It's the data we have: what can we learn from 5-year ACS data about segregation and neighborhoods given the margins of error?

Nancy A. Denton, Hui-shien Tsao, Jin-wook Lee, Ruby Wang and Jeffrey Napierala, University at Albany, SUNY

There is little doubt that the replacement of the Census long-form with the ACS has led to many advantages to researchers, policy makers, and community leaders concerned with demographic data about places big and small. Depending on the size of the place, data are now available annually, every three years, or every five years – all of which are a big improvement over waiting the ten years between censuses. Thus far, the relative newness of the transformation of once-every-ten-year-census-data into what feels like a deluge of data releases every year and the difficulties of comparing data over time because of the overlap of the multi-year data have received the most attention from researchers. And while those issues have also been important to planners and local officials who eagerly look at each new piece of data for their area, something else has attracted their attention: margins of error.

Margins of error (MOE) now accompany every number provided by the Census Bureau. It is also true that long-form census data also had margins of error associated with it, but they had to be calculated by the data user him- or her- self. In addition, the ten-year interval between censuses made it more likely that changes, even to small groups or in small areas, were reflecting a reality rather than sampling error. And even if that were not the case, people tended to act as if it were. If a census tract had 6 blacks in one census and 40 in the next census, that change was treated as a real increase by almost all data users and reflected in the computation of segregation indices. But more frequent data releases, and thus shorter time intervals between comparisons, elevates the importance of the margins of error. Are the changes shown in the data real or are they just reflecting sampling error? People want to know why place x has only two poor children or zero elderly or 5 college graduates. And while it is easy to compute statistical significance for pairs of numbers, doing this for all the cells in a table of six racial groups stratified by education and age for 12 metropolitan and 15 micropolitan areas is more daunting. Plus, how do you compare one small town to the next or one neighborhood to the next, things that planners and city government officials need to do every day. So depending on the uses to which the data are being put, the MOE issue may be more or less important, but it cannot be ignored as seems to generally be the case among many researchers using aggregate data.

One area where research and community use of aggregate census data overlap is in the identification of the characteristics of neighborhoods. Researchers tend to use these data to study residential segregation, or use the neighborhood data in models estimating neighborhood effects. Policy makers and communities need to know how their neighborhoods are changing, where to locate businesses, and which neighborhoods qualify for federal grant programs, among many other things. One topic

relevant to all these groups is residential segregation. Studies of residential segregation from the 2010 Census have generated some controversy in the media. Though all agree that segregation is declining, particularly between blacks and whites, sociologists Logan and Stults (2011), as well as Iceland (2011) and Frey (2010) have a much less optimistic take on what the numbers mean than economists Glaeser and Vigdor (2012) who declared "the end of the segregated century." Remapping Debate, using maps created by Social Explorer, summarized these debates as "segregation is alive and well," and included quotes from the various researchers (Gurian, 2011). But none of these reports discussed how the segregation scores themselves, or changes in them, were affected by the margins of error in the ACS 2005-2009 data that were used to represent 2010 in the early reports. In addition, they concentrated on metropolitan areas, ignoring micropolitan areas, smaller places where 10 percent of the U.S. population lives.

In this paper we focus on segregation to explore some the issues raised by the margins of error for substantive analyses of topics like residential segregation. Our aim is to see how much the margins of error affect the computed segregation indices for places and groups of different sizes. Though statistics is straightforward in predicting that size of place and size of group will both affect the estimates, what margin of error means after all, there has been no investigation of whether this is consistently true given the tremendous variation across groups and places in terms of other characteristics. It is also not clear if other characteristics of places also matter, and if so how. There is also the issue of location within an area: if the estimates are good for certain groups in the city but better for others in the suburbs, how does one compare the groups? As diversity spreads to smaller and smaller areas (Lee et al., 2012) how can we accurately understand and be sure of what is happening? The issue here is less statistical than it is practical: can these numbers be presented and used in a way that they are believable and thus useful for policy, rather than being easily refuted by the next set of data. As Congressional support for the ACS weakens, the question of how useful the numbers are to data users is of increasing importance.

The five-year ACS data (2006-2010) and the 2010 Census data contain the same racial data so it is possible to compute segregation indices using the 100% data and the ACS sample data and compare them. Note that while 2005-2009 ACS data are also available and were used in the initial 2010 segregation reports, those data contain four years in common with the 2006-2010 data so comparing them is not appropriate. The preliminary data we are using are for New York State, making it is easier for us to interpret the results as we know the cities, metropolitan and micropolitan areas. The state has the nearly the full range of diversity in terms of race/ethnic groups and geographical areas that is found in the U.S., though we ultimately plan to look at the entire country. We begin our analysis by computing the two most commonly used segregation indices, Dissimilarity and P-star, using census tracts as neighborhoods in the metropolitan and micropolitan areas. Indices are computed for non-Hispanic whites, blacks, and Asians, as well as Hispanics. We then repeat the same calculations using block groups as one way to assess the effects of scale. Next we compute the

segregation indices addressing the MOE explicitly: first, assigning each tract the maximum as indicated by adding the MOE to the estimate; second, using the minimum calculated by subtracting the MOE from the estimate and setting the result to zero if it is negative to enable calculation of the indices; and third, by randomly assigning a count between the maximum and minimum (or zero) as indicated by the estimate and the MOE. This provides us with five different estimates of the segregation between any two groups for each of the metropolitan and micropolitan areas which we can analyze descriptively. As a second phase in the analyses, we graph the neighborhood level data for selected groups in selected metropolitan and micropolitan areas to illustrate patterns for specific groups or sizes of place. In the final phase we estimate multivariate models to show the effects of various place characteristics (size, density, relative group sizes, socioeconomic status, economic characteristics, location in the state) on the measures of segregation.

At this point, the project is at an exploratory stage, but early results show sizeable variations in the indices, more than we initially expected. While they at first seem to relate mainly to group size and size of place, the patterns are not clear and suggest interactions between size of group, size of place, and type of place. And the variety of results across the different comparisons (whites to blacks, whites to Asians, whites to Hispanics, etc.) is complex to interpret. Early versions of the maps also present complexities and suggest new relationships. We are still at the early stages of the model building but we expect these to help us sort out the results.

In conclusion, it is worth noting that the MOE associated with tracts or block groups can be problematic for studies beyond segregation. In fact, if one is interested in neighborhoods per se, rather than aggregated indices, the MOEs may be even more problematic. So even if the segregation indices are stable, this topic merits attention, though admittedly our results thus far suggest more questions than they answer. In addition, this is the first substantive investigation of the MOE of which we are aware so that increases both our interest in and the importance of our paper.

References:

Frey, William H. 2010. "Segregation Measures for States and Metropolitan Areas." CensusScope.org. http://www.censusscope.org/segregation.html.

Glaeser, Edward and Jacob Vigdor. 2012. "The End of the Segregated Century." Civic Report No. 66, January 2012. Manhattan Institute. http://www.manhattan-institute.org/html/cr_66.htm/pdf/cr_66_appendix.pdf

Gurian, Craig. 2011. Remapping Debate. April 21, 2011. http://www.remappingdebate.org/map-data-tool/new-maps-show-segregation-alive-and-well

Iceland, John. 2010. "Segregation." Changing American Communities Project. http://projects.pop.psu.edu/canc/segregation

Lee, Barrett, John Iceland, and Gregory Sharp. 2012. "Racial and Ethnic Divsersity Goes Local: Charting Change in American Communities over Three Decades." Census Brief prepared for Project US2010. http://www.s4.brown.edu/us2010/Data/Report/report08292012.pdf

Logan, John R. and Brian Stults. 2011. "The Persistence of Segregation in the Metropolis: New Findings from the 2010 Census" Census Brief prepared for Project US2010. http://www.s4.brown.edu/us2010/Data/Report/report2.pdf