52.