

MAY 24 2011



UNITED STATES DEPARTMENT OF COMMERCE
Economics and Statistics Administration
U.S. Census Bureau
Washington, DC 20233-0001

2011 AMERICAN COMMUNITY SURVEY MEMORANDUM SERIES #ACS11-R-03

MEMORANDUM FOR Robert M. Groves
Director, US Census Bureau

From: Section 203 Project Steering Committee
Roderick Little, Co-Chair (DIR) *RAL*
Alfredo Navarro, Co-Chair (DSSD) *AN*
William Bell (ADRM)
Melissa Creech, Ex Officio (Legal Office)
Robert Kominski (SEHSD)
Catherine McCully (DIR)
David Whitford (DSSD)

Prepared by: Michael Beaghen
American Community Survey Estimation Branch
Decennial Statistical Studies Division

Subject: Criteria for Deciding on whether to Use Statistical
Modeling in Producing Section 203 Determinations with
American Community Survey Data

I. Introduction

The Census Bureau will produce the 'Section 203' determinations establishing coverage of jurisdictions which require language assistance at the polls. 42 U.S.C. Section 1973aa-1a as amended in 2006 requires the Census Bureau Director to use American Community Survey (ACS) or decennial census data to determine whether a state or political subdivision falls within the definition of a 'covered area.' A covered area exists when more than 5 percent or when more than 10,000 of the voting age citizens are members of a single-language minority group who do not "speak or understand English adequately enough to participate in the electoral process" and have an illiteracy rate higher than the national illiteracy rate.

Determinations are made based on the total number of voting age citizens who are of limited English proficiency (LEP), the percent of voting age citizens who are LEP, and the percent of voting age citizens who are LEP and have a less than 5th grade education. For a jurisdiction to be covered by Section 203, at least 5 percent of its voting age citizens should be LEP and of that language minority group or there should be at least 10,000 voting age citizens who are LEP and of

that language minority group. In addition, the proportion of LEP voting age citizens of that language minority group who have a less than 5th grade education should be greater than the national average. The Census Bureau will apply this formula for coverage to about 8,000 jurisdictions. These include all counties and county equivalents, minor civil divisions (MCD) in states with functioning MCDs (the New England states, Michigan, and Wisconsin), and American Indian Areas/Alaska Native Areas (AIA/ANA). The determinations are made separately within each jurisdiction for the 68 covered language minority groups which include Spanish/Hispanic/Latino, American Indian/Alaska Native (AIAN), AIAN tribal groups, Asian, and Asian groups.

In 2002 these determinations were made with Census 2000 sample data, also known as 'long form' data (Federal Register, 2002). In 2006, Congress amended 42 U.S.C. Section 1973aa-1a to require the Census Bureau Director to use the ACS data to make future determinations. In 2011 the determinations will be based on the 2005-2009 ACS estimates. Because the ACS has smaller sample sizes and larger variances, questions arose about the potential effect of the use of ACS estimates on the accuracy of determinations, as compared to the use of the Census 2000 long form data.

The Census Bureau is committed to making its data as accurate as possible for all users. Thus the Census Bureau is undertaking research to explore the use of small area estimation models to reduce the variances of the ACS estimates and thus make them more precise. Small area estimation models use auxiliary information to reduce the variances of the estimates. In this case, the auxiliary information would come from ACS estimates of other areas similar with respect to LEP and educational attainment, and from demographic information from the recently completed 2010 Census. There is a long tradition of successful small area modeling to reduce the variances of estimates within the Census Bureau. The Small Area Income and Poverty Estimate (SAIPE) program is a prominent example (U.S. Census Bureau, 2011).

The small area models under consideration are hierarchical random effects (HRE) models. These models have the attractive property that, for larger areas with sufficient sample, the model-based estimates will be very close to the design-based (traditional) ACS estimates; that is, they will borrow little or no information from outside their jurisdiction. For smaller jurisdictions, HRE models allow for borrowing across jurisdictions to reduce the variances of the estimates. To facilitate such borrowing, jurisdictions are stratified by similar demographic characteristics. Note that any approach that relies on models for borrowing information runs some risk of errors arising due to model bias. This is why it is desirable to use models that leave design-based estimates for areas with large samples relatively unchanged, but that rely more on the model and its auxiliary information for areas with small samples.

This document specifies the criteria the Census Bureau will use to decide between using the design-based ACS estimates and model-based ACS estimates for all determinations. In the context of Section 203 determinations, prespecified criteria lend credibility to making changes to methods that may have an impact on public policy. The criteria will apply to the complete set of estimates as an aggregate; either all jurisdictions will have determinations based on model-based estimates or all will have determinations based on design-based estimates. Thus, if the criteria lead to use of model-based estimates, the determinations for both small and large jurisdictions will be based on model-based estimates, though for large areas the improvement in variances may be minimal.

II. The Criteria for Deciding Between Design-Based and Model-Based Estimates

We will identify those 10 percent of jurisdiction and language minority group combinations whose design-based estimates of percent voting age citizens who are LEP have the largest variances. If we see a mean reduction in variances of 20 percent or more among these combinations, then we will strongly consider the use of modeling. In addition, we will take into consideration quality measures, model diagnostics, and a review of face validity of the model results. If we see the specified reductions in variances and find no results which cause concern, we will recommend the model-based estimates. If we do not see the reductions in variance as specified, or if we see these reductions but find results which cause concern in consideration of quality measures, model diagnostics, or face validity, then we will recommend the use of the design-based ACS estimates.

III. Justification of the Decision Criteria

In this section we outline the reasons for the criteria we specified.

A. Why Consider the Variances of the Estimates?

We use variances because they are a standard measure of statistical reliability or precision, and thus are commonly understood and easy to defend. Importantly, they are clearly related to the accuracy of making determinations.

B. Which Characteristic to Use in the Criteria?

There are three measures used in the Section 203 determinations for each jurisdiction and language minority group combination: percent LEP of voting age citizens, number of voting age citizens who are LEP, and percent of voting age citizens who are LEP and have less than a 5th grade education. Of these three measures we will consider in the criteria only

the percent of the voting age population who are LEP. We restrict the criteria to this one characteristic for several reasons. Foremost, the percent LEP is a more relevant measure for smaller jurisdictions than the total LEP because small jurisdictions will not have populations of voting age citizens with LEP larger than 10,000. These smaller jurisdictions will have the higher variances and thus necessitate our focus. Furthermore, we know from the 2002 determinations that educational attainment rarely changed the determinations. In addition, the three measures will be strongly correlated. Thus, a model that reduces the variance of percent LEP should similarly reduce the variances of total LEP, of percent LEP, and of percent LEP with less than 5th grade education. Lastly, the percent reductions in variance should be close for percent LEP and total LEP for the larger counties where most of the population resides, because the total county population is controlled to a known value (U.S. Census Bureau, 2009).

C. Why Focus on a Subset of Jurisdiction and Language Minority Group Combinations?

We expect the larger jurisdictions already have sufficiently reliable design-based ACS estimates, and that the modeling will not change these much. Thus the larger jurisdictions will not benefit from a modeling approach to producing estimates. We expect that only the subset of jurisdiction/language minority group combinations with the largest variances would benefit substantially from modeling, and we want our criteria to focus on the areas whose estimates are most likely to be in need of improvement.

D. Which Values to Use as Thresholds?

The values for the cutoffs we use are based on practical considerations and an analysis of the design-based ACS data. From a practical point of view, nontrivial improvements should be demonstrated to justify changing the methodology. If we can demonstrate substantial improvements in the reliability of the 10 percent of jurisdiction/language minority group combinations with the least reliable estimates, then we believe that we have met this requirement for practical benefits. We define a substantial reduction in variance as a 20 percent or larger reduction.

We also compared the variances of estimates of LEP for jurisdictions with their population size and found natural breaking points, prior to which the variances seemed to increase rapidly with declining population size. This observation lent credence to the approach of looking only at those jurisdictions with the largest variances.

E. Why Consider Measures other than the Numerical Thresholds?

The numerical threshold of 20 percent or greater reduction in variances in the 10 percent of language minority group/jurisdiction combinations with the largest variances captures the essence of the improvements which we would require to adopt a model-based enhancement of ACS estimates. However, in any novel methodology it is conceivable that there are unanticipated results that raise concern. For this reason we also apply more general, common-sense scrutiny to our results, including standard reviews of data quality measures, model diagnostics, and a comparison to the ACS design-based estimates.

F. To Which Jurisdiction and Language Minority Group Combinations are the Criteria Applied?

We will apply the criteria to the combinations of jurisdictions and language minority groups for which the determinations will actually be made. As stated earlier, the jurisdictions will be the AIA/ANAs and all counties, except in New England states and in Michigan and Wisconsin, where the jurisdictions will be the MCDs. From these jurisdiction by language minority group combinations we will select the 10 percent with the largest variances for the proportion of voting age citizens who are LEP.

IV. Bibliography

Federal Register, (2002); Vol. 67, No. 144, Friday, July 26, 2002

http://www.census.gov/rdo/pdf/FRN_VotingRightsDeterminations.pdf

U.S. Census Bureau (2009); "Design and Methodology: American Community Survey".

http://www.census.gov/acs/www/methodology/methodology_main/

U.S. Census Bureau (2011); "Small Area Income and Poverty Estimates" webpage. <http://www.census.gov/did/www/saipc/>