


Nancy Krieger
Department of Health and Social Behavior
Harvard School of Public Health
Boston, MA 02115

David Gordon
School for Policy Studies
Bristol BS8 1TN, England

THE AUTHORS REPLY

In response to Krieger and Gordon (1), let us restate the purpose and conclusions of our studies (2, 3). We analyzed two data sets (the Panel Study of Income Dynamics (PSID) and the National Maternal and Infant Health Survey (NMIHS)) and two distinct health outcomes (self-rated overall health and infant mortality). Our analyses were confined to the question of how well individual-level associations can be inferred when aggregate socioeconomic variables proxy for individual characteristics in health outcome equations. We did not assess the relative importance of contextual versus individual characteristics in relation to health. To our knowledge, there are only three other studies that have even addressed the specific question that we posed (4–6). We offered the following conclusions: 1) census-based aggregate variables are poor proxies for individual characteristics—they have substantially less statistical power than do microlevel measures and are likely to overestimate effects; 2) aggregate socioeconomic variables are highly multicollinear; and 3) somewhat surprisingly, while there is some improvement in precision if a researcher uses aggregate data drawn from a more recent census year compared with one that is 10 years earlier or that uses data aggregated at the census-tract rather than the zip code level, these gains are small.

Krieger and Gordon question our conclusions. They argue that the data we use are unrepresentative, that our statistical methods are flawed, and that our results are at odds with the bulk of research in the area. We disagree.

First, Krieger and Gordon limit their criticisms to only one of the data sets we used, the PSID. Some of their concerns are inapplicable to the NMIHS, including the prominent issue of sample attrition. They also highlight only those PSID subsamples in which we had relatively low rates of matching between the survey sample and the census data (68–72 percent). They fail to mention that we matched 95 percent of the PSID respondents and 89 percent of the NMIHS respondents to 1980 zip codes. Even if one considers only the 1980 zip code analysis, our primary conclusions remain the same.

They question the representativeness of the PSID sample. The many evaluations of the PSID all come to the conclusion that, when weighted, it remains representative of the non-Hispanic US population (7, 8). Of course, these evaluations have not considered the representativeness of the distribution of PSID respondents across zip codes or census tracts, but our own tabulations suggest that respondents are not concentrated in particular zip codes or census tracts or in ones that are unrepresentative of the nation as a whole.

Some issues Krieger and Gordon raise are not pertinent to the validity of estimates based on our PSID samples but, instead, concern their reliability. Thus, for example, the use of sample weights can and typically does reduce the precision of estimates, but does not bias them. Moreover, if one takes into account the fact that weights were used when calculating test statistics, as we did, the decrease in precision will not lead to faulty inference. (Let us add that unweighted tabulations yield results that are similar to the weighted ones we report). Similarly, geographic clustering does not bias estimates, but typically reduces the efficiency of a sample. We experimented with calculating standard errors using methods that account for clustering, finding that the degree of clustering that exists in our data was sufficiently minor to have little effect on calculated standard errors. Furthermore, the fact that, in the end, we ignored clustering when calculating standard errors would bias the standard errors we estimated on the aggregate variables more than they bias the standard errors on the microvariables (9, 10). Thus, had we reported standard errors that took into account the clustering in our data, it is likely that our finding that census-based proxies lack statistical power relative to their microlevel counterparts would have been strengthened.

Most important, Krieger and Gordon fail to note that samples used in previous research comparing results based on aggregate versus micromeasures have less of a claim to representativeness than do the two data sets we used. Of the three other studies that directly compared the use of census-based aggregate with microlevel socioeconomic variables in health outcome equations, the samples studied were 101 women from Alameda County, California (4), a subsample of participants in the Kaiser Health Insurance plan in one state...
Krieger emphasized the similarity between the point estimates obtained on her aggregate occupation or education variables and those on the corresponding microlevel estimates when studying health outcomes (2-6). Two of the remaining three are by Krieger, including the pilot study based on 101 observations.

We cited and discussed all three of the major studies in this area that preceded our Journal article: Krieger’s larger study (5); a study by Greenwald et al. (6), and our study in the Journal of the American Statistical Association (3). Greenwald et al. concluded that error due to imputation of individual characteristics from surrounding populations is a health maintenance organization in California. She analyzed only census-tract and block-group data, leaving open entirely the question of how well zip code data perform, a conclusion consistent with ours. Indeed, one reason Krieger and Gordon contend that ours are the anomalous findings is that we most often find coefficients on aggregate variables used as proxies for individual-level data to be larger than those when the individual-level variables themselves are used. Yet, our statistical model explicates that using an aggregate variable to measure an individual characteristic is subject to two sources of bias, a downward bias due to measurement error and an upward bias due to aggregation (2, 3). Our prediction that most often the aggregation bias would dominate is empirically supported by our findings; still, it is possible in some circumstances to arrive at overestimated. In fact, we discuss one example in which we found this to be true in our paper in the Journal of the American Statistical Association (3). We are concerned about the increasing use of aggregate variables to measure individual characteristics because the resulting bias is likely to be substantial, and an epidemiologist cannot even be confident that he or she can correctly interpret its direction.

On practical issues, block-group data are rarely available to researchers, not because they do not exist in the census, but because health data sets rarely geocode down to that level or make data geocoded to that level publicly available, due to confidentiality concerns. This is often true at the census-tract level as well. Similarly, whether or not block-group data are made available for rural residents, the point is that investigators should question the salience of the construct of a “block group” in sparsely populated rural areas. So, too, “block numbering areas” in rural areas are of a size similar to census tracts in urban areas, but the two groupings are not equivalent in that census tracts, unlike block numbering areas, are relatively permanent statistical subdivisions designed to be homogeneous with respect to population characteristics, economic status, and living conditions.

In our data, the magnitude of the improvement associated with moving from zip codes to census tracts was small. This should be of some comfort to those working with data that include information on zip codes, but not census tracts. As we stated in our article, moving to block-group data might yield substantial gains, but this remains an empirical question. While block groups are more homogeneous than are census tracts or zip codes, the critical question relates to the magnitude of these differences and the extent to which such differences have noticeable effects on either the magnitude or the precision of estimated effects. Krieger’s own calculations leave the answer to this question ambiguous. In half of her calculations, estimates based on block groups were more pre-
ciscely estimated than were those based on census tracts, but in some of her calculations the reverse was true, and in no case were differences in confidence intervals very great (5, table 2). Findings from Australia (11) showing substantially greater homogeneity in "collector's districts" compared with "postcode" suggest that block-group data may be substantially better indicators of individual characteristics than are census-tract data. However, Hyndeman et al. (11) never estimate health outcome equations, and census or postal aggregations in one country may have very different meanings and potential links to health than those in another. Social stratification, racial or ethnic segregation, and the economic behavior of individuals are reflected in geography and will vary among countries, as will approaches to defining government statistical units of aggregation and their relation to these features of the socioeconomic fabric of societies. Studies based on data from other countries cannot replace empirical evaluation of US data on this particular research question.

In short, we see nothing in Krieger and Gordon's comments that would lead us to modify any of our original conclusions. We believe that our findings offer useful practical and conceptual advice to investigators using this approach or interpreting results arrived at in this manner. We also believe that they highlight the severe limitations of this approach and the urgent need for investigators committed to gaining better scientific understanding of the causal mechanisms underlying social disparities in health to have access to health data sets with better-quality microdata on socioeconomic characteristics. We do not argue against stepped-up geocoding of data sets or the critical importance of aspects of geographic areas to the health of individuals. We explicitly state in our article (2) that our findings have limited application to the use of aggregate variables to estimate contextual effects. The very finding Krieger and Gordon call anomalous, namely that the use of aggregate variables will often overstate the effects of individual socioeconomic characteristics, is completely consistent with a view of the world that posits aggregate variables to tap more global or contextual constructs than individual characteristics that can influence health (3).

REFERENCES


Arline T. Geronimus
Department of Health Behavior and Health Education
University of Michigan School of Public Health
Ann Arbor, MI 48109–2029

John Bound
Department of Economics
University of Michigan
Ann Arbor, MI 48109–2029