

Appendix A

Technical Appendix to the Director's Recommendation to Use the Model-Based Estimates for Producing Section 203 Determinations

I. Introduction

This document summarizes the statistical findings and technical foundation for the Steering Committee's recommendation to use the model-based estimates as the basis for producing determinations of Section 203 coverage. There are two key technical elements that have led to the recommendation.

- Of greatest salience are the observed reductions in variance of the relevant estimates, i.e., the LEP rates of the CVAP populations for the jurisdiction by language minority group (LMG) combinations. These reductions well exceed the thresholds of the pre-specified criteria.
- The checks of validity of the estimates reveal no results that undermine confidence in the model-based estimates. Further, the model-based estimates differ from the design-based estimates in interpretable ways that assure the soundness of the model-based estimates.

In addition, all steps that went into the modeling and the production of determinations were independently verified through double programming. This step was implemented as part of using best practices in conducting statistical research, in particular, meeting the standard of "reproducibility." The rest of this document provides more detail in support of the first two points.

II. Results on the Reduction of Variances of LEP Rates to Apply the Decision Criteria

The reductions of variance largely exceed the pre-specified minimum variance reduction to the percent LEP of voting age citizens (CVAP) of LMG for jurisdictions. The criterion required a greater than 20 percent reduction in variances of the estimates of the percent LEP with the largest variances. There were 14,423 jurisdiction/LMG combinations for which both the design-based and model-based estimates of the LEP percentage were non-zero. For the 10 percent of the design-based estimates with the largest variances, we computed the mean variance reduction for model-based estimates. The variance reduction is the percent difference between the design and model-based variance estimates. Among these 10 percent of estimates with the largest variances, 1,442 cases, the mean percent reduction was 61 percent, which well exceeds the 20 percent pre-specified criterion. See the results for the first decile in Table 1 below. Furthermore, the median reduction was 83 percent and the model-based variance was lower in more than 75 percent of all jurisdiction/LMG estimates.

The Census Bureau also looked at the complete distribution by examining the decile slices of the set of 14,423 estimates. These slices were created by ordering the estimates by decreasing variance of the design-based estimates and partitioning them into ten equal slices. The median percent reductions were greater than 60 percent in all the deciles. We found that in all deciles, at least 75 percent of the estimates had some variance reduction. The mean percent reductions

were greater than 20 percent in the seven deciles with the highest estimates of variances. These mean reductions steadily decrease, when going from the deciles with the highest variances to those with the lowest, due to the effect of extreme percent increases (negative reduction), with the bottom deciles having variance increases. Excluding extreme values of percent increase (more than 500 percent increase), all deciles except the lowest one have a mean percent reduction greater than 20 percent. See Table 2.

Table 1: Variance Reduction among Decile Slices of Estimates
(By descending size of ACS Variance)

	Estimates	Mean	95 th Percentile	75 th Percentile	Median	25 th Percentile	5 th Percentile
1 st Decile	1,442	61.15	99.77	97.24	83.16	40.08	-32.80
2nd Decile	1,442	61.45	99.72	96.81	86.93	53.54	-38.63
3rd Decile	1,442	57.91	99.66	97.29	85.05	41.80	-48.36
4th Decile	1,443	46.18	99.48	97.33	84.31	26.63	-82.49
5th Decile	1,442	52.91	99.28	96.32	83.99	39.89	-83.91
6th Decile	1,442	49.76	99.03	94.86	81.22	39.27	-84.11
7th Decile	1,443	31.12	98.61	92.53	78.01	33.18	-140.54
8th Decile	1,442	19.33	98.62	90.67	70.09	26.39	-171.63
9th Decile	1,442	-31.30	98.44	88.89	65.48	-3.57	-369.91
Bottom Decile	1,443	-180.74	98.19	86.66	56.82	-33.69	-707.67

Approximately 17% of all estimates (2,502) had a model-based variance that was larger than the design variance. For these estimates, both the variance and the estimate itself tended to be small. The mean variance of the design-based estimate was 0.09, with a median of 0.0007. The mean design-based estimate of LEP was 0.75 percent, with a median of 0.05 percent. Note that these areas would not have been covered if the determinations had been based on the design estimates and the increase in variance, although not-negligible, is irrelevant in terms of Section 203 coverage.

Table 2: Variance Reduction among Decile Slices of Estimates by Descending Size of ACS Variance. Excludes Extreme Values of Variance Reductions (<-500%)

	Estimates	Mean	95 th Percentile	75 th Percentile	Median	25 th Percentile	5 th Percentile
1st Decile	1,442	61.15	99.77	97.24	83.16	40.08	-32.80
2nd Decile	1,441	64.90	99.72	96.81	86.95	53.61	-37.81
3rd Decile	1,440	59.12	99.66	97.30	85.11	41.82	-46.69
4th Decile	1,434	52.41	99.49	97.34	84.90	27.30	-75.65
5th Decile	1,439	55.20	99.28	96.33	84.02	40.10	-81.61
6th Decile	1,436	53.68	99.04	94.90	81.45	40.39	-73.61
7th Decile	1,420	50.20	98.69	92.77	78.82	36.84	-85.60
8th Decile	1,418	41.84	98.65	90.91	71.14	29.62	-130.89
9th Decile	1,393	24.83	98.55	89.23	68.28	5.01	-203.95
Bottom Decile	1,331	18.47	98.30	88.27	63.31	0.42	-248.25

III. Results Examining the Validity of the Modeled Estimates and Resulting Determinations

We took several approaches to investigate the validity and reasonableness of the model-based estimates. These approaches are:

- We considered those states and jurisdictions which were covered by the model-based estimates, but not the design-based estimates, as well as those jurisdictions which would have been covered by the design-based, but not the model-based, had we elected to use the design-based. That is, we looked at the overlap between the two sets of determinations. A small level of overlap would be an indication of too much disparity between the two sets of estimates and (perhaps) that the stratification scheme did not achieve its intended and desired results. Note that we made this investigation blind to geography and looked only at gross aggregates of LMGs.
- We confirmed that the model-based estimates of voting age population equaled the 2010 Census counts, which should be the case because they are controlled to the 2010 Census population counts. (The design-based estimates are controlled to 2009 population estimates which are based on the 2000 Census).
- We looked at the proportion of model-based estimates of LEP that were within the 90% and 99% confidence intervals (CI) of the design-based estimates. We used this method to identify extreme results. We then examined these outliers more closely.
- We looked at the LEP rate by sample size to examine and assess any indication of selection bias in the decision criteria.

IV. Comparing Determinations Based on Model-based and Design-based Estimates

We compared the determinations based on design-based estimates with those based on model-based estimates. This comparison did not include the geography (or the identity of the jurisdiction) of any potential determinations. The goal of making such comparisons was to examine the plausibility of the different determinations. Any differences emerging from the comparison should be interpreted in ways that assure the soundness of the model-based estimates. Several tables which focus on the differences in determinations made by model-based

and design-based estimates follow. Table 3 shows county/MCD by LMG combinations covered by the model-based estimates, but not by the design-based (excluding those covered because of an American Indian Area/Alaska Native Area)¹. Table 4 shows the AIA/ANAs which implied coverage for counties and MCDs for model-based estimates but not design-based ones. Table 5 shows the county/MCD by LMG combinations for which the design-based are covered, but not the model-based (excluding those covered because of an American Indian Area/Alaska Native Area). Table 6 shows the AIA/ANAs which implied coverage for counties and MCDs for design-based estimates but not model-based ones. Some observations follow.

1. There are two LMG/state combinations which are covered by both the design-based and model-based estimates. (The state comparisons are not presented in any of the tables).
2. There is one state/LMG combination covered only by the model-based estimates. We note that this state coverage leads to no county/MCD by LMG combination being covered that would not have been covered on its own; that is, all covered county/MCD by LMG combinations in this state met the 5% LEP or 10,000 LEP and the illiteracy criteria.
3. There are 270 LMG county/MCD combinations which are covered by both the design-based and model-based estimates. These combinations are not shown in either table.
4. There are 38 combinations covered by the model-based estimates, but not the design-based estimates. The 34 not triggered by AIA/ANAs are listed in Table 3. Table 4 shows those AIA/ANA areas that triggered coverage for the model-based estimates, but not for the design-based. Recall that the AIA/ANA areas trigger coverage in the jurisdictions in which the area is located.
5. There are 40 combinations covered by the design-based estimates, but not the model-based estimates. The 22 are listed in Table 5. Table 6 shows those AIA/ANA areas that triggered coverage for the design-based estimates, but not for the model-based estimates.
6. Of the 38 combinations covered by the model-based estimates only, 16 are close to the 10,000 LEP/CVAP cutoff, and the LEP rate is below 5 percent for all 16. In contrast, there are no such cases among the 40 combinations covered only by the design-based estimates. We believe that this is attributable to the 2010 Census controls to which the model-based estimates are controlled, which are higher than the controls for the 2005-2009 ACS 5-year estimates. The 2005-2009 ACS 5-year estimates are controlled to an average of the 2005-2009 Population Estimate Program (PEP) estimates, which are based on the Census 2000 counts updated with births, deaths, and migration administrative records. This is confirmed by the larger counts for voting age population seen for these jurisdiction/LMG combinations for the model-based estimates. Table 7 shows the voting age population for the modeled- and design-based estimates for the 16 jurisdictions.

¹ The geography of the jurisdictions is not shown as we conducted this part of the investigation without awareness of the geographies involved.

Table 3: Jurisdiction by Language Minority Group Combinations Covered by Model-based but not Design-based (Excluding four covered because of American Indian/Alaska Native Areas)

Voting Age Citizens LEP ²	LEP Rate	Illiteracy Rate ³	Jurisdiction Coverage ⁴	State Coverage
39	6.67	12.82	1	
57	6.92	15.79	1	1
83	5.27	12.05	1	
92	6.23	13.04	1	
107	5.43	13.08	1	
207	10.50	3.86	1	
209	9.01	18.66	1	1
286	5.00	10.49	1	
352	6.41	11.65	1	1
690	5.46	4.64	1	
1,688	5.00	8.29	1	
2,323	6.56	10.12	1	
2,568	5.45	9.54	1	
2,686	5.02	8.94	1	
3,861	5.52	10.85	1	
4,509	5.19	15.37	1	1
4,563	5.08	15.54	1	1
5,897	5.39	6.92	1	
10,128	1.55	10.07	1	
10,142	0.18	10.45	1	
10,203	0.18	5.33	1	
10,381	1.69	7.27	1	
10,388	3.46	13.35	1	
10,541	0.86	3.42	1	
10,598	0.53	8.70	1	
10,703	0.47	5.23	1	
10,704	4.22	9.99	1	
10,739	2.43	9.20	1	
10,752	2.71	13.13	1	
10,780	1.81	2.04	1	
10,805	1.13	3.92	1	
10,867	1.18	16.29	1	
10,982	1.15	6.89	1	
12,078	3.31	17.81	1	

² Voting age citizens who are LEP, LEP rate, and illiteracy rate are calculated with the model-based estimates.

³ Rates are expressed as percentages.

⁴ A “1” indicates coverage.

7. Among both the 38 and 40 combinations we see either the number of LEP and CVAP persons close to 10,000, or the LEP rate close to 5% of the CVAP population. Of the 38 combinations covered by the model-based estimates (but not the design-based), eight range from 1,688 to 5,897 CVAP persons while having LEP rates ranging from 5.02 to 6.56 percent.
8. Of the 40 combinations covered by the design-based estimates (but not the model-based), 18 are covered because of an AIA/ANA area within their boundaries, and have only a small number of LEP and CVAP persons. None are close to the 10,000 LEP cutoff, whereas all are not far from the 5% LEP cutoff. Note that these 18 are not listed in Table 5, but are implied by the 20 AIA/ANAs listed in Table 6.

Table 4: American Indian/Alaska Native Areas Triggering Coverage by Model-based but not Design-based Estimates⁵

Voting Age Citizens LEP ⁶	LEP Rate	Illiteracy Rate
14	5.69	21.43
16	5.99	6.25
17	5.38	5.88
21	6.38	4.76
44	5.50	15.91
44	5.50	15.91
47	8.03	6.38
140	6.56	5.71

⁵ These eight American Indian/Alaska Native Areas require coverage in only four county or MCD jurisdiction by LMG combinations as several LMGs were covered more than once within a given jurisdiction.

⁶ Voting age citizens who are LEP, LEP rate, and illiteracy rate are calculated with the model-based estimates.

Table 5: Jurisdictions by Language Minority Group Combinations Covered by Design-based but not Model-based Estimates (Excluding 18 covered because of American Indian/Alaska Native Areas)

Voting Age Citizens LEP ⁷	LEP Rate	Illiteracy Rate	Jurisdiction Coverage	State Coverage
22	5.93	90.91	1	
28	5.01	21.43	1	
29	5.01	6.90	1	
46	7.18	21.74	1	
57	5.80	5.26	1	
61	8.44	0.00		1
71	5.55	23.94	1	
83	5.75	80.72	1	
90	5.93	11.11	1	
100	5.08	53.00	1	
110	5.68	31.82	1	1
126	5.60	53.17	1	1
146	6.79	8.22	1	
180	5.95	37.22	1	1
193	6.21	10.88	1	
290	5.46	53.79	1	1
325	5.69	32.00	1	1
502	5.12	39.64	1	1
571	5.07	10.51	1	
1,214	6.02	19.19	1	1
2,992	5.05	21.26	1	1
5,179	5.13	7.57	1	

⁷ Voting age citizens who are LEP, LEP rate, and illiteracy rate are calculated with the design-based estimates.

Table 6: American Indian/Alaska Native Areas Triggering Coverage by Design-based but not Model-based Estimates⁸

Voting Age Citizens LEP ⁹	LEP Rate	Illiteracy Rate
3	50.00	100.00
3	5.08	100.00
4	6.45	100.00
5	38.46	100.00
5	6.76	100.00
6	11.54	50.00
6	60.00	50.00
7	17.50	100.00
7	6.73	42.86
7	6.73	42.86
8	10.26	25.00
9	6.16	100.00
10	7.30	40.00
12	20.34	33.33
14	7.25	28.57
23	7.19	13.04
23	8.52	17.39
24	6.30	33.33
26	9.03	11.54
26	9.03	11.54

⁸ These 20 American Indian/Alaska Native Areas require coverage in only 18 county or MCD jurisdiction by LMG combinations as several LMGs were covered more than once within a given jurisdiction.

⁹Voting age citizens who are LEP, LEP rate, and illiteracy rate are calculated with the model-based estimates.

Table 7: Jurisdictions Covered by the Model-based but not the Design-based by Meeting the 10,000 LEP/CVAP Criterion

Language Minority Group	LEP Voting Age Citizens ¹⁰	Model—Based Voting Age Population	Design—Based Voting Age Population
Hispanic	10,128	107,234	94,409
Other Asian - Not specified	10,142	42,006	13,028
Asian Indian	10,203	73,000	64,849
Hispanic	10,381	77,214	70,782
Hispanic	10,388	40,422	31,850
Filipino	10,541	78,459	60,845
Chinese (including Taiwanese)	10,598	48,473	43,684
Chinese (including Taiwanese)	10,703	39,844	36,874
Hispanic	10,704	40,422	31,128
Hispanic	10,739	86,487	80,793
Hispanic	10,752	63,206	56,637
Korean	10,780	44,311	38,115
Filipino	10,805	74,244	67,757
Chinese (including Taiwanese)	10,867	36,097	31,869
Vietnamese	10,982	25,403	22,081
Hispanic	12,078	47,518	39,591

Table 8 shows the breakdowns of the number of jurisdictions covered by the design-based and model-based estimates broken down by three language minority group (LMG) or major race/ethnicity classification, Hispanic, American Indian/Alaska Native (AIAN), and Asian.

- We see there were 67 combinations of jurisdiction/LMG covered for AIAN LMGs by design-based estimates, but only 53 covered by the model-based. Recall that we see in Table 6 that the 20 jurisdictions covered for AIAN by the design-based estimates only have small LEP/CVAP populations ranging from 3 to 26 persons. This decrease is reassuring, as it is these types of small estimates which we hoped to improve with the modeling approach.
- We also see that there are eleven more jurisdictions covered by the model-based for Asian LMGs. In Table 7, we see that nine of these had larger voting age populations for the LMG in the 2010 Census and were pushed above the 10,000 criterion for that reason.

Table 8: Number of Jurisdiction by LMG Combinations Aggregated by Three LMG Groupings

Estimate	Hispanic	AIAN	Asian
Design	211	67	32
Model	212	53	43

¹⁰ Voting age citizens who are LEP is calculated with the model-based estimates.

V. Examining Differences between the Design-based and Model-based Estimates of Percent LEP

We compared the estimates of percent LEP based on design-based estimates with those based on model-based estimates. We made two types of comparisons: direct comparisons of the estimates, and calculations of the proportion of model-based estimates within the 90% confidence interval of the design-based estimates. If either the design-based or model-based estimates of LEP are zero for a jurisdiction by LMG combination, then we exclude that combination from these analyses. This treatment is consistent with the way we handle these cases in the calculation of the decision criteria. Table 9 summarizes the differences. Some observations follow.

1. The differences presented in the Table 9 are for the design-based minus the model-based estimates.
2. The mean and median differences in LEP are small; even at the MCD level where they are largest, they are 0.12% and 0.045%.
3. The differences are closely concentrated around the median with respect to the tails. We found that the box plots we produced were not revealing of additional structure within the distribution.
4. The largest 5th and 95th percentile differences in magnitude were found with MCD, with -0.31% and 0.77%. State and county estimates had smaller 5st and 95th percentile estimates.

Table 9: Percentiles of Differences between Design-based and Model-based Estimates of Percent LEP

	Mean	5th Percentile	25th Percentile	Median	75th Percentile	95th Percentile
State	-0.01284	-0.05983	-0.00836	-0.00112	0.00007	0.00342
County	0.00745	-0.19214	-0.02564	0.00017	0.02697	0.22032
MCD	0.12435	-0.31091	-0.03578	0.04508	0.19386	0.76743

In Table 10, we see that model-based estimates of LEP for LMG/state combinations are generally within the 90-percent confidence interval (CI) of the design-based estimates. 56.5 percent of state estimates, 79.2 percent of county estimates, 88.7 percent of MCD estimates, and 78.4 percent of AIA/ANA estimates are within these CIs. We investigated the proportion voting age population for the LMG over the total voting age population to determine if the new 2010 Census controls were driving these differences. However, we found little differences in these proportions. Regarding the 43.5 percent of model-based state by LMG combinations outside the CI, this is not a concern since the difference is very small as seen in Table 9, with mean and median differences of -0.1284 and -0.00112 of a percent respectively.

Table 10: Counts and Percentages of Model-based LEP Estimates within the 90% Confidence Interval of the Design-based Estimates

	Within the Confidence Interval	Outside of the Confidence Interval	Percentage within the Confidence Interval
State	856	658	56.5
County	10,754	2,827	79.2
MCD	3,703	470	88.7

VI. Variance Reduction by Sample Size of the Estimates

There was concern that by selecting for our decision criterion those estimates with largest variances that the criterion was subject to a selection bias. As an alternative that would not be subject to such a potential bias, we chose to consider the 10% of estimates with the smallest sample sizes. If we saw comparable reductions in variance among these we would be reassured of the robustness or lack of sensitivity of our decision criterion to selection bias. This analysis is presented as one component of the tests for the face validity of the estimates.

In Table 11, we see reductions in variance for the 10 percent of estimates with the smallest samples comparable to those of the largest variances. The 59.3 percent mean reduction easily exceeds the threshold of 20 percent and is close to the 61.2 percent reduction seen earlier (in Table 1), when looking at the reduction among the estimates with the largest variances. The median variance reduction is also similar. Further, the pattern is consistent in the other deciles of sample size (not shown in the table), the first five of which meet 20 percent the threshold for variance reduction. These results suggest that the decision criterion was not afflicted by selection bias.

Table 11: Percent Variance Reduction in the Decile with the Smallest Sample Size

Estimates	Mean	95 th Percentile	75 th Percentile	Median	25 th Percentile	5 th Percentile
1,443	59.28	99.62	97.28	89.22	62.19	-36.81