

I. Background and Overview

While Congressional apportionment arguably is the most well-known use of the data collected during the decennial census, decennial census data meets many other statutory needs. Through the provisions of 42 U.S.C. § 1973aa-1a of the Voting Rights Act of 1965, Congress has mandated one such lesser known use of decennial data. Specifically, the Census Bureau Director is tasked with making certain determinations regarding the size of defined minority language populations within a given state or political subdivision for purposes of determining coverage for minority language voting assistance.¹ In determining how to best meet such needs for decennial data, the Census Bureau recognizes that effective governance, through the implementation of programs attendant to such laws as the Voting Rights Act, requires useful data. To ensure that this requirement is met, the Census Bureau considers data quality, accuracy, reliability, statistical integrity, and precise methodology. These factors informed the approach to the minority language determinations.

Superimposed on these factors were the requirements imposed by federal law, discussed more fully below, that instructs the Census Bureau Director to draw information about the defined characteristics of certain populations from the American Community Survey (ACS) and comparable census data.

The Census Bureau's effort to provide the most accurate coverage determinations possible recognizes, as it must, that the ACS estimates are subject to sampling error that may be large compared to the estimate for a population for rare characteristics and for small geographies.² Therefore, the ACS is limited in accurately identifying the statutorily-defined populations at the geographic levels required. The practical effect of this limitation is the inherent limitation in its ability to accurately measure the small populations potentially entitled to minority language assistance at the polls. As a result, the Census Bureau Director decided to use statistical modeling to address this limitation in order to provide the precise coverage determinations integral to sound implementation of the Voting Rights Act.

II. Summary and Conclusion

To obtain the necessary precision in the coverage determinations, the Census Bureau used model-based estimates of ACS data as the basis for producing coverage determinations pursuant to 42 U.S.C. § 1973aa-1a ("Section 203 determinations"). The decision to use model-based estimates resulted from a comparison of variances³ of estimates from the model with the variances of the direct ACS estimates.⁴

¹ 42 U.S.C. § 1973aa-1a (2006 & Supp. 2011).

² A "rare" characteristic is a characteristic that is not prevalent within the sampled population. Rare characteristics or smaller geographies are subject to more relative sampling error when using direct estimates.

³ Variance and standard deviation measure the distribution of the values in a dataset, i.e. explain the distance between the values for a characteristic of interest in a population. The variance provides information on how far a value deviates from the mean, or average, of the dataset. The standard deviation provides average distances from

These direct estimates commonly are referenced as “design-based ACS estimates.” Upon making the comparison, the Census Bureau achieved a mean variance reduction of 61% among the 10% of design-based estimates with the largest variances. The median variance reduction for these estimates was 83%. The model-based variances were lower for more than 75% of all estimates. Furthermore, a quality review yielded no cause for concern with either the model-based estimates or with the variance estimation methodology used to make the decision. The comparison showed the model-based estimates substantially reduced the variances, improving the accuracy of the estimates and, therefore, the accuracy of the determinations.

In reaching the decision to use modeling, the Census Bureau established variance comparison criterion that recognized three circumstances, based on experience from the ACS and SAIPE programs. First, large jurisdictions with large samples are unlikely to produce large variances using either model-based estimates or design-based estimates. Second, jurisdictions with small samples tend to produce large variances using design-based estimates. Third, model-based estimates rely on more assumptions than design-based estimates.⁵

As a result of the first two considerations, the Census Bureau decided to target jurisdictions with small samples and large variances, in order to gauge the level of improvements achieved in estimates most needing remediation in terms of margins of error. Thus, the Census Bureau identified the 10 percent of jurisdiction⁶ and language minority group⁷ combinations where design-based estimates of percent voting age citizens with limited English proficiency (LEP) have the largest variances. In order to combat any potential for the modeling to provide optimistic results, the Census Bureau required that the model-based estimates produce a substantial reduction in the variances of the design-based estimates. A mean reduction in variances of 20 percent or more among the jurisdiction and language minority group combinations identified above constituted a substantial reduction. In addition to the criterion regarding variance reduction, the Census Bureau examined several quality measures of the estimates. These quality measures were model diagnostics and a review of face validity of the model results. If these quality measures had revealed concerns regarding the estimates produced by the models, the Census

the mean, defining the range of values into which the data falls on one or the other side of the mean. Thus, the standard deviation is a way to determine the values that are normal or expected and to identify the values that fall above or below that expectation. Variance and standard deviation apply to the statistical analysis of data collected from an entire population or from a sample. In contrast, a margin of error is a measure introduced by the analysis of data collected in a sample surveys, measuring how close the sample results (i.e. the estimate) represent the entire population (or the maximum amount by which the results of the sample are expected to differ from the results of a survey of the entire population). The margin of error may be considered a derivative of the variance. Thus, a reduction in the variance will result in the reduction of the margin of error.

⁴ The Census Bureau releases ACS direct estimates annually as a standard Census Bureau data product. See http://www.census.gov/acs/www/data_documentation/2009_release/.

⁵ See Appendix A.

⁶ The jurisdictions under consideration in this decision were counties and minor civil divisions (MCDs).

⁷ There are 68 language minority groups for which language assistance may be provided: Hispanic/Latino/Spanish, 16 Asian groups, and 51 American Indian and Alaska Native groups.

Bureau would have relied on the ACS design-based estimates, even if the model-based estimates met the stipulated reductions in variances.

III. Legal Requirements

Since 1975, Congress has mandated, through Section 203(b) of the Voting Rights Act of 1965, that states and political subdivisions could no longer provide election materials only in the English language, if the Census Bureau Director made certain determinations regarding the existence of defined minority language populations within a given state or political subdivision.⁸

Beginning in 1982, Congress continued to refine the parameters that the Census Bureau Director must use in making these determinations. Among these refinements, Congress required the use of decennial census data and established certain thresholds for coverage.⁹ As part of the enactment of the Fannie Lou Hamer, Rosa Parks, and Coretta Scott King Voting Rights Act Reauthorization and Amendments Act of 2006, Congress amended Section 203(b) of the Voting Rights Act to reflect the implementation of the American Community Survey in 2006.¹⁰ As a result of this enactment, Section 203(b) was amended to read:

(1) Generally

Before August 6, 2032, no covered State or political subdivision shall provide voting materials only in the English Language.

(2) Covered States and political subdivisions

(A) Generally

A State or political subdivision is a covered State or political subdivision for the purposes of this subsection if the Director of the Census Bureau determines, based on the 2010 American Community Survey census data and subsequent American Community Survey data in 5-year increments, or comparable census data, that –

(i)(I) more than 5 percent of the citizens of voting age of such State or political subdivision are members of a single language minority and are limited English-proficient;

(II) more than 10,000 of the citizens of voting age of such political subdivision are members of a single language minority and are limited-English proficient; or

(III) in the case of a political subdivision that contains all or any part of an Indian reservation, more than 5 percent of the American Indian or Alaska Native citizens of

⁸ An Act to Amend the Voting Rights Act of 1965, P.L. 94-73, Title III, § 301, 89 Stat. 402 (1975).

⁹ Voting Rights Act Amendments of 1982, Public Law 97-205, § 4, 96 Stat. 131 (1982).

¹⁰ P.L. 109-246 § 8, 120 Stat. 581 (2006).

voting age within the Indian reservation are members of a single language minority and are limited-English proficient; and

(ii) the illiteracy rate of the citizens in the language minority as a group is higher than the national illiteracy rate.¹¹

This amendment reflects the current statutory language relevant to the Census Bureau.¹² As a result, the upcoming determinations will be made under the language made current by the 2006 amendments. The language of this provision compels the Census Bureau Director to use five-year ACS and census data sets, particularly when considered in light of the legislative history of this 2006 amendment to the Voting Rights Act. In H.R. Rep. 109-478, the Committee of the Judiciary explains the 2006 revision of Section 203(b)(2)(A) as follows:

In reauthorizing Section 203, the Committee was made aware that the United States Census Bureau has changed its data collection methods, eliminating the use of the long-form questionnaire. The long form questionnaire was the detailed document used by the Census Bureau every decade to gather demographic, housing, and social information and upon which determinations with respect to Section 203 were made. In its place, the Census Bureau has been issuing the American Community Survey (ACS). The ACS is a detailed survey conducted by the Census Bureau that will be updated on a rolling basis annually. While issued to a smaller universe of households more frequently, the survey is designed to reach a greater number of individuals per decade (because it is issued more frequently), providing a more detailed look at communities on a more timely basis. In comparing the relevant questions from the long form questionnaires, with the questions provided on the ACS, the Committee finds no substantive difference between the two documents and expects that the ACS will be a suitable substitute for the outdated long form questionnaire. In addition to identifying the ACS as the basis upon which the Director of the Census is required to make Section 203 determinations, the Committee also finds a benefit to reliance on survey results published on a more timely basis, as it better reflects our country's rapidly changing makeup, as opposed to a reliance on results published only every decade. To reflect this need, Section 203 was also amended to require the Director of the United States Census to make determinations every 5 years based on a rolling 5 year average.¹³

Congress enacted, without modification, the revision to Section 203(b)(2)(A) reported out of the House Committee of the Judiciary.¹⁴ The Committee's Report shows that Congress updated this language based on an examination of the content of the ACS and the Decennial long-form questionnaire.¹⁵ This examination is evidence of the congressional intent to define the required data source for Section 203

¹¹ *Id.*

¹² See 42 U.S.C. § 1973aa-1a (2006 & Supp. 2011).

¹³ H.R. Rep. 109-478, at 44 (2006).

¹⁴ Compare H.R. Rep. 109-478, at 6 with 42 U.S.C. § 1973aa-1a (2003 & Supp. 2011).

¹⁵ See H.R. Rep. 109-478, at 44 (2006).

determinations as well as further define the relevant population parameters. In identifying the ACS as a substitute for the long form decennial questionnaire that gathered information “upon which determinations with respect to Section 203 were made,” Congress stated their understanding that the use of the decennial data was well-established.¹⁶

A review of the legislative history of amendments to Section 203(b) of the Voting Rights Act shows that Congress has maintained constant oversight over the Census Bureau’s determinations under Section 203, including remaining abreast of the methodology and the data used by the Bureau. This oversight indicates that Congress expects the Bureau to use only those data sources of which it is aware and has approved. Neither the language of Section 203(b) nor the lengthy legislative history of the minority language assistance provisions suggest that Congress intended the use of any data outside of that data derived from the decennial census.¹⁷

IV. Using the American Community Survey Data to Make “Section 203 Determinations”

Given the Congressional will expressed in the 2006 amendment and evidenced by the House Judiciary Committee Report, the Census Bureau Director was tasked with determining coverage this year based on particular characteristics of defined groups taken from two sources: (1) the ACS and (2) the 2010 Census.¹⁸ Although the 2010 Census is designed to collect information from every person in the United States, the 2010 Census does not collect all of the necessary population characteristics required for the determinations.¹⁹ The detailed questions concerning citizenship, language usage, educational attainment, and English proficiency are collected by the ACS survey.²⁰ The ACS is designed and implemented as a sample survey that collects information from a representative group of individuals to provide direct estimates about different characteristics of the population of the United States.²¹ Therefore, the characteristics required for the coverage determinations are obtained only for this representative sample of individuals within a state or political subdivision. To meet his statutory mandate, the Census Bureau Director must determine the proportion or number of members of voting age citizens within a state or political subdivision who are members of a single language minority, have less than a fifth grade education, and are limited in English proficiency only from the data collected in

¹⁶ See *id.* The “Nickles Amendment” in 1982 introduced an explicit directive to the Census Bureau Director. The language of the amendment instructed the Census Bureau Director to make determinations “based on 1980 and subsequent Census data.” Voting Rights Act Amendments of 1982, Public Law 97-205, § 4, 96 Stat. 131 (1982). By alluding to “subsequent Census data,” the Nickels amendment established, as a continuing statutory requirement, the use of decennial census data for making coverage determinations. Congress would state this intent more directly in the House Judiciary Committee discussion of the 2006 amendments to the Voting Rights Act of 1965. See H.R. Rep. 109-478, at 44.

¹⁷ As set forth in the legislative history of the 2006 amendments, Congress understood the ACS to be a replacement for the long-form decennial questionnaire. H.R. Rep. 109-478, at 44. Although a separate survey, the Census Bureau is authorized to undertake the ACS as part of its decennial program. 13 U.S.C. §§ 141, 193 (2006).

¹⁸ See Appendix A. The determinations used the 2005-2009 ACS five-year estimates, weighted with population data from the 2010 Census. Weighting assists in reducing reliance on the model to estimate values.

¹⁹ http://www.census.gov/acs/www/about_the_survey/american_community_survey_and_2010_census/

²⁰ *Id.*; see also http://www.census.gov/acs/www/methodology/questionnaire_archive/

²¹ See http://factfinder.census.gov/jsp/saff/SAFFInfo.jsp?_pagelid=sp1_acs&_submenuId=

the ACS and the 2010 Census. Thus, The Census Bureau Director must make determinations regarding all states and political subdivisions from a representative sample of individuals, and not “observed” data from all individuals within these jurisdictions.

The necessity of relying on data from a sample to make the determinations is not a new occurrence; however, use of the ACS data is novel. The long-form decennial questionnaire was sent to an average of one out of every six households during the decennial census.²² This one-in-six average provided a larger sample than the ACS, which averages one in-eight households in the 5-year samples.²³ As a result of the smaller sample size, the ACS 5-year estimates have larger variances. Thus, the ACS 5-year estimates have larger margins of error than the long-form estimates, particularly for determinations involving the small populations defined in Section 203.²⁴

V. Translating the Requirements of Section 203(b) Into Statistical Determinations

As a result of the potential statistical issues implicated by this new use of the ACS data, the Census Bureau began outlining the requirements of Section 203(b) and researching ways to translate those statutory requirements into a methodology that would produce precise estimates for the small populations delineated in 42 U.S.C. §1973aa-1a. The Census Bureau reviewed the previous determinations, published in the Federal Register in 2002, to help inform its approach to the current determinations. These previous determinations used direct estimates of the data collected on the long-form questionnaire. Because of the reduction in sample size that occurred when the Census Bureau moved from the long-form questionnaire to the ACS, the Census Bureau believed that the use of direct estimates would not yield precise results.²⁵ As a result, the Census Bureau began considering the use of and researching for an alternative to direct estimation. Because the coverage determinations would use ACS data to draw conclusions about small populations and geographies, the Census Bureau tapped knowledge developed through the Small Area Income and Poverty Estimates (SAIPE) program to begin formulating its approach to the coverage determinations. SAIPE uses statistical models applied to ACS direct estimates, decennial data, and administrative records data to produce poverty estimates for states, counties, and school districts to meet various federal mandates.²⁶

²² *A Compass for Understanding and Using American Community Survey Data: What Federal Agencies Need to Know*, U.S. Government Printing Office, Washington, D.C., 2008, at 5, available at <http://www.census.gov/acs/www/Downloads/handbooks/ACSFed.pdf>.

²³ *Id.*

²⁴ *Id.*; see also *supra* note 19 (providing additional information on ACS sampling and methodology). Additionally, the ACS design-based estimates experience higher variances than long-form decennial estimates because the ACS results are controlled to the intercensal population estimates, instead of being controlled to the detailed population figures obtained from the decennial census, as were the long-form decennial results.

²⁵ See *supra* note 23 (change in population figures used to control ACS results also contributed to increased variances in ACS direct estimates).

²⁶ See <http://www.census.gov/did/www/saipe/about/index.html>. SAIPE modeling and estimation differ from the modeling used to make the Section 203 determinations. SAIPE uses administrative records and other non-survey data to form continuous variables to use in the models. Because 42 U.S.C. §1973aa-1a speaks only to the ACS and

Because of the similarities between the data sets and geographies involved in the SAIPE program and those data sets and geographies relevant to the determinations of coverage under Section 203(b), the Census Bureau considered hierarchical random effects (HRE) modeling, as an appropriate alternative to using the direct ACS estimates for making the Section 203 coverage determinations. Hierarchical random effects modeling allows for the cross-sectional comparison of certain populations nested within larger populations,²⁷ while permitting the use of certain qualities or explanatory variables about the populations at any level of nested populations. HRE models used in a survey context also can take advantage of the most effective aspects of both direct estimates and models. Where ACS sample is sizeable, the estimates resulting from an HRE model will be weighted heavily toward the direct estimates. Where ACS sample size is small, an HRE model permits the addition of explanatory variables in combination with the design-based survey estimates to result in an estimate where the variance is smaller than the original design-based survey estimate. As a result, HRE models are proficient in handling different concentrations of ACS sample across jurisdictions. Additionally, a HRE model allows for a consistent modeling approach across all jurisdictions, while addressing the imprecision of the design-based estimates for small sample sizes.

In order to test the efficacy of the model for making the 203 determinations, the Census Bureau established measures for implementation. Since the potential use of modeling arose as a way to reduce the margins of error associated with small ACS samples, the Census Bureau concluded that the model must provide estimates with lower variances than the variances resulting from the design-based ACS estimates. To ensure confidence that the model provided a meaningful improvement, the Census Bureau decided that the model would need to produce a substantial reduction in the variances of the design-based estimates having the largest variances to merit use of the modeling results for making the determinations. The Census Bureau expressed this need as requiring that the model produce at least a 20% mean variance reduction in the 10% of jurisdiction and language minority group combinations having design-based ACS estimates of percent voting age citizens of limited English proficiency (LEP) with the largest variances.²⁸

comparable census data, the Census Bureau did not include administrative records or other non-survey data as a source of data for making the coverage determinations. Instead, the Census Bureau used stratification to perform shrinkage estimation. See Appendix A.

²⁷ Although not populations utilized in making the Section 203 determinations, an example of nested populations would be voters within districts, districts within counties, counties within states.

²⁸ The selection of the 10% subset of the combinations with the largest variances recognized certain circumstances and observations. First, this 10% subset represented design-based estimates with the most potential for improvement through the use of the model-based approach. Under the HRE model, estimates for large jurisdictions would vary little, if at all, from the design-based estimates and the model would not be expected much improve estimates based on large samples. Second, the limitation of the subset characteristics to the percent voting age population with limited English proficiency recognized that small areas will not have a voting age, LEP population that meets the 10,000 person criteria of Section 203(b). Third, the limitation to these characteristics reflects an examination of the previous Section 203 determinations that showed that the other statutory criteria rarely impacted coverage. Additionally, a comparison of the variances of estimate of LEP for jurisdictions with the population size of the jurisdiction found rapidly increasing variances with declining population size. See Appendix A.

With this criteria established, the Census Bureau began constructing the model. In order to construct the model, the Census Bureau identified data from the ACS and the 2010 Census that related to the population qualities delineated in Section 203(b). Specifically, the Census Bureau identified information from ACS estimates related to language spoken at home, English-speaking ability, citizenship, age, race, and ethnicity, and educational attainment and identified age, race and ethnicity from the 2010 Census as data indicative of the population parameters set forth in Section 203(b). After selection of the relevant data, the Census Bureau assembled this data in a manner consistent with the calculation described by the language of the statute. This combined data then was stratified, at a jurisdictional level, by a separately modeled rate of citizenship and limited English proficiency for each minority group enumerated in the statute.²⁹ The constructed stratification was then used as input into the HRE model.³⁰

VI. Decision to Use the Model to Make the “Section 203 Determinations”

Upon making the comparison between the variances produced by the design-based estimates and the variances produced by the HRE model-based estimates, the Census Bureau achieved a mean variance reduction of 61% among the 10% of estimates with the largest variances. The median variance reduction of these estimates was 83%. The model-based variance was lower in more than 75% of all estimates. Furthermore, a quality review yielded no serious causes of concern with either the model-based estimates or with the variance estimation methodology used to make the decision.

The comparison showed the model-based estimates substantially reduced the variances, and therefore, the margins of error, associated with the use of design-based estimates. As a result of these large reductions that satisfied the variance criteria, the Census Bureau Director decided that the model-based estimates would be used to make the coverage determinations pursuant to 42 U.S.C. § 1973aa-1a.

²⁹ See Appendix A. Generally, stratification is used to organize populations with several characteristics into independent “subpopulations” that share one of the characteristics.

³⁰ See *id.*