



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

September 12, 2019

2018 AMERICAN COMMUNITY SURVEY RESEARCH AND EVALUATION REPORT MEMORANDUM  
SERIES # ACS18-RER-01

DSSD 2018 AMERICAN COMMUNITY SURVEY MEMORANDUM SERIES #ACS18-MP-06

MEMORANDUM FOR American Community Survey Research and Evaluation Workgroup

From: Donna M. Daily  
Chief, American Community Survey Office

Patrick J. Cantwell  
Chief, Decennial Statistical Studies Division

Karen Battle  
Chief, Population Division

Prepared by: Gregory J. Mills  
Survey Analytics and Measures Branch  
American Community Survey Office

Sarah K. Heimel  
American Community Survey Experiments Branch  
Decennial Statistical Studies Division

Angela B. Buchanan  
Ethnicity and Ancestry Branch  
Population Division

Subject: Response Consistency between the Detailed Race and Hispanic  
Origin Questions from the 2016 ACS Content Test and the  
Ancestry Question

Attached is the American Community Survey (ACS) Research and Evaluation report, "Response Consistency between the Detailed Race and Hispanic Origin Questions from the 2016 ACS

Content Test and the Ancestry Question.” As currently constructed, each individual residing in every sampled housing unit and group quarter in the ACS is asked three separate questions regarding Hispanic origin, race, and ancestry. Questions arise about the extent to which data collected in the ancestry item are distinct from data collected in the race and Hispanic origin items, particularly in light of planned updates that will ask detailed information for all race categories. Using 2016 ACS Content Test data, this research first measured the degree of similarity between individuals’ race/Hispanic origin and ancestry responses. Second, using 2016 ACS production data with 2016 ACS Content Test data, this research compared group estimates derived from race and Hispanic origin responses with estimates derived from ancestry responses. Overall, this research found that when asked in conjunction with the expanded race and Hispanic origin survey items, similar data are collected in the ancestry question as in the Hispanic origin and race questions; in many cases it collects information that was already collected from the race and Hispanic origin survey items. In addition, the revised race and Hispanic origin items, when considered jointly, tended to collect more detailed and more comprehensive information than the ancestry item overall. When deciding about the future of the ancestry survey item on the ACS, decision makers must take into account several competing considerations. This research is an important step in making sure the ACS survey items on race and ethnic origins continue to be well-researched and driven by overarching concerns for data quality and respondent burden.

If you have any questions about this report, please contact Gregory Mills at 301-763-6172, Sarah Heimel at 202-384-8548, or Angela Buchanan at 301-763-6040.

Attachment

cc:

Rodney Terry	CBSM
Roberto Ramirez	POP
Merarys Rios	POP
Hyon Shin	POP
Nicholas Jones	POP
Jason Lee	SEHSD

September 12, 2019

# **Response Consistency between the Detailed Race and Hispanic Origin Questions from the 2016 ACS Content Test and the Ancestry Question**

FINAL REPORT



**Gregory J. Mills**  
American Community Survey Office

**Sarah K. Heibel**  
Decennial Statistical Studies Division

**Angela B. Buchanan**  
Population Division

---

Intentionally Blank

## TABLE OF CONTENTS

<b>EXECUTIVE SUMMARY</b> .....	<b>v</b>
<b>1. Introduction</b> .....	<b>1</b>
<b>2. Background and Literature Review</b> .....	<b>4</b>
2.1 The ACS and the 2016 ACS Content Test .....	4
2.2 Race and Hispanic Origin Questions .....	4
2.3 Ancestry Question.....	9
2.3.1 Background on the Ancestry Question.....	9
2.3.2 Ancestry Data Products .....	10
2.4 Past Research on the Interaction of Race/Hispanic Origin and Ancestry.....	11
<b>3. Research Questions and Methodology</b> .....	<b>14</b>
3.1 Research Questions .....	14
3.2 Data and Methods .....	15
3.2.1 Data .....	15
3.2.2 Coding and Detailed Group Construction .....	16
3.2.3 Research Question Metrics .....	18
3.2.4 Standard Error Calculation and Statistical Testing.....	23
<b>4. Assumptions and Limitations</b> .....	<b>24</b>
<b>5. Results</b> .....	<b>25</b>
5.1 Research Question 1 .....	25
5.2 Research Question 2 .....	29
5.3 Research Question 3 .....	31
5.4 Research Question 4 .....	35
5.5 Research Question 5 .....	40
<b>6. Conclusions</b> .....	<b>45</b>
<b>7. References</b> .....	<b>48</b>
<b>Appendix A. Mail Materials from the Test Treatment of the 2016 ACS Content Test</b> .....	<b>51</b>
<b>Appendix B. Internet Materials from the 2016 ACS Content Test</b> .....	<b>52</b>
<b>Appendix C. CATI and CAPI Questionnaire Materials</b> .....	<b>57</b>
<b>Appendix D. Table B04006: People Reporting Ancestry</b> .....	<b>61</b>
<b>Appendix E. Expanded Version of Table 11</b> .....	<b>65</b>
<b>Appendix F. Analysis Tables using the 2016 ACS Content Test, Test Treatment</b> .....	<b>69</b>
Research Question 1 .....	69
Research Question 2 .....	70

Research Question 3.....	72
Research Question 4.....	74
Research Question 5.....	77
<b>Appendix G. Limitations in Identifying Matches Between Race and Ancestry.....</b>	<b>85</b>

### List of Tables

Table 1. OMB Categories and Definitions for Data on Race and Ethnicity.....	5
Table 2. Percentage of People with Consistent Ancestry Responses Compared to Race or Hispanic Origin in 2016 ACS Content Test .....	13
Table 3: Presence and Validity of Race/Hispanic Origin and Ancestry Responses.....	27
Table 4: Distribution of Hispanic Origin By Validity of Race/Ancestry Data.....	28
Table 5: Distribution of Race By Validity of Race/Ancestry Data .....	29
Table 6. Match Rates for Persons in the Control Treatment of the 2016 ACS Content Test .....	30
Table 7. Match Rates for all Reported Ancestries .....	31
Table 8. Match Rates, Difference in Major Category, and Specificity of Information Provided by Ancestry Response Compared With Race/Hispanic Origin Response, by Race or Hispanic Origin of the Ancestry Response.....	33
Table 9. Location of Major Race and Hispanic Origin Group Responses Across Survey Items ....	36
Table 10. Location of Detailed Responses Across Survey Items, by Major Category and Detailed Group.....	38
Table 11. Comparing Estimates Derived from Race/Hispanic Origin and Ancestry, by Detailed Group, Showing Only Groups that were Significantly Different .....	41
Table 12. Comparing Estimates Derived from Race/Hispanic Origin and Ancestry, by Major Category or Detailed Group .....	43
Table 13. Comparing Estimates Derived from Race/Hispanic Origin and Ancestry, by World Region.....	45

### List of Figures

Figure 1. Planned Hispanic Origin and Race Questions for the 2020 ACS.....	3
Figure 2. Hispanic Origin and Race Questions from the 2016 ACS Production .....	6
Figure 3. Race and Ethnicity Questions from the Control Treatment of the 2016 ACS Content Test.....	9
Figure 4. Ancestry Question on the ACS.....	10

## List of Appendix Tables

Table A-1. CATI and CAPI Versions of the Race and Ethnicity Questions .....	57
Table A-2. Reproduction of AFF Table B04006, People Reporting Ancestry .....	61
Table A-3. Comparing Group Estimates Derived from Race/Hispanic Origin and Ancestry .....	65
Table A-4: Presence and Validity of Race/Hispanic Origin and Ancestry Responses (Test Treatment) .....	70
Table A-5: Match Rates For All Persons in the Test Treatment of the 2016 ACS Content Test ...	71
Table A-6. Match Rates for all Reported Ancestries (Test Treatment) .....	71
Table A-7. Match Rates, Difference in Major Category, and Specificity of Information Provided by Ancestry Response Compared With Race/Hispanic Origin Response, by Race or Hispanic Origin of the Ancestry Response (Test Treatment) .....	72
Table A-8. Location of Detailed Responses Across Survey Items, by Major Category .....	74
Table A-9. Location of Detailed Responses Across Survey Items, by Major Category and Detailed Group .....	75
Table A-10. Comparing Estimates Derived from Race/Hispanic Origin and Ancestry, by Detailed Group .....	78
Table A-11. Comparing Estimates Derived from Race/Hispanic Origin and Ancestry, by Major Category and Detailed Group .....	82
Table A-12. Comparing Group Estimates Derived from Race/Hispanic Origin and Ancestry, by World Region .....	84
Table A-13. Matching Results for Select Multiracial Responses .....	86
Table A-14. Matching Results for Select AIAN Responses .....	86
Table A-15. Matching Results for Select Hispanic Responses .....	87
Table A-16. Matching Results for Select White Responses .....	88
Table A-17. Matching Results for Select Black Responses .....	90

## List of Appendix Figures

Figure A-1. Mail Version of the Race and Ethnicity Question from the Test Treatment of the 2016 ACS Content Test .....	51
Figure A-2. Internet Version of the Control Treatment Hispanic Origin Question .....	52
Figure A-3. Internet Version of the Control Treatment Race Question .....	52
Figure A-4. Image of the Internet Version of the Test Treatment Combined Race and Hispanic Origin Question .....	53

Figure A-5. White Detailed Question for the Test Treatment Race and Hispanic Origin Question .....	54
Figure A-6. Hispanic Detailed Question for the Test Treatment Race and Hispanic Origin Question .....	54
Figure A-7. Black Detailed Question for the Test Treatment Race and Hispanic Origin Question .....	55
Figure A-8. Asian Detailed Question for the Test Treatment Race and Hispanic Origin Question .....	55
Figure A-9. AIAN Detailed Question for the Test Treatment Race and Hispanic Origin Question .....	55
Figure A-10. MENA Detailed Question for the Test Treatment Race and Hispanic Origin Question .....	56
Figure A-11. NHPI Detailed Question for the Test Treatment Race and Hispanic Origin Question .....	56
Figure A-12. SOR Detailed Question for the Test Treatment Race and Hispanic Origin Question .....	56



## EXECUTIVE SUMMARY

The U.S. Census Bureau's mission is to provide quality data about the nation's people and economy. Part of that mission is to ask only what is necessary and to avoid redundancy in survey items. Currently, each individual surveyed in the American Community Survey (ACS) is asked three separate questions to capture Hispanic origin, race, and ancestry. The major race categories as defined by the Office of Management and Budget's *1997 Revisions to the Standards for the Classification of Federal Data* are White, Black or African American, Asian, American Indian or Alaska Native, and Native Hawaiian or Other Pacific Islander (U.S. Office of Management and Budget, 1997). The Census Bureau also uses the category Some Other Race. The major ethnicity categories are Hispanic or Latino and Not Hispanic or Latino.

Starting in 2020, the race question will ask respondents to provide detailed information about any of the major race categories they report. For instance, a respondent who reports their race as White will be asked their specific White origin. Examples will be provided, in this case: German, Irish, English, Italian, Lebanese and Egyptian. Previously, detailed information was not elicited from the White or Black categories; rather, it was collected from the ancestry item. With this revision to the race item, questions arise about the extent to which data collected in the ancestry item are distinct from data collected in the race and Hispanic origin items and whether retaining the ancestry item on the ACS causes unnecessary burden on respondents.

The 2016 ACS Content Test provided a unique opportunity to compare detailed reporting on the race, Hispanic origin, and ancestry items because detailed information was collected from the Hispanic origin question and all five major race categories.<sup>1</sup> This report compares responses to the revised race/Hispanic origin questions with information collected from ancestry in both the 2016 ACS Content Test and in regular ACS production.<sup>2</sup> There were some limitations to this research, most notably that the 2016 ACS Content Test was not conducted in Alaska and Hawaii, that it was only collected in English and Spanish, and that the processing procedures allowed for more race and Hispanic groups to be collected than for ancestry groups.

Our analyses revealed several important findings. First, we measured the degree of consistency between individuals' race/Hispanic origin and ancestry responses in the 2016 ACS Content Test. In this analysis, we found that:

---

<sup>1</sup> The 2016 ACS Content Test was split into test and control treatments. In the analysis provided in the main paper, we used the data from the control treatment. The control treatment contained a race question that collected detailed groups in a format similar to what is being planned for 2020 Census data collection.

<sup>2</sup> Throughout this report, the terminology "race/Hispanic origin" refers to a combined response from both the race and Hispanic origin questions. For example, a respondent was considered Brazilian for race/Hispanic origin if they indicated they were Brazilian on the Hispanic origin question, on the race question or on both questions.

- 84.7 percent of individuals provided a valid response to both the race/Hispanic origin and ancestry items.<sup>3</sup>
- Among those that reported a valid response to both questions, 72.5 percent had an ancestry or ancestries that matched exactly to their reported race/Hispanic origin information, another 7.1 percent half matched (one of their reported ancestries matched and the other did not), and 20.4 percent provided an ancestry or ancestries that did not match at all.

We then turned to looking at individual ancestry responses, as some people reported more than one ancestry, and could have matched on one but not the other. The goal of this analysis was to see how non-matching ancestry responses were different from a respondent's reported race or Hispanic origin.

- A quarter of all ancestries reported did not exactly match what was reported in the race/Hispanic origin items (24.9 percent).
- While these ancestries did not exactly match what was reported in race/Hispanic origin, the majority were in the same major category as reported by the respondent (19.4 percent of all ancestries). An example of this is a respondent who selected the White box in the race item, did not report Irish in race/Hispanic origin, and reported "Irish" in the ancestry item. Thus, the ancestry response of Irish is not an exact match, but is in the same major category as the race/Hispanic origin response.
- In comparison, a smaller proportion of the non-matching ancestries were not in the same major race/Hispanic origin category (2.5 percent of the ancestries). For example, a respondent may have selected only the American Indian or Alaska Native box in race, and reported "Irish" in ancestry.

We then analyzed the non-matching ancestry responses that were in the same major category as a race/Hispanic origin report (the 19.4 percent from above). The goal of this analysis was to learn whether people tended to provide a more specific response in race/Hispanic origin or ancestry.<sup>4</sup> In this analysis, we found that:

- The majority of the ancestry responses that did not match exactly but were in the same major category were *more specific* than the response provided in the race/Hispanic origin items (12.0 percent of all ancestries). The situation described above, of "Irish" being provided in ancestry while only White was reported in race, is an example of an ancestry response that was more specific than the race/Hispanic response.
- 6.5 percent of all ancestry responses did not match exactly to a race/Hispanic origin response, but were in the same major category and provided the *same level of specificity* as the race/Hispanic origin response.

---

<sup>3</sup> Responses such as "Adopted," "Don't know," or a name were not considered valid for this analysis.

<sup>4</sup> We compared the level of specificity in ancestry responses only to race/Hispanic origin responses that were in the same major category due to the difficulty in comparing levels of specificity across major categories.

- 0.9 percent of all ancestry responses did not match exactly to a race/Hispanic origin response, were in the same major category, and were *less specific* than the race/Hispanic origin response.

Next, we examined the locational variation in the reporting of groups, noting how often a particular group was reported in both items, or in only the race/Hispanic origin or ancestry items. We analyzed major race/Hispanic origin categories, and the six most common detailed groups within each major category with large enough sample sizes to be compared.<sup>5</sup>

- 11 of the 25 detailed groups examined were reported more often from race/Hispanic origin than from ancestry.
- 14 groups were reported as often from race/Hispanic origin as they were from ancestry.
- No groups had higher reporting from ancestry.

Finally, we analyzed 104 groups with sufficient sample size to examine how production estimates would differ if derived from the race/Hispanic origin data from the 2016 ACS Content Test. We also examined differences in detailed reporting by race and Hispanic origin category.

- A majority of groups, 88 of the 104, were not statistically different.
- Eight groups had higher estimates from race/Hispanic origin data (English, Irish, German, Scottish, Dutch, Asian Indian, Puerto Rican, and Unclassified).
- Eight had higher estimates from ancestry data (French Canadian, British, Haitian, Pennsylvania German, Czechoslovakian, Ethiopian, Kenyan, and Not Reported).
- More people provided a detailed response to race/Hispanic origin in the 2016 ACS Content Test than did in response to ancestry in ACS production (74.1 vs. 70.4 percent).
- Detailed reporting on the race/Hispanic origin question from the 2016 ACS Content Test was higher for White (44.7 vs. 42.1 percent), Asian (6.7 vs. 5.4 percent), and Hispanic (16.0 vs. 14.3 percent) respondents overall. Differences in detailed reporting among Black and Native Hawaiian or Pacific Islander were not statistically different.

When deciding about the future of the ancestry survey item on the ACS, decision makers must take into account several considerations, including the burden that the ancestry question places on the public, the utility and necessity of the data collected, as well as the views of affected groups and community stakeholders. While this research cannot speak to all these issues, it indicates that there is redundancy in the ancestry data when compared with data from the revised race and Hispanic origin questions. These results should be considered as the ACS program strives to minimize respondent burden while maintaining high data quality standards.

---

<sup>5</sup> Detailed American Indian and Alaska Native groups that were reported in response to the ancestry question were not retained during the coding operation. For this reason, those groups were not included in this analysis.



## 1. INTRODUCTION

The U.S. Census Bureau conducts the American Community Survey (ACS) in order to provide communities around the nation with social, economic, housing, and demographic information about their residents on an annual basis. The content of the survey is also designed to meet the needs of federal government agencies. Every question on the ACS has a required purpose, with periodic reviews of the survey content. The ACS asks questions pertaining to the household and housing unit, as well as each person within the housing unit. Among the questions asked of each person, there are separate questions regarding: 1) Hispanic origin, 2) race, and 3) ancestry or ethnic origin.

The ACS questions on Hispanic origin and race are the same as those asked on the decennial census. The major race and ethnicity categories are based on the U.S. Office of Management and Budget's (OMB) *1997 Revisions to the Standards for the Classification of Federal Data*. The Hispanic origin and race questions are asked in order to create statistics about race and to gather detailed information within certain categories—specifically Hispanic or Latino, White, Black or African American, Asian, Native Hawaiian and Other Pacific Islander (NHPI), American Indian and Alaska Native (AIAN), and Some Other Race (SOR).

The question on ancestry was included on the decennial census long form from 1980 through 2000 and, along with the other long-form items, is now only included on the ACS. The ancestry question has been asked on the ACS since its first year of data collection.<sup>6</sup> The question asks about one's "ancestry or ethnic origin." Although ancestry may be intended to measure a different concept from race, hence differences in question wording, instructions, and classification of groups, some of the examples provided with the ancestry question on the ACS also appear as checkboxes in the Hispanic origin or race questions (such as Mexican and Korean). Currently, however, data for detailed Hispanic, Asian, NHPI, and AIAN groups are not tabulated from ancestry. Only the detailed White, Black, and SOR groups (such as Lebanese, Haitian, and Brazilian, respectively) are in the ancestry data products.

The Census Bureau is planning to make the following changes to the way Hispanic origin and race are asked in the 2020 ACS, which will also be reflected in the 2020 Census questionnaire:

- Examples of detailed origins will be revised for the Hispanic origin question.
- The instruction to the race question will include "**AND** print origins."
- Detailed origin(s) will be collected for the White and Black categories using write-in lines.
- Examples of detailed origins will be added for the White, Black, and AIAN categories.

---

<sup>6</sup> To see an archive of past ACS questionnaires, visit <https://www.census.gov/programs-surveys/acs/methodology/questionnaire-archive.html>.

- Removal of the term “Guamanian” from the Chamorro checkbox and other revisions to the detailed origins and examples in the Asian and Other Pacific Islander categories.

Thus, starting in 2020 the ACS race question will be different in that it will collect detailed groups that are considered White or Black, such as Italian, Lebanese, African American, or Haitian, along with detailed information from all other groups. The paper questionnaire version of the new question can be seen in Figure 1 on the next page. The added examples to the side of “White” and “Black or African American” checkboxes will aid in the collection of more detailed data, as past Census Bureau research has found these to be helpful to respondents (Compton et al., 2012; Mathews et al., 2017).

Because of the wide range of detailed groups that will be collected and reported from the newly revised race question, the question on ancestry may be increasingly redundant, and therefore, burdensome. This concern has been expressed previously by those familiar with the ancestry question, and was also suggested by results from the 2016 ACS Content Test, which found a high degree of matching among race and ancestry responses (Harth et al., 2017).

The research in this paper was proposed in order to address this issue of redundancy. It will examine how people respond to ancestry after being asked the revised race and Hispanic origin questions, and how the data for the groups currently published in ancestry data products would differ if derived from the revised race and Hispanic origin questions. The results will inform decisions about whether the ancestry question should be retained or removed in the future.

Most of the data used in this report are from the 2016 ACS Content Test. The 2016 ACS Content Test questions on Hispanic origin and race asked for detailed origins from all major Hispanic origin and race categories on both the control and test treatments (thus both versions differed from the questions used in production ACS). The control treatment asked Hispanic origin and race in separate questions, with examples given for every category and space available for reporting detailed origins for every category including White and Black. This question is similar to the race question planned for the 2020 Census and the 2020 ACS, with differences only in instruction wording and the number of examples used. The test treatment asked Hispanic origin and race in one combined “race or ethnicity” question, included a separate category for Middle Eastern or North African, and requested detailed groups within each major category. Since the layout of the control version is most similar to the layout of the Hispanic origin and race questions being proposed for 2020 ACS, this report will mainly analyze and discuss data from the control treatment. Finally, the 2016 ACS production data on ancestry are also used in this analysis to compare relative sizes of detailed groups when estimates are created using the revised race and Hispanic origin questions.

**Figure 1. Planned Hispanic Origin and Race Questions for the 2020 ACS**

**5 Is Person 1 of Hispanic, Latino, or Spanish origin?**

No, not of Hispanic, Latino, or Spanish origin

Yes, Mexican, Mexican Am., Chicano

Yes, Puerto Rican

Yes, Cuban

Yes, another Hispanic, Latino, or Spanish origin – *Print, for example, Salvadoran, Dominican, Colombian, Guatemalan, Spaniard, Ecuadorian, etc.* ↴

**6 What is Person 1's race?**  
*Mark (X) one or more boxes AND print origins.*

White – *Print, for example, German, Irish, English, Italian, Lebanese, Egyptian, etc.* ↴

Black or African Am. – *Print, for example, African American, Jamaican, Haitian, Nigerian, Ethiopian, Somali, etc.* ↴

American Indian or Alaska Native – *Print name of enrolled or principal tribe(s), for example, Navajo Nation, Blackfeet Tribe, Mayan, Aztec, Native Village of Barrow Inupiat Traditional Government, Nome Eskimo Community, etc.* ↴

<input type="checkbox"/> Chinese	<input type="checkbox"/> Vietnamese	<input type="checkbox"/> Native Hawaiian
<input type="checkbox"/> Filipino	<input type="checkbox"/> Korean	<input type="checkbox"/> Samoan
<input type="checkbox"/> Asian Indian	<input type="checkbox"/> Japanese	<input type="checkbox"/> Chamorro
<input type="checkbox"/> Other Asian – <i>Print, for example, Pakistani, Cambodian, Hmong, etc.</i> ↴	<input type="checkbox"/> Other Pacific Islander – <i>Print, for example, Tongan, Fijian, Marshallese, etc.</i> ↴	

Some other race – *Print race or origin.* ↴

Source: Questions Planned for the 2020 American Community Survey (2018).

## 2. BACKGROUND AND LITERATURE REVIEW

### 2.1 The ACS and the 2016 ACS Content Test

The ACS is a nationwide survey that collects information on demographic, social, economic, and housing characteristics about the nation's population. The survey utilizes a multimode methodology to collect data from both housing units and group quarters (GQs), and is carried out in all 50 states, plus the District of Columbia and Puerto Rico.<sup>7</sup> The ACS has an annual sample size of about 3.5 million addresses, with survey information collected nearly every day of the year. Data are pooled across a calendar year to produce estimates for that year. As a result, ACS estimates reflect data that have been collected over a period of time rather than for a single point in time as in the decennial census, which is conducted every 10 years and provides population counts as of April 1. ACS 1-year estimates are data that have been collected over a 12-month period and are available for geographic areas with at least 65,000 people.

The 2016 ACS Content Test assessed changes to ten ACS topics to determine whether proposed changes to question wording, response categories, and definitions of underlying constructs improved the quality of the data collected. Race and Hispanic origin was one of the topics tested in the 2016 ACS Content Test. The test was a split-panel experiment with one-half of addresses assigned to a control treatment and the other half assigned to a test treatment. The same data collection modes used in the ACS production were used in the 2016 ACS Content Test; an internet questionnaire, paper questionnaire, computer-assisted telephone interviews (CATI), and computer-assisted personal interviews (CAPI).

### 2.2 Race and Hispanic Origin Questions

Since the first census in 1790, the Census Bureau has collected information on race. A question on Hispanic origin was added in 1970. The census questionnaire has changed over time, reflecting changes in society. Today, as mentioned previously, the Census Bureau collects race and ethnicity data following OMB standards. The standards state that the two-question format should be used whenever feasible, and that ethnicity should be collected before race. The standards also define five broad categories for data on race and two broad categories for data on ethnicity as shown in Table 1 (U.S. Office of Management and Budget, 1997).<sup>8</sup> The Census Bureau also includes a category called Some Other Race for those who do not identify with any of the categories in the standard.

---

<sup>7</sup> The Census Bureau classifies all people not living in housing units as living in group quarters. There are two types of group quarters: institutional group quarters (for example, correctional facilities for adults, nursing homes, and hospice facilities) and noninstitutional group quarters (for example, college/university student housing, military quarters, and group homes).

<sup>8</sup> Since OMB instructs that the two minimum categories on ethnicity should be "Hispanic or Latino" and "Not Hispanic or Latino," this report uses the terms "ethnicity" and "Hispanic origin" interchangeably.



**Table 1. OMB Categories and Definitions for Data on Race and Ethnicity**

Race and Ethnicity Categories	Definitions
<b>Race</b>	
<b>American Indian or Alaska Native</b>	A person having origins in any of the original peoples of North and South America (including Central America) and who maintains tribal affiliation or community attachment.
<b>Asian</b>	A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam.
<b>Black or African American</b>	A person having origins in any of the Black racial groups of Africa.
<b>Native Hawaiian or Other Pacific Islander</b>	A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands.
<b>White</b>	A person having origins in any of the original peoples of Europe, the Middle East, or North Africa.
<b>Ethnicity</b>	
<b>Hispanic or Latino</b>	A person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin, regardless of race. The term, "Spanish origin," can be used in addition to "Hispanic or Latino."
<b>Not Hispanic or Latino</b>	(No definition provided)

Source: U.S. Office of Management and Budget; Federal Register Vol. 62, No. 210, 1997.

The OMB standards further encouraged the collection of greater detail from race and ethnicity, as long as the additional detailed categories could be aggregated into the minimum categories outlined above. When tabulating the data, OMB recommended that federal agencies should use procedures that “result in the production of as much detailed information on race and ethnicity as possible,” as long as data quality or confidentiality standards are not compromised.

The Hispanic origin and race questions from the 2016 ACS production paper questionnaire are shown in Figure 2. These questions are analogous to those used in the 2010 Census.<sup>9,10</sup>

<sup>9</sup> Interviewer-assisted versions of the ACS questionnaires had the same content as shown in the paper versions, but with somewhat simplified pathing appropriate for the mode. For example, in the Hispanic question the respondent answers “yes” or “no” to the question, and is only read the detailed options if they respond “yes.” Similarly, in the race question the respondent is only read the major race groups (e.g. “Asian”), which is followed by the corresponding detailed groups (e.g. “Asian Indian,” “Japanese,” etc.) if the major race group is indicated.

<sup>10</sup> The only difference between the 2010 Census and 2016 ACS questions is that the term “Negro” was removed from the Black or African American category in the ACS after 2010. However, people who wrote-in “Negro” were still tabulated in that category.

**Figure 2. Hispanic Origin and Race Questions from the 2016 ACS Production**

→ **NOTE:** Please answer BOTH Question 5 about Hispanic origin and Question 6 about race. For this survey, Hispanic origins are not races.

**5** Is Person 1 of Hispanic, Latino, or Spanish origin?

- No, not of Hispanic, Latino, or Spanish origin
- Yes, Mexican, Mexican Am., Chicano
- Yes, Puerto Rican
- Yes, Cuban
- Yes, another Hispanic, Latino, or Spanish origin – *Print origin, for example, Argentinian, Colombian, Dominican, Nicaraguan, Salvadoran, Spaniard, and so on.* ↴

**6** What is Person 1's race? Mark (X) one or more boxes.

- White
- Black or African Am.
- American Indian or Alaska Native – *Print name of enrolled or principal tribe.* ↴

<input type="checkbox"/> Asian Indian	<input type="checkbox"/> Japanese	<input type="checkbox"/> Native Hawaiian
<input type="checkbox"/> Chinese	<input type="checkbox"/> Korean	<input type="checkbox"/> Guamanian or Chamorro
<input type="checkbox"/> Filipino	<input type="checkbox"/> Vietnamese	<input type="checkbox"/> Samoan
<input type="checkbox"/> Other Asian – <i>Print race, for example, Hmong, Laotian, Thai, Pakistani, Cambodian, and so on.</i> ↴	<input type="checkbox"/> Other Pacific Islander – <i>Print race, for example, Fijian, Tongan, and so on.</i> ↴	

Some other race – *Print race.* ↴

Source: U.S. Census Bureau, 2016 ACS Production Paper Questionnaire

Research since 2010 has shown that a growing number of people find the current race and ethnicity categories confusing (Compton et al. 2012). One factor contributing to the confusion is that there are no examples listed with the White, Black, or American Indian or Alaska Native categories. Neither is there a place for people with specific White or Black origins, such as Lebanese or Haitian, to report their specific origins.

Over the past decade, the Census Bureau has conducted extensive research to address quality concerns with race and Hispanic origin reporting and issues raised by data users and community organizations. During the 2010 Census, as part of the 2010 Alternative Questionnaire Experiment (AQE), researchers conducted focus groups and cognitive interviews on variations of the Hispanic origin and race questions. Additional research was conducted to test prospective question designs for the content of the 2020 Census, particularly with the new emphasis on using web-based designs for data collection (Childs et al., 2010; Compton et al., 2012; Fernandez et al., 2009). This work culminated in the 2015 National Content Test (NCT) and later the 2016 ACS Content Test, which tested several versions of the race and ethnicity questions, using various question formats, wording, examples, and methods for detailed origin collection within all categories (Matthews et al., 2017).

After consultation with OMB on the results of the testing and the needs of agencies across the federal government, the Census Bureau moved forward with the two-question format for the collection of race and ethnicity for the 2020 Census and the 2020 ACS.<sup>11</sup> The Census Bureau also revised the wording, examples, and write-in spaces associated with the questions. The revisions serve to increase understanding of each racial category and improve the data collected on detailed origins, while continuing to follow the current federal guidelines for the collection of race and ethnicity data set by OMB in 1997.

For the Hispanic origin question, the examples were changed. Guatemalan and Ecuadorian were added to the Hispanic origin examples, while Argentinean and Nicaraguan were removed (see Figure 1 for the Hispanic origin and race questions starting in 2020, and see Figure 2 for the Hispanic origin and race questions in 2016 ACS). The Hispanic origin examples were also reordered.

For the race question, there was an instruction wording change as well as changes to examples for every category except SOR. The instruction “Mark (X) one or more boxes” was expanded to “Mark [X] one or more boxes **AND** print origins.” Examples and write-in areas were added to both the White category and the Black or African American category. The examples for White are German, Irish, English, Italian, Lebanese, and Egyptian. The examples for Black are African American, Jamaican, Haitian, Nigerian, Ethiopian, and Somali. For the American Indian or Alaska Native category, the instruction was changed from “Print name of enrolled or principal tribe” to “print name of enrolled or principal tribe(s), for example, Navajo Nation, Blackfeet Tribe, Mayan, Aztec, Native Village of Barrow Inupiat Traditional Government, Nome Eskimo Community, etc.” The set of checkboxes that covered the Asian and NHPI groups did not change in content except that “Guamanian or Chamorro” was changed to “Guamanian.” The groups were reordered, however. The examples for “Other Asian” were reduced; Laotian and Thai were deleted. For Other Pacific Islander, Marshallese was added. The SOR category continues to have no examples, but the instructions in that category were changed from “Print race” to “Print race or origin.”

The most noteworthy of these revisions, in terms of changes to the type of data that will be collected and produced from 2020 Census and ACS 2020, is the addition of write-in areas to the White and Black categories, which makes it possible for respondents to report White and Black detailed origins. The control treatment of the 2016 ACS Content Test was very similar to the revised race and Hispanic origin questions in that it had write-in areas for White and Black. Thus,

---

<sup>11</sup> On January 26, 2018, the Census Bureau issued a memorandum to document the 2020 Census Program decision on race and ethnicity questions for the 2018 End-to-End Census Test and the proposed design for the 2020 Census. The Census Bureau needed to make a decision on the design of the race and ethnicity questions by December 31, 2017 in order to prepare 2020 Census systems, and deliver the final 2020 Census question wording to Congress by March 31, 2018. In accordance with the current 1997 OMB standards for race and ethnicity, the 2020 Census will use two separate questions for collecting data on race and ethnicity. Source: [https://www.census.gov/programs-surveys/decennial-census/2020-census/planning-management/memo-series/2020-memo-2018\\_02.html](https://www.census.gov/programs-surveys/decennial-census/2020-census/planning-management/memo-series/2020-memo-2018_02.html).

data from the control treatment can be used to compare race and Hispanic origin with ancestry in order to understand how respondents might answer both questions together and how the data from the revised question may compare to current ancestry data. All analysis using the control treatment in this report was also done using the test treatment; those results are shown in Appendix F.

Figure 3 shows the control version of the Hispanic origin and race questions from the 2016 ACS Content Test, as they appeared on the paper questionnaire mailed to sampled households.<sup>12</sup> The internet instrument for the control treatment contained identical checkbox categories and write-in areas in both the race and Hispanic origin questions as was found in the paper questionnaire. However, the race question on the internet mode of the test treatment offered six checkboxes to elicit specific groups within each major race and Hispanic origin category, which the control treatment did not offer.<sup>13</sup> Interviewer-assisted versions of the ACS questionnaires had the same general content as shown in the paper versions, but were modified to be appropriate for the mode. For instance, the paper questionnaires for both the control and test treatments listed only three examples for each major OMB category due to space constraints, while the CATI and CAPI modes provided six examples.<sup>14</sup>

---

<sup>12</sup> Mail materials from the test treatment of the 2016 ACS Content Test were distinct from the control treatment, and are shown in Appendix A.

<sup>13</sup> Internet questionnaires for both the test and control treatments are shown in Appendix B.

<sup>14</sup> Interviewer-assisted versions of test and control treatment questionnaires are shown in Appendix C.

**Figure 3. Race and Ethnicity Questions from the Control Treatment of the 2016 ACS Content Test**

→ **NOTE: Please answer BOTH Question 5 about Hispanic origin and Question 6 about race. For this survey, Hispanic origins are not races.**

**5 Is Person 1 of Hispanic, Latino, or Spanish origin?** Mark all boxes that apply **AND** print origins. Note, you may report more than one group.

No, not of Hispanic, Latino, or Spanish origin

Yes, Mexican, Mexican Am., Chicano

Yes, Puerto Rican

Yes, Cuban

Yes, another Hispanic, Latino, or Spanish origin – Print details, for example, Salvadoran, Dominican, Colombian, etc. ↴

**6 What is Person 1's race?** Mark all boxes that apply **AND** print origins in the spaces below. Note, you may report more than one group.

**White** – Print details, for example, German, Lebanese, Egyptian, etc. ↴

**Black or African Am.** – Print details, for example, African American, Jamaican, Nigerian, etc. ↴

**American Indian or Alaska Native** – Print name of enrolled or principal tribe(s), for example, Navajo Nation, Mayan, Native Village of Barrow Inupiat Traditional Government, etc. ↴

Chinese       Vietnamese       Native Hawaiian

Filipino       Korean       Samoan

Asian Indian       Japanese       Chamorro

Other Asian – Print details, for example, Pakistani, Cambodian, Hmong, etc. ↴       Other Pacific Islander – Print details, for example, Tongan, Fijian, Marshallese, etc. ↴

**Some other race** – Print race(s) or origin(s). ↴

Source: U.S. Census Bureau, 2016 ACS Content Test Paper Questionnaire

## 2.3 Ancestry Question

In this section, we present information about the current ancestry question, its history on the decennial census long-form questionnaire and the ACS, published data products that include ancestry data, and past research on the interaction between ancestry and race and ethnicity.

### 2.3.1 Background on the Ancestry Question

The Census Bureau asks about a person's ancestry or ethnic origin to create statistics about ancestry groups in America. Ancestry data allow researchers and policy makers to measure the characteristics of ethnic groups and to tailor services to accommodate cultural differences. These data are used in planning and evaluating government programs and policies, and to facilitate the fair and equitable access to programs across ethnic groups. They also facilitate the enforcement of laws, regulations, and policies against ethnic discrimination.

Ancestry is a broad concept that may be interpreted in a variety of ways by different individuals. People may view it as where they or their parents are from, where their ancestors originated, or as how they see themselves ethnically (Farley, 1991). The Census Bureau defines ancestry as a person's ethnic origin, heritage, descent, or "roots," which allows for any of these interpretations of the concept (Brittingham and de la Cruz, 2004). In large part, reported ancestries tend to be

geographic or ethnic terms, although many responses include racial, religious, or unclassifiable terms (Hobbs and Brittingham, 2007).

The ancestry question was asked on the decennial census long form from 1980 to 2000, replacing a question on parents' place of birth, which had been asked on the census since 1870 (U.S. Census Bureau, 2002). Ancestry has been asked on the ACS since its full implementation in 2005. The ancestry question is shown in Figure 4 as it appears on the ACS paper questionnaire.

**Figure 4. Ancestry Question on the ACS**



**13 What is this person's ancestry or ethnic origin?**

(For example: Italian, Jamaican, African Am., Cambodian, Cape Verdean, Norwegian, Dominican, French Canadian, Haitian, Korean, Lebanese, Polish, Nigerian, Mexican, Taiwanese, Ukrainian, and so on.)

Source: U.S. Census Bureau, 2016 American Community Survey and 2016 ACS Content Test

### 2.3.2 Ancestry Data Products

The Census Bureau currently publishes several ACS data products relating to ancestry. Current data products provide ancestry group estimates for people reporting one ancestry alone, multiple ancestries, as well as total ancestry group estimates that are reported either alone or in combination. For example, the products show the number of people who reported German as their only ancestry, the number who reported German as one of multiple ancestries, and the overall number who reported German either alone or with another ancestry. The tables also show the number of people who only provided an uncodable ancestry response (such as “human”) and how many skip the question entirely. Additionally, there are tables that provide demographic, social, and economic data for selected ancestry groups.

The most recent public data products on ancestry can be found on the American FactFinder (AFF).<sup>15</sup> The tables that show ancestry on AFF are listed here, by table number and title:

- DP02 – Data Profile (shows only the larger ancestry groups)
- B04004 – People Reporting Single Ancestry
- B04005 – People Reporting Multiple Ancestry
- B04006 – People Reporting Ancestry (the total number of people who reported each ancestry)
- B04007 – Ancestry (i.e., single, multiple, or not reported)
- S0201 – Selected Population Profiles (demographic, social, and economic characteristics of ancestry groups)

---

<sup>15</sup> The American FactFinder can be accessed at <https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml>.

- Selected Population Tables (several hundred social and economic characteristic tables for ancestry groups)
- C04004- C04007 – Ancestry tables for a condensed number of ancestry groups

One of the main published data products derived from ACS ancestry data is the detailed Table B04006: *People Reporting Ancestry*. The 2016 ACS national estimates for this table are shown in Appendix D. Notably, although 106 ancestry groups and subgroups are shown in the table, over 122 million people are classified into the group denoted as “other groups.” This is because the table categories only include ancestries that correspond with White, Black, or Some Other Race identities. All other ethnic identities are only shown in the detailed race and Hispanic origin tables using data from the race and Hispanic origin questions. For the published AFF ancestry tables, responses to the ancestry question that fall into one of those groups (such as an Asian or Hispanic response) are aggregated into “Other groups.”

All ancestries, including those that are race or Hispanic groups, are included in the ACS Public Use Microdata Sample (PUMS) file. Ancestries in the PUMS data with estimates below a certain threshold are combined into larger groups for disclosure avoidance purposes. There have been two ancestry publications by the Census Bureau that used all reported ancestry responses: the report *Ancestry: 2000* and a subsequent tabulation of ancestries reported in Census 2000 by detailed code (Brittingham and de la Cruz, 2004; U.S. Census Bureau, 2007).

#### **2.4 Past Research on the Interaction of Race/Hispanic Origin and Ancestry**

In all ACS survey instruments, basic person-level demographics including race and Hispanic origin are collected at the start of the instrument, followed by all housing-level questions and a few additional person-level questions before the ancestry question is presented.<sup>16</sup> Prior research has documented that at least some respondents consider one or more of the race, Hispanic origin, and ancestry questions redundant (e.g., McKay and de la Puente, 1996; Schwede, Leslie, and Griffin, 2002).

The decennial census short form has never included a question about ancestry, although one was tested in the 2005 National Census Test. The 2005 National Census Test experimented with using abbreviated Hispanic origin and race questions followed immediately by an open-ended ancestry question. The question on ancestry was intended to elicit detailed information on all race and ethnic groups; no checkboxes were provided to capture detailed information in the Hispanic origin or race questions. Results showed less complete data on forms with this experimental three-question format for persons identifying as Asian, NHPI, and Hispanic (Alberti, 2006). As a result, the ancestry question was not pursued as an option to supplement simplified race and Hispanic origin questions as it did not yield data that are more complete.

---

<sup>16</sup> For the entire ACS questionnaire, visit <https://www.census.gov/programs-surveys/acs/about/forms-and-instructions/2019-form.html>.

The issue of redundancy between race/Hispanic origin and ancestry data was documented in an analysis of the 2016 ACS Content Test (Harth, et al., 2017).<sup>17</sup> Ancestry responses were compared to race and Hispanic origin responses, for those who had provided both an ancestry and a race or Hispanic origin, to assess how consistent the responses were. To be considered consistent, the one or two ancestry responses provided must have matched exactly with responses provided in the race and Hispanic origin questions. For example, if the two resulting codes from an individual's ancestry response identified them as German and Mexican, then a response to race and Hispanic origin that also identified them as German and Mexican was considered consistent (even if additional identities were reported in race and Hispanic origin, and regardless of the order of reporting). However, if the individual identified as German and Hispanic (with no more detail) in race or Hispanic origin, that response was not considered consistent with a response of German and Mexican in ancestry.

Table 2 shows that 72.5 percent of persons in the control treatment of the 2016 ACS Content Test provided responses in ancestry that were consistent with their reported race and Hispanic origin. Modal differences were noted however. Overall, interviewer-assisted modes (CATI and CAPI) had an 81.9 percent match rate, while self-response modes (internet and mail) had a 65.8 percent match rate. Within self-response modes, internet and mail had distinctly different rates; mail had the lowest match rate of all response modes, at 48.8 percent while 73.3 percent of persons on internet returns had ancestry responses that matched to their reported race and Hispanic origin responses.

---

<sup>17</sup> Throughout this report, we most often refer to the Hispanic origin and race items in a combined way as “race/Hispanic origin.” This is because in our analysis we combine responses to both questions in order to capture a more comprehensive self-identification. However, we also sometimes refer to these topics as “race and Hispanic origin” or “Hispanic origin and race” when referring specifically to the distinct questions or responses, as they are separate items on the questionnaire. All the above phrases are considered interchangeable for the purposes of this report.



**Table 2. Percentage of People with Consistent Ancestry Responses Compared to Race or Hispanic Origin in 2016 ACS Content Test**

Mode	Sample Size	Percent with Consistent Responses
All Modes	36,000	72.5 (0.5)
Self-Response	25,000	65.8 (0.5)
Mail	8,200	48.8 (1.1)
Internet	17,000	73.3 (0.5)
Interviewer Assisted	11,000	81.9 (0.9)
CATI	1,700	83.6 (1.5)
CAPI	9,300	81.7 (0.9)

Source: U.S. Census Bureau, 2016 American Community Survey Content Test, Control Treatment.

Note: Standard errors are shown in parentheses. DRB Approval Number CBDRB-FY19-RAGLIN-B0005.

While the majority of the responses to the ancestry item were consistent with race and Hispanic origin responses, responses might differ for several reasons. For one, some respondents might think since they are being asked another question about their ethnic background that they should provide a different answer than what they provided previously. They may provide responses with a higher or lower degree of specificity without necessarily opposing or contradicting their earlier response. For example, a respondent might report “Mexican” in the Hispanic origin item, while in the ancestry item report “Hispanic” ancestry. The reverse of this scenario might also occur, with the more specific “Mexican” response being reported in the ancestry item, with a more general “Hispanic” response in the Hispanic origin item. While these are valid and nonmatching responses, they are also not contradictory. Duncan and Trejo (2008) found evidence of both the above patterns, and found them to be correlated with specific human capital characteristics. Of course, some respondents also report disparate groups with little or no obvious commonality. Alba and Islam (2009) found that in the 1980, 1990, and 2000 censuses about three percent of U.S.-born Mexican Americans reported being “non-Hispanic” in the Hispanic origin item, but reported Mexican ancestry on the ancestry item. They argued that forces of assimilation might cause later generations of ethnic groups in particular to identify with an ethnic group in terms of ancestry rather than racial identity.

Past work also indicates that a high proportion of respondents leave the ancestry question blank compared to the race and Hispanic origin questions. Considering that ancestry data are not imputed for missing responses, as Hispanic origin and race data are, this could yield varying results for group estimates. For example, when the ancestry question was asked for the first time on the long-form of the 1980 Census, 26.5 million people identified as Black or African American on the race question, while only 21 million identified as such on the ancestry question (Lieberson and Waters, 1988). Hispanics have also been found to be more likely to identify themselves as such on the Hispanic origin question than the ancestry question (Lieberson and Waters, 1988; Duncan and Trejo, 2008). In the 2016 ACS Content Test control treatment, 13.2 percent of all respondents did

not provide a response to the ancestry question. A significantly higher percent (15.7 percent) did not provide an ancestry response in the test treatment. Thus, there is some indication that when probed for more details in the race and Hispanic origin questions, as was done in the test treatment, respondents view the ancestry question as being more redundant. However, ancestry has one of the highest rates of missing data of all ACS questions (Heimel, 2014).<sup>18</sup> In 2016 ACS production, shown in Table A-2, it was estimated that 51 million out of 323 million people had an unreported ancestry (about 15.7 percent of the population).

An additional indicator that suggests a general elevated level of confusion with the ACS ancestry question is the frequency with which internet respondents access the help screen on the internet questionnaire. Help screens are put in place throughout the internet questionnaire in order to provide respondents with guidance if they are having trouble understanding a particular question. If a certain question's help screen is accessed at a particularly high rate, it may indicate that it lacks clarity, either in question wording, response categories, instructions, or in other aspects of the survey design. A 2013 study found that the ancestry question elicited 13.1 percent of all help screen requests in the ACS internet instrument, which was the highest of the screens seen by all people (Horwitz, Tancreto, Zelenak and Davis, 2013). Although help screen rates of the race and Hispanic origin questions were not directly shown in that report, an additional internal analysis of paradata for the calendar year of 2016 revealed that the race question elicited 2.1 percent of help screen retrievals, and the Hispanic origin question elicited 0.9 percent, compared to 10.1 percent for ancestry. Taken together, help screen analyses suggest that some respondents have issues understanding or responding to ancestry question, which may also be an underlying cause of the high item missing data rates.

### **3. RESEARCH QUESTIONS AND METHODOLOGY**

In this section, we present the research questions and discuss the data and methods used for the analysis, including coding and metrics.

#### **3.1 Research Questions**

The research questions are listed below and clarified in subsequent sections.

1. What is the response quality, in terms of response presence and validity, to the race/Hispanic origin and ancestry questions? How does response quality vary by mode and by race and Hispanic origin?
2. How many people have ancestry responses that matched exactly, half matched (one of the two ancestries match exactly), or did not match to race/Hispanic origin responses? How did these match rates vary by response mode?

---

<sup>18</sup> The median county-level item missing data rate for ancestry across all 3,143 U.S. counties in 2008-2012 was 14.4 percent. This was the 8<sup>th</sup> highest rate of the 125 ACS questions analyzed.

3. Of the ancestry responses that did not match the reported race/Hispanic origin, how often were the ancestry responses in the same OMB race/Hispanic origin category? When in the same OMB category, how often did the reported ancestry provide more specificity, the same level of specificity, or less specificity as the reported race/Hispanic origin?

The second part of our analysis looks into how deriving ancestry products with detailed data from the revised race and Hispanic origin question would influence the estimates for a wide range of groups.

4. Of the people who identified with a particular group in either the race/Hispanic origin question or ancestry question, what percent identified with the group (1) in both responses, (2) in only the detailed race/Hispanic origin response, or (3) in only the ancestry response?
5. How do group estimates derived from the 2016 ACS Content Test race and Hispanic origin questions compare to those derived from the production ancestry question?

## **3.2 Data and Methods**

The following section describes the data we used for this analysis, some limitations of the data, and the methods we used to answer each research question.

### **3.2.1 Data**

For this analysis, we used person-weighted data from two datasets:

- (1) 2016 ACS Content Test control treatment data from the race, ethnicity, and ancestry questions<sup>19</sup>
- (2) 2016 ACS production data from the ancestry question

The 2016 ACS Content Test consisted of a nationally representative sample of 70,000 residential addresses in the United States, independent of the production ACS sample. The 2016 ACS Content Test sample universe did not include GQs, nor did it include housing units in Alaska, Hawaii, or Puerto Rico. The 2016 ACS Content Test had a sample of 35,000 control cases and 35,000 test cases; approximately 20,000 housing units responded from each treatment.

We utilized 1-year 2016 ACS production data but excluded data from Alaska, Hawaii, Puerto Rico, and GQs in order to better align the production data with the 2016 ACS Content Test data. The only survey response data that we analyzed from ACS production was the ancestry survey item; the production ancestry data was the baseline against which we compared estimates from the

---

<sup>19</sup> Data from the test treatment are shown in Appendix F.

2016 ACS Content Test. We used ACS production data for the group comparisons in Research Question 5 in order to control for any Content Test effects on ancestry survey item response.

All estimates from the 2016 ACS Content Test were weighted using the final weights from the Content Test, which take into account the initial probability of selection (the base weight) and CAPI subsampling. These are different from production ACS final weights, however, which also control for population estimates, seasonal variations in response patterns, nonresponse bias, and under-coverage bias.<sup>20</sup> ACS production data were edited or imputed for respondents who did not provide a race or Hispanic origin response, while the 2016 ACS Content Test data were not.

It also should be noted that households were rostered at the beginning of the questionnaire or interview to include everyone residing in the household. In many cases, there is one person who answers the survey questions not only for themselves, but also for other or all of the people in the household. In this analysis, when we refer to “respondents,” “individuals,” or “people” reporting an ancestry, race, or Hispanic origin, we included all individuals rostered in the survey, regardless of whether they provided their own responses or if their responses were provided by a proxy.

### **3.2.2 Coding and Detailed Group Construction**

The *2015 Race, Ethnicity, and Origin Code List* was developed to evaluate responses to the race and Hispanic origin question in the 2015 NCT research. This code list was also used for the race and Hispanic origin responses from the 2016 ACS Content Test. The *Race, Ethnicity, and Origin Code List* applies three-digit numeric or alphanumeric codes to race and ethnicity groups (e.g., 335 for a write-in of “Ghanaian” or A24 for a write-in of “Arapaho”). An important note is that each group is associated with a major OMB race/Hispanic origin category. For example, codes 300-399 of the code list contained all Black or African American detailed groups (e.g. Ethiopian or Jamaican). In our analysis, we included all groups indicated by the respondent that represented a major category, as well as the associated major category of any provided detailed group. For example, if a respondent only checked “Black” and wrote “Cuban” into the associated field, we applied both Black and Hispanic as major categories (because Cuban is associated with Hispanic as the major category), and Cuban as the detailed category. Additionally, some write-in responses do not fall within a code list grouping and are considered “Unclassified” (e.g., write-ins such as “Adopted” or “Refused”).

The ACS production *Ancestry Code List* was used to evaluate the ancestry responses and has been used since the 1980 Census with minor updates since then.<sup>21</sup> The *Ancestry Code List* employed

---

<sup>20</sup> ACS production sample design and weighting are described in Chapters 4 and 11, respectively, of the American Community Survey Design and Methodology Report (U.S. Census Bureau, 2014). For more detail on how the 2016 ACS Content Test sample design differed from ACS production, see the 2016 ACS Content Test Evaluation Report for Race and Ethnicity (Harth et al., 2017).

<sup>21</sup> The ACS Ancestry Code List is publically available at [https://www2.census.gov/programs-surveys/acs/tech\\_docs/code\\_lists/2016\\_ACS\\_Code\\_Lists.pdf](https://www2.census.gov/programs-surveys/acs/tech_docs/code_lists/2016_ACS_Code_Lists.pdf).

only numeric three-digit codes and is more specific in some areas than race while less specific in other areas. For instance, the *Ancestry Code List* is less specific for AIAN responses than the *Race, Ethnicity, and Origin Code List* but more specific for some European heritages. Since all AIAN data from the ACS come from the race question, there is not a need for the ancestry data to capture every distinct tribe. Instead, most write-in responses to ancestry that are an AIAN tribe receive the same generic ancestry code, regardless of how specific the write-in is. However, in ancestry coding, the groups Italian, Bolognese, Roman, and Umbrian are assigned four distinct codes. Those groups would all be coded with a single code for Italian in race.

For this analysis, each ancestry code was associated with one and only one race/ethnicity code that was considered to be the corresponding match. This crosswalk was developed by subject matter experts at the Census Bureau. When comparing the ancestry response to the race/Hispanic origin response, only the codes were compared using the assigned crosswalk mapping, without regard to the original written text. For the analyses in this paper that include ancestry data tabulated into major race/Hispanic origin groups, we assigned major race/Hispanic origin groups to detailed ancestry responses in a similar way as described above, using the same coding structure as in the *Race, Ethnicity, and Origin Code List*.

It is important to note that neither the *Ancestry Code List* nor the *Race, Ethnicity, and Origin Code List* align exactly with the definitions used when reporting ancestry in the published data products. For example, the estimate for Irish in AFF Table B04006 includes people that responded with either “Irish” or “Northern Irish” to the ancestry question. The association of respondent-provided ancestry codes to AFF data product groups is provided in a coding specification, which we followed when building AFF Table B04006 groups from ancestry data in this analysis. For groups in our analysis that are not included in AFF Table B04006, we consulted with subject matter experts at the Census Bureau to confirm the corresponding ancestry code(s). To associate race codes with the ancestry groups in AFF Table B04006 for our additional analysis, we again consulted with subject matter experts, the crosswalk of race-to-ancestry mapping, and occasionally adapted AFF Table B04006 to work for our data. For example, in the *Ancestry Code List*, West Indian, Dutch West Indian, British West Indian, and French West Indian all have separate codes, several of which are distinct rows in AFF Table B04006. In the *Race, Ethnicity, and Origin Code List*, however, all four groups have the same code, making it impossible to directly compare the individual groups. In such circumstances, we combined certain groups to be able to make direct comparisons between the race data and the ancestry data.

One important difference between the race/Hispanic origin and ancestry questions was the coding of the write-in fields in data processing. In both the race and Hispanic origin questions, a maximum of 10 codes were assigned to each write-in field, essentially based on the first 10 codable groups provided. The ancestry question is open-ended in all ACS modes. However, open-ended questions present obvious coding challenges of responses. The ACS ancestry coding rules

attempt to decipher the first two detailed responses provided. A maximum of two ancestries are coded for each respondent.

### 3.2.3 Research Question Metrics

This section includes descriptions of the metrics used to answer the research questions detailed above. All proportions use weighted numbers.

1. *What is the response quality, in terms of response presence and validity, to the race/Hispanic origin and ancestry questions? How does response quality vary by mode and by race and Hispanic origin?*

For this research question, we describe the interaction of race/Hispanic origin responses with ancestry responses for each person. Each person is classified into one of the following:

- Provided a valid response to both the race/Hispanic origin series and the ancestry question.
- Provided a valid response to the race/Hispanic origin series but the ancestry question was completely blank.
- Provided a valid response to the race/Hispanic origin series but the ancestry question was an uncodable response.
- Provided no response to the race/Hispanic origin series and no response to the ancestry question.
- Provided a different combination of responses, such as leaving the race/Hispanic origin series entirely blank but providing a valid response to ancestry.

The resulting proportions for Research Question 1 are defined in Equation 1.1.

$$\begin{array}{l}
 \text{[1.1] Percent of} \\
 \text{People who} \\
 \text{Provided a Certain} \\
 \text{Level of Response} \\
 \text{Presence and} \\
 \text{Validity}
 \end{array}
 =
 \frac{\begin{array}{l}
 \text{\# of people with a given classification for response} \\
 \text{presence and validity to the race/Hispanic origin and} \\
 \text{ancestry questions in the ACS Content Test treatment} \\
 \text{in a given mode}
 \end{array}}{\begin{array}{l}
 \text{\# of people in the ACS Content Test treatment in a} \\
 \text{given mode}
 \end{array}}
 *100$$

2. *How many people have ancestry responses that matched exactly, half matched (one of the two ancestries matched exactly), or did not match to race/Hispanic origin responses? How did these match rates vary by response mode?*

For those people with valid data in both questions, the following possible categories were defined in order to compare race/Hispanic origin information to ancestry information.

- Exact match – this classification is applicable when the information found in ancestry was also found in race/Hispanic origin as prescribed by the crosswalk identified by subject matter experts within the Census Bureau. Exact matches are subsequently broken out by whether one or two ancestry codes were provided. If two ancestry codes were provided, then both had to be in the race/Hispanic origin data in order to be considered an exact match.
- Half match – this classification is only applicable to persons who provided two ancestry codes and is applied when one ancestry code is found to be consistent with the race/Hispanic origin data while the other ancestry code is not consistent.
- No match -- this classification is applicable when the ancestry code(s) are not considered to be consistent with the race/Hispanic origin data as prescribed by the crosswalk file. These are subsequently broken out by whether one or two ancestry codes were provided. If two ancestry codes were provided, then neither code was found to be consistent with the race/Hispanic origin data.

The resulting proportions for Research Question 2 are defined in Equation 2.1.

$$\begin{aligned}
 & \text{[2.1] Percent of} \\
 & \text{People who [exactly} \\
 & \text{matched/half} \\
 & \text{matched/did not} \\
 & \text{match] Between} \\
 & \text{Race/Hispanic origin} \\
 & \text{and Ancestry} \\
 & = \frac{\text{\# of people with valid data who [exactly} \\
 & \quad \text{matched/half matched/did not match] between the} \\
 & \quad \text{ancestry responses and the race/Hispanic origin} \\
 & \quad \text{responses in a given mode}}{\text{\# of people who provided both a valid race/Hispanic} \\
 & \quad \text{origin and a valid ancestry in the ACS Content Test} \\
 & \quad \text{treatment in a given mode}} * 100
 \end{aligned}$$

□

$$\begin{aligned}
 & \text{[2.2] Percent of} \\
 & \text{Ancestries that} \\
 & \text{(did/did not) Match} \\
 & \text{Exactly to a} \\
 & \text{Response in} \\
 & \text{Race/Hispanic origin} \\
 & = \frac{\text{\# of ancestries in a given mode that (did/did not)} \\
 & \quad \text{match to a reported race/Hispanic origin response}}{\text{\# of ancestries provided in a given mode}} * 100
 \end{aligned}$$

3. *Of the ancestry responses that did not match the reported race/Hispanic origin, how often were the ancestry responses in the same OMB race/Hispanic origin category? When in the same OMB category, how often did the reported ancestry provide more specificity, the same level of specificity, or less specificity as the reported race/Hispanic origin?*

The OMB categories used in this analysis are White, Black, Asian, AIAN, NHPI, Some Other Race (SOR) and Hispanic. The resulting proportions for the first part of the research question are defined in the equations (3.1, 3.2) below.

$$\begin{aligned}
 & \text{[3.1] Percent of} \\
 & \text{Ancestries that} \\
 & \text{(did/did not) Match} \\
 & \text{Exactly to a} \\
 & \text{Response in} \\
 & \text{Race/Hispanic origin,} \\
 & \text{by major category} \\
 & = \frac{\text{\# of ancestries that (did/did not) match to a reported} \\
 & \quad \text{race/Hispanic origin response}}{\text{\# of ancestries provided that classify as a given major} \\
 & \quad \text{category}} * 100
 \end{aligned}$$

$$\begin{aligned}
 & \text{[3.2] Percent of} \\
 & \text{Ancestries that did} \\
 & \text{not Match Exactly to} \\
 & \text{a Response in} \\
 & \text{Race/Hispanic origin} \\
 & \text{but (were/were not)} \\
 & \text{in the same major} \\
 & \text{category} \\
 & = \frac{\text{\# of ancestries that were not an exact match but} \\
 & \quad \text{(did/did not) have the same major race classification as} \\
 & \quad \text{a race/Hispanic origin response}}{\text{\# of ancestries that classify as a given major category}} * 100
 \end{aligned}$$

The second half of this research question looks at the specificity of responses. All ancestry responses and all race/Hispanic origin responses were classified by how specific the information was. There were four possible classifications, outlined below. They are presented from most specific to least specific.

- Specific: Responses such as “German,” “Arab,” “Cherokee,” etc., were considered specific responses
- Regional: Responses such as “European,” “Caribbean,” etc., were considered regional responses, a somewhat-specific level of response
- Generic: Responses such as “White,” “Caucasian,” “Black,” “African American,” “Asian,” etc., were considered generic responses, the least specific level possible within a major category
- Unclassified: Responses such as “American,” “Mixed,” “Jewish,” etc., were not used in this analysis because they are not associated with any distinct major race/Hispanic origin group

After classifying the specificity of all ancestry and all race/Hispanic origin responses, we compared the specificity of each ancestry response to the specificity found in race/Hispanic origin of responses from the same major category.



To illustrate this analysis, consider the following example of one person’s response to the ACS:

- The White checkbox was marked in race (the least specific type of response) and the Chinese checkbox was also marked (a specific response),
- The ancestry write-in yielded both German (a specific response) and Chinese (a specific response).

Then the following comparisons and conclusions would be made:

- The ancestry response of German would be compared only to the race/Hispanic origin response of White, since they are in the same major category (White).
  - German does not exactly match any reported race/Hispanic origin responses,
  - German is in the same OMB category as a reported race/Hispanic origin response,
  - German is more specific than the race/Hispanic origin response of White.
- The ancestry response of Chinese would be compared only to the race/Hispanic origin response of Chinese, since they are in the same major category (Asian). Since the ancestry response does exactly match to race/Hispanic origin information, no further analysis would be conducted.

Equation 3.3 defines the proportions used to analyze specificity of responses.

$$\begin{array}{l}
 \text{[3.3] Percent of} \\
 \text{Ancestries that were} \\
 \text{(more specific/less} \\
 \text{specific/the same} \\
 \text{amount of} \\
 \text{specificity) than} \\
 \text{what was found in} \\
 \text{race/Hispanic origin} \\
 \text{responses}
 \end{array}
 = \frac{\begin{array}{l} \text{\# of ancestries that were not an exact match but did} \\ \text{have the same major race classification as was found} \\ \text{in race/Hispanic origin responses with the same major} \\ \text{category and (were more specific/were less} \\ \text{specific/had the same amount of specificity)} \end{array}}{\begin{array}{l} \text{\# of ancestries that classify as a given major category} \end{array}} * 100$$

4. *Of the people who identified with a particular group in either the race/Hispanic origin question or ancestry question, what percent identified with the group (1) in both responses, (2) in only the detailed race/Hispanic origin response, or (3) in only the ancestry response?*

For this research question, we looked at each major race/Hispanic origin category, as well as the six largest detailed groups within each major category. To construct each detailed group for this research question, we combined people who indicated the group in either the race/Hispanic origin questions *or* the ancestry question. We then divided that group into three mutually exclusive categories: 1) those who indicated the group in both questions, 2) those who indicated the group only in the race/Hispanic origin questions, and 3) those who indicated the group only in the

ancestry question. For each group, we statistically compared the difference between those who reported the group in only race/Hispanic origin (4.2) and those who reported the group in only ancestry (4.3).

$$\begin{aligned}
 & \text{[4.1] Percent that} \\
 & \text{Reported the Group} \\
 & \text{in Both} \\
 & \text{Race/Hispanic Origin} \\
 & \text{and Ancestry} \\
 & \text{Responses} \\
 & = \frac{\text{\# of people that reported the particular group in both} \\
 & \quad \text{the race/Hispanic origin and ancestry questions}}{\text{\# of people reported the particular group in either the} \\
 & \quad \text{race/Hispanic origin or ancestry questions}} * 100
 \end{aligned}$$

$$\begin{aligned}
 & \text{[4.2] Percent that} \\
 & \text{Reported the Group} \\
 & \text{in Race/Hispanic} \\
 & \text{Origin Only} \\
 & = \frac{\text{\# of people that reported the particular group in the} \\
 & \quad \text{race/Hispanic origin question and not the ancestry} \\
 & \quad \text{question}}{\text{\# of people reported the particular group in either the} \\
 & \quad \text{race/Hispanic origin or ancestry questions}} * 100
 \end{aligned}$$

$$\begin{aligned}
 & \text{[4.3] Percent that} \\
 & \text{Reported the Group} \\
 & \text{in Ancestry Only} \\
 & = \frac{\text{\# of people that reported the particular group in the} \\
 & \quad \text{ancestry question and not the race/Hispanic origin} \\
 & \quad \text{question}}{\text{\# of people reported the particular group in either the} \\
 & \quad \text{race/Hispanic origin or ancestry questions}} * 100
 \end{aligned}$$

5. *How do group estimates derived from the 2016 ACS Content Test race and Hispanic origin questions compare to those derived from the production ancestry question?*

To answer Research Question 5, we categorized the ancestry data using three different formats. First, we examined the data for the groups that are shown only on AFF Table B04006. Second, we compared other groups not found in B04006, including several detailed Hispanic, Asian, and NHPI groups, plus other groups that would be racially classified as SOR. And third, we examined groups aggregated by world region. For all three formats, we compared 2016 production ACS estimates with estimates from the 2016 ACS Content Test, and estimates from both sources followed formula 5.1. For all estimates, the denominator includes all survey respondents, whether or not a respondent answered the particular question under analysis. To ease comparison of estimates across and within data sources, we report estimates as percentages.

In the first part of our analysis for Research Question 5, we compared many of the groups included in AFF Table B04006. To render the data sources comparable, however, it was at times necessary

to alter how the groups were formed (as described with the “West Indian” group in Section 3.2.2). In the second part of our analysis for this research question, we examine groups from all racial categories as well as Hispanic origin groups. We included “unclassified” and “missing or not reported” as groups in this part of the analysis because they are included as a combined row in AFF Table B04006, even though it could be argued that they are not “ancestry groups” in the traditional sense. We separated them in our analysis to better gauge differences between the data sources. In the third part of our analysis for Research Question 5, we combined groups together by world geographic area based on the ACS Ancestry code list, with some modifications.<sup>22</sup> If a group did not fit cleanly into a geographic area, it was not included in the analysis (e.g. responses such as “White” or “Black” were excluded from the third part of the analysis for Research Question 5). The rows in the tables for Research Questions 4 and 5 will not add to 100 percent because people appeared in multiple columns. In addition, we excluded the catchall “Other Groups” row that appears in AFF Table B04006 from this analysis.

$$\begin{array}{l}
 \text{[5.1] Percent of} \\
 \text{People in a Detailed} \\
 \text{Race/Hispanic Origin} \\
 \text{or Ancestry Group}
 \end{array}
 = \frac{\text{\# of people whose response was in a particular group}}{\text{\# of survey respondents}} * 100$$

### 3.2.4 Standard Error Calculation and Statistical Testing

When applicable, variances were calculated using the Successive Differences Replication method with replicate weights, which is the standard method used in the ACS (see U.S. Census Bureau, 2014, Chapter 12). The variance for each estimate and difference was calculated with the formula below.

$$Var(X_o) = \frac{4}{80} \sum_{r=1}^{80} (x_r - x_o)^2$$

where:

$X_o$  = the estimate calculated using the full sample,

$X_r$  = the estimate calculated for replicate  $r$ .

The standard error (SE) of the estimate ( $X_o$ ) is the square root of the variance. In the report tables, we show margins of error (MOEs) calculated with a 90 percent confidence interval, using the following formula:

$$\text{Margin of Error} = \text{Standard Error} \times 1.645$$

---

<sup>22</sup> Our list of categories differs from the ancestry codelist in a few ways. The regional breakdown in the ancestry codelist separates out Hispanic categories; we have divided these groups up by geography. We also included most of the Asian categories into a single group, except those countries typically classified as “Central Asia.”

In some research questions, we conducted statistical testing to identify differences between rates or percentages. For these comparisons, we first calculated p-values using two-tailed tests for differences at the  $\alpha = 0.10$  level. Then, because we were often comparing many percentages simultaneously that share a common universe, we adjusted the calculated p-values for multiple comparisons using the Holm-Bonferroni method (Holm, 1979).

#### 4. ASSUMPTIONS AND LIMITATIONS

When comparing race and Hispanic origin data to ancestry data in the 2016 ACS Content Test, there are a few methodological differences between the survey items to consider, beyond the question design and wording.

- On the internet version, people who did not respond to the race and Hispanic origin questions received a prompt to supply a response before moving on to the next question. This did not happen for the ancestry question, which could have resulted in higher nonresponse for ancestry compared with race and Hispanic origin in the internet mode.
- An individual could have identified a maximum of 13 detailed groups in the Hispanic origin question and a maximum of 59 detailed groups in the race question on the paper questionnaire (more could be captured in automated instruments). Such an outcome would have been possible if every checkbox had been marked and every write-in captured the maximum of ten coded answers. However, the maximum number of ancestries that could be captured, regardless of how much was written, was two. Therefore, some respondents had a higher number of race/Hispanic origin responses than ancestry responses.
- The modified race/Hispanic origin questions on the 2016 ACS Content Test may have affected responses to the ancestry question due to people being provided examples of detailed White and Black origins and being asked to provide detailed responses earlier in the questionnaire. For this reason, when we compare group estimates created from race/Hispanic origin data to those created from ancestry data in Research Question 5, we use production ancestry data.
- Most American Indian and Alaska Native groups, such as individual tribes, tribal groupings, villages and corporations were not retained in the ancestry data. Therefore, only the overall AIAN total will be shown from ancestry data.

The 2016 ACS Content Test methodology differed in several ways from that of production ACS:

- The 2016 ACS Content Test was not conducted in Alaska or Hawaii, while production data covers all of the United States. In order to address this difference, we excluded data from Alaska and Hawaii from the ACS production data.<sup>23</sup> Thus, conclusions should not be drawn

---

<sup>23</sup> The 2015 National Content Test included Alaska and Hawaii. The results of the analysis of detailed reporting for American Indian, Alaska Native, Native Hawaiian, and Pacific Islander groups can be found in the report by Mathews, et al, 2017.

from this paper for the American Indian, Alaska Native, Native Hawaiian, and Pacific Islander populations, or groups that tend to reside in Hawaii or Alaska.

- The 2016 ACS Content Test data were collected from March to May of 2016, while the ACS production data used in this report come from the entire 2016 tabulation year. This shorter time interval may affect the data for areas that have seasonal population fluctuations. To mitigate this effect, we performed all analyses at the national-level. However, national-level analyses leave unexplored variation in smaller-level geographies.
- During the 2016 ACS Content Test data collection period, interviewers were assigned both control and test cases for the Content Test, as well as production cases. While interviewers were trained to read questions verbatim, administering three slightly different versions of the ACS could still have led to unintentional miscues by interviewers.
- The 2016 ACS Content Test data were collected in English and Spanish only. ACS production data are collected in 11 languages in the self-response modes, and even more in the interviewer-assisted modes. Since language may be correlated with self-identification in the race/Hispanic origin and ancestry items, this could be a source of bias in the results.
- The 2016 ACS Content Test data were not edited or imputed. The ACS production data undergo thorough editing and imputation prior to release.
- The 2016 ACS Content Test data went through a less complex weighting application than the production data.<sup>24</sup>

## 5. RESULTS

This section presents the results of analysis on the ACS ancestry question in conjunction with the questions on race and Hispanic origin. Research Questions 1 to 4 analyze data strictly from the 2016 ACS Content Test control treatment. Research Question 5 compares 2016 ACS production data to 2016 ACS Content Test control treatment data.

### 5.1 Research Question 1

*What is the response quality, in terms of response presence and validity, to the race/Hispanic origin and ancestry questions? How does response quality vary by mode and by race and Hispanic origin?*

To assess the respondent burden that can result from answering similar questions on the same survey, we analyzed data from the 2016 ACS Content Test control treatment. A previous analysis of that data presented the item nonresponse rates for Hispanic origin, race, and ancestry (Harth et al., 2017). The results showed that:

---

<sup>24</sup> For more discussion on the weighting, please see Section 3.2.1.

- 1.7 percent skipped the Hispanic origin question or provided only an uncodable answer,
- 1.4 percent skipped the race question or provided only an uncodable answer,
- 0.4 percent skipped both Hispanic origin and race or provided only an uncodable answer in that series, and
- 13.2 percent did not respond to the ancestry question.

These results are in line with historical data on nonresponse to these questions.<sup>25</sup>

However, minimal analysis has been done on the interaction of responses to race and ancestry. For this research question, we developed categories to more completely describe the interaction of race/Hispanic origin responses with ancestry responses. The categories are:

- The individual provided a valid response to both the race and Hispanic origin series and the ancestry question.
  - *Valid response for the race and Hispanic origin series* means that a response was provided in either or both of the questions that allowed for the person to be categorized as a race or Hispanic group. Notably, a person who marked the checkbox for “No, not of Hispanic, Latino, or Spanish origin” and provided no other useful information in race is not considered a valid response for this analysis.
  - *Valid response for ancestry* means that text was provided in the write-in and it was not classified as uncodable text (examples of write-in responses that are classified as uncodable include ‘Unknown’, ‘NA’, or ‘Already Answered’).
- The individual provided a valid response to the race and Hispanic origin series but the ancestry question was an uncodable response.
- The individual provided a valid response to the race and Hispanic origin series but the ancestry question was completely blank.
- The individual provided no response to both the race and Hispanic origin series and to the ancestry question.
- The individual provided a different combination of responses, such as leaving the race and Hispanic origin series entirely blank but providing a valid ancestry.

Table 3 presents the results from the 2016 ACS Content Test control treatment. The first column of results shows the distribution of all person records, followed by distinct columns for the distribution from each of the four response modes used.

---

<sup>25</sup> See allocation rates table at [https://www.census.gov/acs/www/methodology/sample-size-and-data-quality/item-allocation-rates/index.php#basic\\_demographics](https://www.census.gov/acs/www/methodology/sample-size-and-data-quality/item-allocation-rates/index.php#basic_demographics).

**Table 3: Presence and Validity of Race/Hispanic Origin and Ancestry Responses**

<b>Validity of Response to Race/Hispanic Origin and Ancestry</b>	<b>All Persons (N=43,500)</b>	<b>Internet (N=21,000)</b>	<b>Mail (N=10,500)</b>	<b>CATI (N=1,900)</b>	<b>CAPI (N=10,500)</b>
Valid response to both	84.7 (0.7)	83.0 (0.8)	79.3 (1.0)	87.1 (2.8)	89.3 (1.3)
One ancestry provided	56.1 (1.0)	47.1 (1.1)	52.5 (1.5)	58.5 (3.1)	68.0 (2.1)
Two ancestries provided	28.6 (0.8)	35.9 (1.1)	26.8 (1.3)	28.6 (2.8)	21.2 (1.9)
Valid Race/Hispanic origin, uncodable Ancestry	1.8 (0.2)	2.8 (0.4)	2.6 (0.5)	0.4 (0.3)	0.3 (0.2)
Valid Race/Hispanic origin, blank Ancestry	12.6 (0.6)	13.3 (0.7)	17.2 (1.0)	11.5 (2.7)	9.6 (1.2)
Nonresponse to both	0.6 (0.2)	0.8 (0.2)	0.4 (0.1)	0.3 (0.3)	0.5 (0.3)
Other combinations	0.3 (0.1)	0.2 (0.1)	0.6 (0.2)	0.7 (0.6)	0.2 (0.3)
<b>TOTAL</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>

Source: U. S. Census Bureau, 2016 American Community Survey Content Test, Control Treatment.

Notes: Margin of error is provided in parentheses. Minor additive discrepancies are due to rounding. DRB Approval Number CBDRB-FY19-RAGLIN-B0005.

Overall, 84.7 percent of people provided a valid response to both the race/Hispanic origin questions and the ancestry question while 12.6 percent provided a valid race/Hispanic origin but left ancestry blank. Another 1.8 percent of people provided a valid race/Hispanic origin response but an uncodable ancestry response. Less than one percent of people did not respond to any of the race, Hispanic origin, or ancestry questions (0.6 percent).

Individuals who responded through the mail were the most likely to provide a valid race/Hispanic origin but then leave ancestry blank (17.2 percent). Individuals who used self-response modes (internet and mail) were the most likely to provide a valid race/Hispanic origin but then give an uncodable response to ancestry (2.8 percent in internet and 2.6 percent in mail compared to 1.8 percent overall).

Since the following research questions analyze cases where both valid ancestry and valid race/Hispanic origin data were provided, we wanted to understand whether people in certain race or Hispanic origin groups tended to not respond to ancestry as often as the total population. If certain groups responded less frequently to ancestry, our results would be limited. For example, if people with Hispanic origin were more likely to skip the ancestry question, our comparisons of ancestry to race/Hispanic reports would not be as applicable to that group.

Table 4 shows the distribution of Hispanic origin for persons in each category (when sufficient information was provided), and the distribution of Hispanic origin for all persons in the bottom row of the table. Only race/Hispanic origin data (not ancestry) was used to determine whether someone was of Hispanic origin. The table rows are retained from Table 3 with rows summing to 100 percent. This analysis seeks to show whether the people who provided valid responses to

both survey questions resemble the overall population who provided valid responses to both questions.

**Table 4: Distribution of Hispanic Origin By Validity of Race/Ancestry Data**

<b>Validity of Response to Race/Hispanic Origin and Ancestry</b>	<b>N</b>	<b>Hispanic</b>	<b>Not Hispanic</b>	<b>Total</b>
Valid response to both	36,000	18.5 (0.9)	81.5 (0.9)	<b>100.0</b>
One ancestry provided	25,000	22.3 (1.2)	77.7 (1.2)	<b>100.0</b>
Two ancestries provided	11,000	11.0 (1.1)	89.0 (1.1)	<b>100.0</b>
Valid Race/Hispanic origin, blank Ancestry	6,200	11.4 (1.5)	88.6 (1.5)	<b>100.0</b>
Valid Race/Hispanic origin, uncodable Ancestry	850	2.6 (1.8)	97.4 (1.8)	<b>100.0</b>
Nonresponse to both	300	(X)	(X)	(X)
Other combinations	100	(X)	(X)	(X)
<b>Overall</b>	<b>43,671</b>	<b>17.1 (0.8)</b>	<b>82.9 (0.8)</b>	<b>100.0</b>

Source: U. S. Census Bureau, 2016 American Community Survey Content Test, Control Treatment.

Notes: Margin of error is provided in parentheses. An '(X)' means that the estimate is not applicable or not available. Minor additive discrepancies are due to rounding. DRB Approval Number CBDRB-FY19-RAGLIN-B0005.

While 17.1 percent of all people in the 2016 ACS Content Test control treatment were of Hispanic origin, 18.5 percent of people who provided a valid response to both questions were of Hispanic origin. The difference is not statistically significant so there is no evidence that our research is less applicable to those of Hispanic origin.

Table 5 shows the racial distribution of persons in each category delineated in Table 3. The racial distribution of all persons in the control treatment is shown in the bottom row of Table 5. Only race/Hispanic origin data (not ancestry) was used to determine someone's race. Multiracial individuals were included in multiple rows so the racial distributions do not add to 100 percent.



**Table 5: Distribution of Race By Validity of Race/Ancestry Data**

<b>Validity of Response to Race/Hispanic Origin and Ancestry</b>	<b>White</b>	<b>Black</b>	<b>Asian</b>	<b>AIAN</b>	<b>NHPI</b>	<b>SOR</b>
Valid response to both	73.9 (0.9)	13.3 (0.7)	7.3 (0.6)	4.4 (0.5)	0.4 (0.2)	8.5 (0.7)
One Ancestry provided	63.8 (1.1)	16.8 (1.0)	8.9 (0.8)	2.8 (0.6)	0.3 (0.3)	10.2 (0.8)
Two Ancestries provided	93.7 (0.8)	6.5 (0.8)	4.2 (0.6)	7.6 (0.8)	0.7 (0.3)	5.2 (0.7)
Valid Race/Hispanic origin, blank Ancestry	78.0 (1.9)	10.5 (1.4)	5.4 (1.2)	2.6 (0.7)	0.5 (0.3)	7.3 (1.3)
Valid Race/Hispanic origin, uncodable Ancestry	88.5 (3.6)	7.2 (3.2)	2.2 (1.6)	2.9 (1.6)	0.2 (0.3)	5.0 (2.8)
Nonresponse to both	(X)	(X)	(X)	(X)	(X)	(X)
Other combinations	(X)	(X)	(X)	(X)	(X)	(X)
<b>Overall</b>	<b>74.0 (0.8)</b>	<b>12.8 (0.6)</b>	<b>6.9 (0.6)</b>	<b>4.1 (0.4)</b>	<b>0.4 (0.2)</b>	<b>8.2 (0.6)</b>

Source: U. S. Census Bureau, 2016 American Community Survey Content Test, Control Treatment.

Notes: Margin of error is provided in parentheses. Race categories are not mutually exclusive. An '(X)' means that the estimate is not applicable or not available. Minor additive discrepancies are due to rounding. DRB Approval Number CBDRB-FY19-RAGLIN-B0005.

Of the people in the control treatment of the 2016 ACS Content Test, 74.0 percent identified at least partially as White. Of the people who provided a valid response to both questions, 73.9 percent were White. The difference between those results is not statistically significant. No significant differences were found when making the same comparison for the remaining five major groups. There is no evidence that our research is less applicable to particular racial groups.

## 5.2 Research Question 2

*How many people have ancestry responses that matched exactly, half matched (one of the two ancestries matched exactly), or did not match to race/Hispanic origin responses? How did these match rates vary by response mode?*

Initial comparisons of the ancestry data to the race data for those persons with both a valid race/Hispanic origin and a valid ancestry was reported in prior research on the 2016 ACS Content Test (Harth et al., 2017). That research showed how often a person's ancestry data was consistent with what they had reported under race and Hispanic origin. In the control treatment, 72.5 percent of persons provided consistent answers in race/Hispanic origin to what was reported in ancestry.

This research presents a more nuanced look at consistency than was presented in the 2016 ACS Content Test report, in that the previous report only considered exact matches. For those people with valid data in both questions, the following possible categories were defined for this research in order to further compare race/Hispanic origin information to ancestry information.

- Exact match – this classification is applicable when the information found in ancestry was also found in race as prescribed by the crosswalk file. This definition is the same as was used in the analysis of the 2016 ACS Content Test. Examples of responses that matched exactly include:
  - Writing in White in ancestry after marking the White checkbox in race,
  - Writing in Mexican in ancestry after marking the Mexican checkbox in Hispanic origin,
  - Writing in Thai in ancestry after writing in Thai in any race or Hispanic origin write-in line

Exact matches are subsequently broken out by whether one or two ancestry codes were provided. If two ancestry codes were provided, then both had to be in the race and Hispanic origin data in order to be considered an exact match.

- Half match – this classification is only applicable to persons who provided two ancestry codes. A half match is when one ancestry code is found to match with something in the race/Hispanic origin data while the other ancestry code does not match.
- No match - this classification is applicable when the ancestry code(s) are not considered to be consistent with the race/Hispanic origin data as prescribed by the crosswalk file. These are subsequently broken out by whether one or two ancestry codes were provided. If two ancestry codes were provided, then neither code was found to be consistent with the race and Hispanic origin data.

The results to the research question are shown in Table 6 below. Only those persons with a valid response to both ancestry and the race/Hispanic origin series are included in this analysis.

**Table 6. Match Rates for Persons in the Control Treatment of the 2016 ACS Content Test**

Match Results	Overall (N=36,000)	Internet (N=17,000)	Mail (N=8,200)	CATI (N=1,700)	CAPI (N=9,300)
Exact Match	72.5 (0.8)	73.3 (0.9)	48.8 (1.7)	83.6 (2.5)	81.7 (1.5)
One ancestry provided	51.6 (0.9)	45.8 (1.1)	35.7 (1.5)	59.7 (3.3)	64.5 (2.0)
Two ancestries	20.8 (0.8)	27.5 (1.1)	13.1 (1.1)	23.9 (2.9)	17.2 (1.7)
Half Match	7.1 (0.4)	10.1 (0.6)	5.0 (0.6)	7.6 (1.6)	5.0 (0.9)
No Match	20.4 (0.7)	16.6 (0.8)	46.2 (1.7)	8.8 (1.9)	13.3 (1.5)
One ancestry provided	14.6 (0.6)	10.9 (0.7)	30.5 (1.5)	7.5 (1.7)	11.7 (1.4)
Two ancestries	5.8 (0.4)	5.7 (0.5)	15.7 (1.4)	1.3 (0.7)	1.6 (0.6)
<b>Total</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>

Source: U. S. Census Bureau, 2016 American Community Survey Content Test, Control Treatment.

Notes: Margins of error are shown in parentheses. Minor additive discrepancies are due to rounding. DRB Approval Number CBDRB-FY19-RAGLIN-B0005.

As was reported in the 2016 ACS Content Test Report, of all persons who provided a valid response in both ancestry and race/Hispanic origin, 72.5 percent provided ancestry responses that were exactly the same as information found in race/Hispanic origin. This table shows that 7.1

percent of persons who provided a valid response in both ancestry and race/Hispanic origin had a half match and the remaining 20.4 percent had no response in ancestry that also exactly matched to their race/Hispanic origin responses. Note that this research only compares ancestry to race; it does not compare race to ancestry. If a respondent marked more than one race checkbox or provided multiple race write-ins, a single ancestry response would be considered a match as long as one of the race responses matched to the ancestry response. Since there is a maximum of two ancestry codes, we do not assess whether all reported race/Hispanic origin information is found in ancestry.

Individuals who were enumerated on a mail questionnaire were less likely to have exact matches from ancestry to race/Hispanic origin, compared to other modes of response. On mail questionnaires, 15.7 percent of persons provided two ancestries and neither ancestry was found to be reported in race/Hispanic origin. From internet responses, only 5.7 percent of individuals fit that description.

Table 6 showed how ancestry match rates varied when analyzing the data at the person-level. We also analyzed the match rates for each reported ancestry where a person also gave valid race and Hispanic origin information. Table 7 presents the ancestry-level match rates.

**Table 7. Match Rates for all Reported Ancestries**

	<b>Overall (N=47,500)</b>	<b>Internet (N=23,500)</b>	<b>Mail (N=10,500)</b>	<b>CATI (N=2,100)</b>	<b>CAPI (N=11,000)</b>
Ancestries Matched to Race	75.1 (0.7)	77.3 (0.8)	50.0 (1.8)	86.7 (2.1)	83.9 (1.4)
Ancestries Not Matched to Race	24.9 (0.7)	22.6 (0.8)	50.0 (1.8)	13.3 (2.1)	16.0 (1.4)
<b>TOTAL</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>

Sources: U. S. Census Bureau, 2016 American Community Survey Content Test, Control Treatment.

Notes: Margins of error are shown in parentheses. Minor additive discrepancies are due to rounding. DRB Approval Number CBDRB-FY19-RAGLIN-B0005.

Table 7 shows that 75.1 percent of ancestries reported were found to be identical matches to something reported in race or Hispanic origin, while 24.9 percent were not. Ancestries reported on the paper (mail) questionnaire were the least likely to be matched to a race or Hispanic origin response (50.0 percent).

### 5.3 Research Question 3

*Of the ancestry responses that did not match the reported race/Hispanic origin, how often were the ancestry responses in the same OMB race/Hispanic origin category? When in the same OMB category, how often did the reported ancestry provide more specificity, the same level of specificity, or less specificity as the reported race/Hispanic origin?*

Research Question 3 continues looking at the individual ancestry responses and dives deeper into how an ancestry response compared to the reported race and Hispanic origin responses for a person. Research Question 2 showed that 75.1 percent of reported ancestries were found to be exact matches to something reported in race or Hispanic origin. We focus now on the 24.9 percent of ancestries that did not match exactly to what was reported in the race and Hispanic origin series.

One attribute of ancestry responses is the major race or Hispanic origin category that the ancestry response would be classified into if provided under the race question. For example, an ancestry response of Jamaican is associated with the major race category Black. This is of interest since the ancestry responses published in the data products are almost exclusively White or Black. For this analysis, each ancestry response is assigned to only one major category. Ancestry responses that are either religious responses, U.S. states or “American” do not fit into any major category and are considered unclassified for this analysis.

Additionally, each ancestry was classified into one of four hierarchical categories to describe the specificity of the ancestry. The hierarchy, from most specific to least specific, was:

- Specific: Responses such as “German,” “Arab,” “Scotch Irish,” etc., were considered specific responses
- Regional: Responses such as “European,” “Caribbean,” etc., were considered regional responses, a somewhat-specific level of response
- Generic: Responses such as “White,” “Caucasian,” “Black,” “African American,” “Asian,” etc., were considered generic responses, the least specific level possible within a major race category
- Unclassified: Responses such as “American,” “Mixed,” etc., were not used in this analysis because they are not associated with any distinct major race/Hispanic origin group

All information provided in the race and Hispanic origin series was also assessed using this hierarchy of specificity. The specificity seen in ancestry responses was then compared to the specificity of race and Hispanic origin responses within the same major category; an ancestry response of German would be in the same major OMB category as the White checkbox from the race question and German is more specific than White.

Table 8 first identifies the frequency of ancestry responses that matched exactly to a race or Hispanic origin response, by the major race category of the ancestry response; the total column in Table 8 is the same as was reported in Table 7. Table 8 then provides further classification of ancestries that did not match exactly to a race and Hispanic origin response.

- First, we classified the response as one of the following:
  - In the same major race category as something reported in race and Hispanic origin
  - In an entirely different major race category than anything reported in race and Hispanic origin

- An unclassified response (e.g., a write-in such as “Adopted” or “Human”) that is not associated with a major race category
- Second, for ancestries that did not match exactly to an individual’s race and Hispanic origin information but were in the same major race category, we assessed the specificity of the provided information, assessing whether the ancestry response was either:
  - More specific than anything provided in race and Hispanic origin
  - Had the same level of specificity, or was
  - Less specific.

**Table 8. Match Rates, Difference in Major Category, and Specificity of Information Provided by Ancestry Response Compared With Race/Hispanic Origin Response, by Race or Hispanic Origin of the Ancestry Response**

	White (N=26,500)	Black (N=6,000)	Asian (N=2,200)	AIAN (N=1,100)	NHPI (N=90)	SOR (N=800)	Hispanic (N=8,300)	Total (N=47,500)
Ancestry matched	74.8 (0.9)	92.8 (1.6)	90.9 (1.7)	51.6 (6.0)	87.1 (9.6)	37.4 (6.1)	79.9 (2.1)	<b>75.1 (0.7)</b>
Ancestry did not match	25.2 (0.9)	7.2 (1.6)	9.1 (1.7)	48.4 (6.0)	12.9 (9.6)	62.6 (6.1)	20.2 (2.1)	<b>24.9 (0.7)</b>
Unclassified	-	-	-	-	-	-	-	<b>3.0 (0.2)</b>
Different major category	0.6 (0.1)	0.7 (0.4)	2.5 (1.0)	28.7 (4.8)	10.8 (8.5)	57.4 (6.6)	2.3 (0.5)	<b>2.5 (0.2)</b>
Same major category	24.6 (0.9)	6.5 (1.5)	6.6 (1.4)	19.8 (4.2)	2.1 (3.3)	5.3 (3.1)	17.9 (2.2)	<b>19.4 (0.7)</b>
Ancestry more specific	17.9 (0.8)	4.8 (1.3)	0.8 (0.4)	1.0 (0.7)	1.4 (2.2)	0.2 (0.2)	1.3 (0.4)	<b>12.0 (0.6)</b>
Same specificity	6.6 (0.4)	1.7 (0.7)	4.1 (1.0)	17.1 (4.2)	-	-	12.2 (1.9)	<b>6.5 (0.4)</b>
Ancestry less specific	0.1 (<0.1)	-	1.7 (0.8)	1.7 (0.9)	0.6 (1.1)	5.0 (3.1)	4.3 (1.3)	<b>0.9 (0.2)</b>
<b>Total ancestry responses</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>

Source: U. S. Census Bureau, 2016 American Community Survey Content Test, Control Treatment.

Notes: Margins of error are shown in parentheses. An '-' entry in the estimate column indicates that either no sample observations or too few sample observations were available to compute an estimate. Minor additive discrepancies are due to rounding. DRB Approval Number CBDRB-FY19-RAGLIN-B0005.

Looking at the top two rows of Table 8, over 90 percent of Black ancestry responses matched exactly to a response in the race and Hispanic origin series, with Asian and NHPI ancestry match rates also around 90 percent (although the margin of error for NHPI reflects the low sample size). Hispanic ancestry responses matched around 80 percent of the time and White ancestries matched around 75 percent of the time. However, roughly half (51.6 percent) of AIAN ancestry responses matched to a race and Hispanic origin response and 37.3 percent of SOR responses matched. AIAN and NHPI responses should be interpreted with caution due to the omission of Alaska and Hawaii from the 2016 ACS Content Test sample.

As discussed in Appendix G, the crosswalk to match an ancestry response to a race response is not flawless and results based on it are considered conservative. Thus some of the “non-matching” ancestries in Table 8 would be matches if assessed using a more generous crosswalk or if assessed after the data have gone through post-processing. For instance, the largest contributor to SOR non-matches are the ancestry responses of “Indian” and “Mixed”. Appendix G shows how “Mixed” could be a non-match simply due to the conservative nature of the matching. An ancestry response of “Indian” (which receives an initial classification of SOR) will go through post-processing and is often then clarified as Asian Indian based on available race data, which it would match to as an Asian response instead of SOR. Additionally, Appendix G shows how most write-in responses to ancestry that are an AIAN tribe receive the same numeric ancestry code, thus making them indistinct from one another for analysis. These ancestry responses are unable to be successfully matched against the race data, which negatively impacts the match rate for AIAN ancestries.

The Total column in Table 8 shows that of all the ancestries in the 2016 ACS Content Test control treatment,

- 75.1 percent matched exactly to something reported in race and Hispanic origin, while 24.9 percent did not.
- 3.0 percent did not match exactly and provided new information but of an ancestry considered to not be in a major race category (i.e., ancestry was American).
- 2.5 percent did not match exactly and the information provided was in an entirely new major race category. An example is Navajo in ancestry and no indication of AIAN identification in race.
- 12.0 percent did not match exactly but were not only in the same major category as something in race but provided a more specific description of a major race category that had been reported. For instance, Irish in ancestry with only the White checkbox marked in race.
- 6.5 percent did not match exactly but were in the same major category as something reported in race or Hispanic origin with comparably specific descriptions of that major category. For example, someone wrote in “Native American” for ancestry after marking the AIAN checkbox for race.
- 0.9 percent did not match exactly but were in the same major category as something reported in race and Hispanic origin with less specific descriptions than was reported in race and Hispanic origin. For example, if a person had marked the Chinese checkbox in race but their ancestry response was “Asian.”

The distribution of those categories changes based on the major race classification of the reported ancestries. For instance, Table 8 shows while 11.9 percent of all ancestries provided more specific descriptions of a known major race category, for White ancestries it was 18.1 percent. We note

again that AIAN and NHPI responses should be interpreted with caution due to the omission of Alaska and Hawaii from the 2016 ACS Content Test sample.

It appears that persons who provide a non-matching ancestry response and identify as White or Black provide more specific answers in ancestry than in the race and Hispanic origin series. However, non-matching ancestries from the other major race groups provide the same level of specificity in ancestry as in race. For these non-matching responses from Asian, AIAN or Hispanic that have the same level of specificity, the lack of a match might be due either to the process of turning open-ended write-ins into codes for analysis or to the matching approach used in the analysis. For instance, a person who marked the Mexican checkbox in Hispanic origin and provided an ancestry write-in of Oaxaca would be considered to have the same level of specificity in their answers, based on how the coding process and our analysis were structured. Additionally, a person who just indicated they were of Hispanic origin without a more detailed response in the race/Hispanic origin series, and then provided an ancestry write-in of Latino would also not be an exact match but would have the same level of specificity. Thus, a non-matching response is not necessarily indicative of new information being gathered.

#### **5.4 Research Question 4**

*Of the people who identified with a particular group in either the race/Hispanic origin question or ancestry question, what percent identified with the group (1) in both responses, (2) in only the detailed race/Hispanic origin response, or (3) in only the ancestry response?*

In Research Questions 4 and 5, we explore how obtaining detailed group information from the race and Hispanic origin items would potentially impact group estimates and data products. Thus, the estimates presented in the analysis for these questions includes blank, invalid, and otherwise uncodable responses. Research Question 4 looks at the location of major race and Hispanic origin category and selected detailed group responses across survey items on the 2016 ACS Content Test. Table 9 summarizes the findings across the seven major race and Hispanic origin OMB groups. Thus, the rows in Table 9 include all detailed and generic responses that fall under the particular category.

The rate at which reporting occurred in both items, detailed in Table 9, varied across race/Hispanic origin categories between 8.5 and 82.1 percent.<sup>26</sup> Among respondents that reported a particular category in only one location, there was variation in where that data was reported. For White, Black, Asian, and Hispanic major categories, it was very rare for a response to only occur in the

---

<sup>26</sup> In Table 9, the elevated rate of SOR respondents that reported only in Race/Hispanic origin is an artifact of Hispanics tending to mark the SOR checkbox in the race question. In fact, 86 percent of those who provided a generic but not detailed SOR response in the race question also reported as Hispanic. In Table 10, when we subset by those who reported a *detailed* group response, this anomaly disappears.

ancestry item. For all major categories, the race/Hispanic origin item was more successful than the ancestry item at eliciting the response for information that was provided in only one location. Differences for all categories showed more reporting in the race/Hispanic origin item, and all differences were statistically significant.

**Table 9. Location of Major Race and Hispanic Origin Group Responses Across Survey Items**

Major Category	N	Reported in			(Race/HO only) minus (Ancestry only)	Adjusted P-value
		Both Race/HO and Ancestry	Reported in Race/HO Only	Reported in Ancestry Only		
White	29,500	66.4 (1.1)	32.9 (1.2)	0.7 (0.1)	32.3 (1.2)	<0.01 *
Black or African American	7,500	79.9 (1.9)	19.3 (1.9)	0.7 (0.3)	18.6 (1.9)	<0.01 *
Asian	2,900	72.7 (3.3)	25.4 (3.5)	1.9 (0.8)	22.3 (3.8)	<0.01 *
AIAN	2,000	38.9 (4.8)	45.1 (4.5)	16.0 (2.5)	29.1 (5.5)	<0.01 *
NHPI	150	56.4 (17.2)	34.5 (16.0)	9.0 (5.4)	25.5 (16.6)	0.01 *
SOR	4,900	8.5 (1.7)	79.8 (2.6)	11.6 (2.0)	68.2 (4.3)	<0.01 *
Hispanic	9,500	82.1 (1.9)	15.9 (1.9)	2.0 (0.5)	12.0 (1.9)	<0.01 *

Source: U. S. Census Bureau, 2016 American Community Survey Content Test, Control Treatment.

Note: Margins of error are shown in parentheses. Significant at  $\alpha=0.1$  level. P-values with an asterisk (\*) indicate a significant difference between the two rates. P-values have been adjusted for multiple comparisons using the Holm-Bonferroni method. DRB Approval Number CBDRB-FY19-RAGLIN-B0005.

While Table 9 looks at overall race/Hispanic origin reporting, Table 10 focuses on detailed group reporting and more closely examines selected detailed groups. It is important to note that on average, the race/Hispanic origin and ancestry survey items elicited detailed group reporting from a similar proportion of respondents. Of respondents who reported at least one detailed group on either survey item, 76.5 percent reported a detailed group in both items, 11.5 percent reported a detailed group only in race and Hispanic origin, while 12.0 percent reported a detailed group only in ancestry.<sup>27</sup>

There was, however, variation in the location of detailed group reporting across major race/Hispanic origin categories. For those who reported a detailed Hispanic or Asian group, the race/Hispanic origin survey items elicited significantly more overall and detailed group reporting than ancestry. Of those who reported a detailed Hispanic group, 74.9 percent did so in both items. However, a higher percentage reported a detailed Hispanic group in the race/Hispanic origin item (22.9 percent) than in the ancestry item (2.2 percent). Those who reported a detailed Asian group followed a similar trend, with 72.2 percent reporting in both items, 25.7 percent responding only in the race/Hispanic origin item, and only 2.1 percent only in the ancestry item.

<sup>27</sup> We note again that for the purposes of this analysis, we combined responses from the race and Hispanic origin questions as if they were a single item. Thus, an individual that reported a Hispanic group in either race or Hispanic origin items would be considered Hispanic.



White and Black detailed group reporting followed somewhat distinct trends. Above in Table 9, it was shown that the race/Hispanic origin item provided significantly more overall reporting than ancestry for both White (32.9 vs 0.7 percent, respectively) and Black (19.3 vs 0.7 percent, respectively) group responses. However, when looking at detailed group reporting in Table 10, these differences disappeared. Among White detailed group responses, the differences reversed, with the ancestry item eliciting a statistically higher rate of detailed reporting than the race/Hispanic origin item (16.5 vs. 10.7 percent). Among Black detailed group responses, the differences became statistically insignificant. Further analysis showed that for the most part, those who reported either a White or Black detailed group in ancestry but not in race/Hispanic origin marked the respective checkbox in the race item and either left the write-in field blank, or wrote in some other non-detailed response.<sup>28</sup> Thus, for those who reported a White or Black detailed group, the ancestry item was just as or somewhat more successful at eliciting a detailed response than the race/Hispanic origin items, particularly compared to the other major groups.<sup>29,30</sup>

Table 10 shows that overall 11 of the 25 selected detailed groups large enough to be compared had significantly more responses come from the race/Hispanic origin items, while the other 15 groups showed no statistical differences in the location of responses. No detailed groups that had significantly more reporting come from the ancestry survey item. Six of the 11 detailed groups with significant differences also had checkbox response options in the race/Hispanic origin items (Chinese, Filipino, Asian Indian, Vietnamese, Mexican, and Puerto Rican), which may have facilitated higher reporting rates for those groups. However, there were other detailed groups with checkbox options did not show such differences (Korean, Japanese, and Native Hawaiian).

---

<sup>28</sup> For those that reported a White detailed ancestry group, but not a White detailed race/ethnicity group, 95 percent checked the White box in the race question. Of those, 64 percent left the write-in blank. For those who reported a Black detailed ancestry group but not a Black detailed race/ethnicity group, 90 percent marked the Black or African American checkbox. Of those, 56 percent left the write-in blank, and another 31 percent wrote in "Black."

<sup>29</sup> The category of "Some Other Race" also followed this pattern, but this was driven to large extent by the write-in response of "Indian," which is classified as "Some Other Race" in the code list. Our analysis suggests that many of those who wrote in "Indian" (classified as Some Other Race) for ancestry also marked "Asian Indian" (classified as Asian) in the Race question, which were classified as nonmatching responses in our analysis.

<sup>30</sup> It should be noted that in the test treatment, which provided detailed checkbox categories on the internet mode, detailed reporting for White and Black was higher in the race/Hispanic origin survey items. See Table A-9.

**Table 10. Location of Detailed Responses Across Survey Items, by Major Category and Detailed Group**

Major Category or Detailed Group	N	Reported in Both	(Race/HO only)				Adjusted P-value
			Reported in Race/HO Only	Reported in Ancestry Only	minus (Ancestry only)		
Any detailed	36,500	76.5 (0.8)	11.5 (0.7)	12.0 (0.6)	-0.5 (1.0)	1.00	
White, detailed	20,000	72.8 (1.0)	10.7 (0.8)	16.5 (0.8)	-5.9 (1.2)	<0.01	*
German	7,300	63.9 (1.6)	19.3 (1.4)	16.7 (1.3)	2.6 (2.2)	0.83	
Irish	6,000	59.5 (2.1)	25.1 (1.8)	15.3 (1.2)	9.8 (2.2)	<0.01	*
English	5,700	53.4 (2.0)	29.3 (1.8)	17.3 (1.5)	12.0 (2.7)	<0.01	*
Italian	2,600	64.3 (2.9)	16.8 (2.4)	18.9 (2.7)	-2.2 (4.3)	1.00	
Polish	1,400	59.9 (3.3)	24.3 (3.4)	15.9 (2.4)	8.4 (4.8)	0.10	*
French	1,400	50.4 (3.1)	28.3 (2.8)	21.2 (2.9)	7.1 (4.9)	0.35	
Black, detailed	6,200	70.9 (2.3)	13.3 (1.8)	15.8 (1.8)	-2.5 (2.8)	1.00	
African American	5,100	66.5 (2.4)	15.5 (1.9)	18.0 (1.9)	-2.4 (3.0)	1.00	
Jamaican	300	75.5 (8.3)	12.0 (4.9)	12.6 (6.1)	-0.6 (7.4)	1.00	
Haitian	200	59.8 (14.6)	11.3 (8.5)	28.9 (20.8)	-17.5 (28.2)	1.00	
Nigerian	80	47.2 (30.2)	43.4 (33.1)	9.4 (9.5)	34.0 (38.3)	1.00	
Ethiopian	30	82.9 (17.6)	-	-	-	-	
Somali	-	-	-	-	-	-	
Asian, detailed	2,800	72.2	25.7 (3.5)	2.1 (0.8)	23.6 (3.7)	<0.01	*
Chinese	750	74.7	22.3 (7.0)	3.0 (1.3)	19.3 (7.0)	<0.01	*
Filipino	500	77.7	18.6 (4.7)	3.7 (2.3)	14.8 (5.5)	<0.01	*
Asian Indian	550	45.1	53.8 (7.6)	1.1 (0.9)	52.7 (7.9)	<0.01	*
Vietnamese	250	81.4 (7.6)	15.9 (7.1)	2.7 (2.2)	13.2 (7.2)	0.06	*
Korean	250	82.8 (6.8)	13.0 (5.1)	4.2 (5.7)	8.8 (8.4)	1.00	
Japanese	150	67.9 (8.3)	23.8 (7.7)	8.4 (5.0)	15.4 (10.0)	0.25	
NHPI, detailed	150	52.7 (17.8)	38.3 (16.5)	9.0 (5.4)	29.3 (16.8)	0.86	
Native Hawaiian	50	56.8 (15.3)	18.8 (8.5)	24.4 (15.4)	-5.5 (19.7)	1.00	
Samoan	40	78.1 (26.9)	-	-	-	-	
Chamorro	-	-	-	-	-	-	
Tongan	-	-	-	-	-	-	
Fijian	-	-	-	-	-	-	
Marshallese	-	-	-	-	-	-	

(Table continued on next page)

Major Category or Detailed Group	N	Reported in Both		Reported in Race/HO Only		Reported in Ancestry Only		(Race/HO only) minus (Ancestry only)	Adjusted P-value
SOR, detailed	1,300	27.0	(4.7)	31.4	(4.8)	41.5	(5.0)	-10.1 (8.6)	0.87
Brazilian	100	64.1	(14.6)	31.0	(13.5)	4.9	(4.8)	26.1 (14.0)	0.06 *
Belizean	-	-		-		-		-	-
Cabo Verdean	20	88.9	(10.7)	-		-		-	-
Guyanese	50	92.5	(7.4)	0.7	(1.3)	6.9	(6.9)	-6.2 (6.6)	1.00
Creole	-	-		-		-		-	-
North American	-	-		-		-		-	-
Hispanic, detailed	9,000	74.9	(2.2)	22.9	(2.3)	2.2	(0.6)	20.7 (2.5)	<0.01 *
Mexican	5,800	75.8	(2.8)	22.2	(3.0)	2.1	(0.8)	20.1 (3.3)	<0.01 *
Puerto Rican	1,000	65.6	(5.9)	33.0	(5.9)	1.4	(0.7)	31.6 (5.9)	<0.01 *
Cuban	450	64.3	(9.1)	23.1	(7.5)	12.6	(8.7)	10.5 (13.5)	1.00
Salvadoran	400	55.4	(9.1)	39.7	(9.7)	4.9	(3.2)	34.8 (11.2)	<0.01 *
Dominican	400	78.6	(6.0)	16.3	(5.8)	5.2	(2.3)	11.1 (6.5)	0.11
Colombian	250	61.3	(15.7)	29.6	(14.7)	9.1	(6.0)	20.5 (16.0)	0.70

Source: U. S. Census Bureau, 2016 American Community Survey Content Test, Control Treatment.

Note: Margins of error are shown in parentheses. Significant at  $\alpha=0.1$  level. P-values with an asterisk (\*) indicate a significant difference between the two rates. P-values have been adjusted for multiple comparisons using the Holm-Bonferroni method. A '-' entry in the estimate column indicates that either no sample observations or too few sample observations were available to compute an estimate. DRB Approval Number CBDRB-FY19-RAGLIN-B0005.

## 5.5 Research Question 5

*How do group estimates derived from the 2016 ACS Content Test race and Hispanic origin questions compare to those derived from the production ancestry question?*

Research Question 5 asks how estimates derived from detailed race and Hispanic origin data from the 2016 ACS Content Test compare to production ancestry estimates of parallel groups. For the first part of the analysis for Research Question 5, we compared estimates represented in AFF Table B04006 “People Reporting Ancestry,” which shows overall reported ancestry and provides the most comprehensive publicly-available source of ancestry group estimates from the ACS.<sup>31</sup> We were interested in how AFF Table B04006 would differ when using race and Hispanic origin data from the 2016 ACS Content Test in place of ACS production ancestry data. Groups that lack a 2016 ACS Content Test estimate in our tables had either no or too few observations to calculate an estimate.

We included 105 groups, 87 of which had sample sizes large enough to be compared in our analysis. Of the 87 groups we compared, 73 were not statistically different (see Appendix Table A-3 for the full table). The 14 groups that showed statistical differences are shown in Table 11. Of the 14 groups that showed statistical differences, the largest change was the “Not Reported or Missing” group, which decreased by from 15.53 to 0.83 percent.<sup>32</sup> Part of this change was due to the fact that the race/Hispanic origin items have lower item nonresponse rates than the ancestry item. Additionally, for this project a response was accepted for race/Hispanic origin if the respondent had responded to either question (rather than both). Other reasons for higher nonresponse in the ancestry item might be that it requires a write-in response, while both the race and Hispanic origin items contain checkboxes with optional write-in fields, and that the ancestry item appears at a much later part of the instrument, potentially more often being avoided by respondents who do not complete the survey or break off early on the instrument.

Some of the groups with largest differences were Dutch, English, German, Irish, and Scottish, which are all Western European groups that already had relatively large ACS production ancestry estimates. Group estimates for English, German, and Irish were all larger by at least three percentage points in 2016 ACS Content Test race/Hispanic origin compared to ancestry in ACS production. The estimate for “Unclassified” was also higher in race/Hispanic origin, likely due to some responses with race connotations lacking a specific ancestry code, such as “Biracial.”

The groups with estimates that were lower in the 2016 ACS Content Test were more diverse in geographic background. These groups included British, Czechoslovakian, Ethiopian, French

---

<sup>31</sup> Since the 2016 ACS Content Test was not carried out in Alaska, Hawaii, Puerto Rico, or in GQs, we also removed these groups from the production ACS estimates.

<sup>32</sup> Estimates in Research Question 5 are reported to the hundredths decimal due to the small size of some groups, where differences could otherwise not be reported in a table format.

Canadian, Kenyan, Haitian, and Pennsylvania German. There was also some evidence of shifting between groups. For example, the estimates for “Pennsylvania German” and “British” were higher in production ancestry, while “German” and “English” were higher in 2016 ACS Content Test race/Hispanic origin.

**Table 11. Comparing Estimates Derived from Race/Hispanic Origin and Ancestry, by Detailed Group, Showing Only Groups that were Significantly Different**

Ancestry or Detailed Race/Hispanic Origin Group	2016 ACS Production Ancestry (N=4,880,000)		2016 ACS Content Test, Race/Hispanic Origin (N=43,500)		Difference		Adjusted P-Value	
	Estimate	Margin of Error	Estimate	Margin of Error	Estimate	Margin of Error	P-Value	Significance
Not reported or missing	15.54	(0.1)	0.83	(0.2)	-14.70	(0.2)	<0.01	*
English	7.47	(<0.1)	11.75	(0.6)	4.28	(0.6)	<0.01	*
Irish	10.05	(<0.1)	13.99	(0.7)	3.94	(0.8)	<0.01	*
German	13.96	(<0.1)	16.99	(0.7)	3.03	(0.7)	<0.01	*
Unclassified	1.95	(<0.1)	2.86	(0.3)	0.91	(0.3)	<0.01	*
Scottish	1.77	(<0.1)	2.59	(0.3)	0.83	(0.3)	<0.01	*
Dutch	1.26	(<0.1)	1.70	(0.2)	0.44	(0.2)	0.03	*
French Canadian	0.66	(<0.1)	0.37	(0.1)	-0.29	(0.1)	<0.01	*
British	0.48	(<0.1)	0.20	(0.1)	-0.28	(0.1)	<0.01	*
Haitian	0.33	(<0.1)	0.19	(<0.1)	-0.14	(0.1)	<0.01	*
Pennsylvania German	0.10	(<0.1)	0.03	(<0.1)	-0.08	(<0.1)	<0.01	*
Czechoslovakian	0.10	(<0.1)	0.04	(<0.1)	-0.06	(<0.1)	<0.01	*
Ethiopian	0.10	(<0.1)	0.04	(<0.1)	-0.05	(0.1)	<0.01	*
Kenyan	0.02	(<0.1)	0.01	(<0.1)	-0.02	(<0.1)	<0.01	*

Sources: U.S. Census Bureau, 2016 American Community Survey 1-year Estimates and 2016 American Community Survey Content Test, Control Treatment.

Note: Margins of error are shown in parentheses. Significant at  $\alpha=0.1$  level. P-values with an asterisk (\*) indicate a significant difference between the two rates. P-values have been adjusted for multiple comparisons using the Holm-Bonferroni method. Data for housing units in Alaska and Hawaii were excluded from the 2016 American Community Survey 1-year estimates in this table to make them comparable to the 2016 ACS Content Test data, which did not sample in these states. DRB Approval Number CBDRB-FY19-RAGLIN-B0005.

Overall in analyses from Tables 11 and 12, we compared 126 unique detailed groups, 104 of which had samples sizes large enough to compare. Of the 104 groups, 88 showed no statistical differences, while eight showed larger estimates from the expanded race/Hispanic origin questions of the 2016 ACS Content Test, and eight (including the “not reported or missing” category) showed higher estimates from the ancestry question of the ACS production.

Table 12 shows similar comparisons with a more equitable list of detailed groups from each major race/Hispanic origin category. In this second part of the analysis of Research Question 5, we compared aggregated major race/Hispanic origin category, along with six detailed groups from each major category. For the major race/Hispanic origin categories, aggregated estimates of White, Black or African American, AIAN, and SOR all showed more reporting when using the 2016 ACS Content Test race/Hispanic origin item, while Hispanic and NHPI all showed no change. No larger race/Hispanic origin groups had statistically less reporting when derived from the race/Hispanic origin items in the 2016 ACS Content Test versus ACS production.

Notably, however, some of the major race/Hispanic origin groups include checkbox responses (see Figure 2) which may or may not coincide with a detailed response.<sup>33</sup> Thus if respondents are simply marking a checkbox without giving further details, it could be that the ancestry survey item is more successful at gathering *detailed* responses. To test this, we compared estimates of detailed responses, both overall and within six of the larger race/Hispanic origin groups. A notable finding shown in the top row of Table 12 is that detailed reporting overall occurred more often on the race/Hispanic origin items than on the ancestry item (74.14 percent vs 70.80 percent). Additionally, there was more detailed reporting of White, Asian, and Hispanic groups in the race/Hispanic origin items.<sup>34</sup> No major race/Hispanic origin groups had more detailed reporting in the ancestry item.

We also examined six detailed groups within each major race/Hispanic origin group, for a total of 36 detailed groups.<sup>35</sup> Of the 36 detailed groups shown in Table 12, 32 had sample sizes large enough to be compared, 25 of which showed no statistical difference when derived from the race and Hispanic origin item from the 2016 ACS Content Test versus ACS production. For five groups, the estimate from the 2016 ACS Content Test was larger, while the estimate from ACS production was larger for two groups. Most of the detailed groups that showed differences were also shown in Table 11. Additional groups that showed differences were Asian Indian and Puerto Rican, which had higher estimates in the 2016 ACS Content Test.

Overall in analyses from Tables 11 and 12, we compared 126 unique detailed groups, 104 of which had samples sizes large enough to compare. Of the 104 groups, 88 showed no statistical differences, while eight showed larger estimates from the expanded race/Hispanic origin questions

---

<sup>33</sup> Each response was also categorized as “detailed” or “not detailed,” as a sizable number of people respond to the race/ethnicity questions without giving a detailed response (i.e. marking “White” checkbox only). Detailed responses from the American Indian/Alaskan Native group were not tabulated.

<sup>34</sup> One confounding finding is that detailed White reporting in the 2016 ACS Content Test was higher in the ancestry item, whereas when comparing race/Hispanic origin data from the 2016 ACS Content Test with Production ancestry it was *lower* in the ancestry item. This was because detailed ancestry reporting for White groups was higher in the 2016 ACS Content Test than it was in production. Although this phenomenon was not a focus of this report, this could be in part due to the added effects of asking detailed race for every major category before ancestry is asked, which is not done in production.

<sup>35</sup> Of the 36 groups analyzed, 15 are repeated from Table A-3, and 21 of which are unique to Table 12.

of the 2016 ACS Content Test, and eight (including the “not reported or missing” category) showed higher estimates from the ancestry question of the ACS production.

**Table 12. Comparing Estimates Derived from Race/Hispanic Origin and Ancestry, by Major Category or Detailed Group**

Major Category or Detailed Group	2016 ACS		2016 ACS Content		Difference		Adjusted P-Value
	Production Ancestry (N=4,880,000)		Test, Race/Hispanic Origin (N=43,500)				
Any Detailed	70.80	(0.1)	74.14	(0.8)	3.34	(0.8)	<0.01 *
White	45.57	(0.1)	74.02	(0.8)	28.42	(0.8)	<0.01 *
White, detailed	42.12	(0.1)	44.66	(1.0)	2.54	(1.0)	<0.01 *
German	13.96	(<0.1)	16.99	(0.7)	3.03	(0.7)	<0.01 *
Irish	10.05	(<0.1)	13.99	(0.7)	3.94	(0.7)	<0.01 *
English	7.47	(<0.1)	11.75	(0.6)	4.28	(0.6)	<0.01 *
Italian	5.26	(<0.1)	5.96	(0.4)	0.71	(0.5)	0.31
Polish	2.89	(<0.1)	3.51	(0.4)	0.62	(0.4)	0.12
French (except Basque)	2.48	(<0.1)	2.67	(0.2)	0.19	(0.2)	1.00
Black or African American	11.11	(<0.1)	12.75	(0.6)	1.64	(0.6)	<0.01 *
Black or African American, detailed	10.04	(<0.1)	9.21	(0.5)	-0.83	(0.5)	0.33
African American	8.01	(<0.1)	7.61	(0.5)	-0.41	(0.5)	1.00
Jamaican	0.35	(<0.1)	0.40	(0.1)	0.04	(0.1)	1.00
Haitian	0.33	(<0.1)	0.19	(<0.1)	-0.14	(<0.1)	<0.01 *
Nigerian	0.12	(<0.1)	0.17	(0.1)	0.05	(0.1)	1.00
Ethiopian	0.08	(<0.1)	0.03	(<0.1)	-0.05	(<0.1)	<0.01 *
Somali	0.05	(<0.1)	0.03	(<0.1)	-0.02	(<0.1)	1.00
Asian	5.56	(<0.1)	6.88	(0.6)	1.21	(0.5)	<0.01 *
Asian, detailed	5.36	(<0.1)	6.71	(0.5)	1.35	(0.5)	<0.01 *
Chinese	1.25	(<0.1)	1.64	(0.3)	0.40	(0.3)	0.38
Asian Indian	1.06	(<0.1)	1.55	(0.3)	0.50	(0.3)	0.06 *
Filipino	0.93	(<0.1)	1.08	(0.2)	0.15	(0.2)	1.00
Vietnamese	0.53	(<0.1)	0.69	(0.2)	0.15	(0.2)	1.00
Korean	0.49	(<0.1)	0.61	(0.1)	0.12	(0.1)	1.00
Japanese	0.32	(<0.1)	0.42	(0.1)	0.10	(0.1)	1.00
AIAN	2.86	(<0.1)	4.10	(0.4)	1.24	(0.4)	<0.01 *
NHPI	0.19	(<0.1)	0.44	(0.2)	0.25	(0.2)	0.90
NHPI, detailed	0.17	(<0.1)	0.37	(0.2)	0.27	(0.2)	0.60
Native Hawaiian	0.06	(<0.1)	0.09	(<0.1)	0.03	(<0.1)	1.00
Samoan	0.04	(<0.1)	0.18	(0.2)	0.14	(0.2)	1.00
Chamorro	0.01	(<0.1)	0.06	(<0.1)	0.05	(<0.1)	0.90
Tongan	0.01	(<0.1)	-		-		-
Fijian	0.01	(<0.1)	-		-		-
Marshallese	0.01	(<0.1)	-		-		-

(Table continued on the next page)



**Table 12. Comparing Estimates Derived from Race/Hispanic Origin and Ancestry, by Major Category or Detailed Group, continued**

Major Category or Detailed Group	2016 ACS	2016 ACS Content		Difference		Adjusted P-Value
	Production Ancestry (N=4,880,000)	Test, Race/Hispanic Origin (N=43,500)				
Some other race	1.27 (<0.1)	8.21 (0.6)	6.93 (0.6)		<0.01 *	
Some other race, detailed	1.27 (<0.1)	1.60 (0.2)	0.33 (0.2)		0.15	
Brazilian	0.13 (<0.1)	0.25 (0.1)	0.11 (0.1)		0.97	
Guyanese	0.08 (<0.1)	0.08 (<0.1)	0.00 (<0.1)		1.00	
Cabo Verdean	0.04 (<0.1)	0.03 (<0.1)	0.00 (<0.1)		1.00	
Belizean	0.02 (<0.1)	0.01 (<0.1)	-0.01 (<0.1)		1.00	
Creole	0.02 (<0.1)	-	-		-	
North American	0.01 (<0.1)	0.03 (<0.1)	0.02 (<0.1)		1.00	
Hispanic	16.11 (<0.1)	17.14 (0.8)	0.64 (0.8)		1.00	
Hispanic, detailed	14.29 (<0.1)	16.03 (0.8)	1.74 (0.8)		<0.01 *	
Mexican	9.35 (<0.1)	10.24 (0.6)	0.89 (0.6)		0.46	
Puerto Rican	1.33 (<0.1)	1.97 (0.3)	0.63 (0.3)		0.01 *	
Cuban	0.61 (<0.1)	0.73 (0.1)	0.11 (0.1)		1.00	
Salvadoran	0.66 (<0.1)	0.60 (0.1)	-0.06 (0.1)		1.00	
Dominican	0.57 (<0.1)	0.62 (0.2)	0.05 (0.2)		1.00	
Colombian	0.32 (<0.1)	0.36 (0.1)	0.04 (0.1)		1.00	

Sources: U. S. Census Bureau, 2016 American Community Survey 1-year Estimates and 2016 American Community Survey Content Test, Control Treatment.

Note: Margins of error are shown in parentheses. Significant at  $\alpha=0.1$  level. P-values with an asterisk (\*) indicate a significant difference between the two rates. P-values have been adjusted for multiple comparisons using the Holm-Bonferroni method. An '-' entry in the estimate column indicates that either no sample observations or too few sample observations were available to compute an estimate. Data for housing units in Alaska and Hawaii were excluded from the 2016 American Community Survey 1-year estimates in this table to make them comparable to the 2016 ACS Content Test data, which did not sample in these states. AIAN = American Indian and Alaskan Native, SOR = Some other race. DRB Approval Number CBDRB-FY19-RAGLIN-B0005.

Finally, for the third part of our analysis for Research Question 3, Table 13 compares group estimates aggregated by world region. Groups that could be specifically linked to a geographic region were included in the analysis. Responses such as “White,” “Black,” “Hispanic,” or others that could not conclusively be linked to a region were excluded from the analysis in Table 13. When aggregated this way, estimates of detailed groups from Western Europe; Mexico, Central America, and South America; the Caribbean; and East, Southeast, and South Asia were larger when derived from race/Hispanic origin data than from ACS production ancestry data. Estimates of groups from Eastern Europe and Central Asia, the Middle East and North Africa, Sub-Saharan Africa, and the Pacific Islands showed no statistical differences. Estimates of groups from North America (which essentially included Canadian, regional Canada, Greenland, and Cajun) were lower with race/Hispanic origin data. One potential explanation for this findings is people of Canadian descent reporting a more distant, perhaps European, origin in the ancestry item.

**Table 13. Comparing Estimates Derived from Race/Hispanic Origin and Ancestry, by World Region**

Ancestry or Detailed Race/Hispanic Origin Group	2016 ACS Production Ancestry (N=4,880,000)		2016 ACS Content Test, Race/HO (N=43,500)		Difference		Adjusted P-Value
	Estimate	Margin of Error	Estimate	Margin of Error	Estimate	Margin of Error	
Western Europe	36.43	(0.1)	40.40	(1.0)	3.97	(1.0)	<0.01 *
Eastern Europe and Central Asia	5.97	(<0.1)	6.49	(0.5)	0.53	(0.4)	0.16
Mexico, Central America, and South America	12.76	(<0.1)	13.67	(0.7)	0.90	(0.7)	0.15
Caribbean	3.36	(<0.1)	3.95	(0.4)	0.59	(0.4)	0.06 *
Middle East and North Africa	1.07	(<0.1)	1.31	(0.2)	0.24	(0.2)	0.16
Sub-Saharan Africa	1.09	(<0.1)	0.85	(0.2)	-0.24	(0.2)	0.16
East, Southeast, and South Asia	5.34	(<0.1)	6.69	(0.5)	1.35	(0.5)	<0.01 *
Pacific Islands	0.23	(<0.1)	0.50	(0.2)	0.26	(0.2)	0.15
North America (except the U.S., Mexico, and Caribbean)	0.91	(<0.1)	0.61	(0.1)	-0.30	(0.1)	<0.01 *

Sources: U. S. Census Bureau, 2016 American Community Survey 1-year Estimates and 2016 American Community Survey Content Test, Control Treatment.

Note: Margins of error are shown in parentheses. Significant at  $\alpha=0.1$  level. P-values with an asterisk (\*) indicate a significant difference between the two rates. P-values have been adjusted for multiple comparisons using the Holm-Bonferroni method. Data for housing units in Alaska and Hawaii were excluded from the 2016 American Community Survey 1-year estimates in this table to make them comparable to the 2016 ACS Content Test data, which did not sample in these states. DRB Approval Number CBDRB-FY19-RAGLIN-B0005.

## 6. CONCLUSIONS

This research provides a detailed view into the relationship between data collected from the race, Hispanic origin, and ancestry survey items. We compared race and Hispanic origin data to ancestry data for individual 2016 ACS Content Test respondents and also compared the resulting estimates from race and Hispanic data to the estimates obtained from ancestry data.

Within an individual person's ACS data, the information reported in the ancestry item is often the same as information reported in the race and Hispanic origin items. With the expanded format of the race and Hispanic origin items, 72.5 percent of respondents provided the exact same information in ancestry as was in race/Hispanic origin (when responses were provided to both questions). An additional 7.1 percent of people provided two ancestries but only one matched exactly to information in race/Hispanic origin. The remaining 20.4 percent of people had ancestry information that did not exactly match to their race/Hispanic origin information. These calculations omit cases where a valid response was not recorded for both the race/Hispanic origin and ancestry items. In the 2016 ACS Content Test control treatment, 12.6 percent of people had a valid race/Hispanic origin response but left ancestry entirely blank. The prevalence of missing data from the ancestry question plus the Census Bureau's policy to retain a maximum of two codes from a person's ancestry write-in are important limitations to keep in mind when comparing ancestry responses to race/Hispanic origin responses.

Of all the ancestries collected, 25 percent were not an exact match to information provided in race/Hispanic origin. The likelihood of an ancestry response matching exactly to information in the race/Hispanic origin response for a person was found to vary based on the major race or Hispanic origin category of the ancestry response. This analysis sought to categorize the relationship between non-matching ancestry responses and the reported race and Hispanic origin responses for a person. Of the ancestries that did not exactly match a race/Hispanic origin report, the majority were in the same major category as the race or Hispanic origin (for example, a response of Haitian in race and response of Caribbean in ancestry). It appeared that a non-matching ancestry response classified as White or Black tended to be more specific than what was in the race and Hispanic origin series. However, non-matching ancestries classified as other major race/Hispanic groups tended to provide the same level of specificity in ancestry as in race.

When looking for how often and where a specific detailed group (such as Korean) was reported, we saw that some groups were more often collected in the race and Hispanic origin items rather than in the ancestry item, and no group that we analyzed appeared more often in ancestry than in race/Hispanic origin. Six of the 11 detailed groups that were reported more often in race/Hispanic origin also had checkbox response options in the race/Hispanic origin items (Chinese, Filipino, Asian Indian, Vietnamese, Mexican, and Puerto Rican), which may have facilitated higher reporting rates for those groups. A sizable number of people who reported their race as White or Black report a detailed response in the ancestry item while only marking the checkbox for White or Black in the race item. Collecting detailed information will likely continue to pose a challenge for some groups, particularly for some White and Black respondents, who are perhaps more accustomed to marking a single generic checkbox in the race item, or who do not identify with a more specific origin.

When comparing estimates from the expanded race and Hispanic origin questions on the 2016 ACS Content Test to ACS production ancestry estimates, the estimates tended not to differ at the national level. Out of 104 groups that we compared, 16 showed statistical differences, with eight estimates being larger when derived from 2016 ACS Content Test race data and eight estimates being smaller when derived from 2016 ACS Content Test race data.<sup>36</sup> Overall, more detailed reporting occurred on the race and Hispanic origin items of the 2016 ACS Content Test compared to the ancestry item from production. Detailed reporting in the White, Asian, and Hispanic categories was higher in the 2016 ACS Content Test race and Hispanic origin items as well. Detailed reporting in the Black, NHPI, and SOR categories of the 2016 ACS Content Test did not

---

<sup>36</sup> We compared 105 detailed groups in Table 12 and Table A-3.

differ from production ancestry data.<sup>37</sup> Part of this may be attributed to the fact that the 2016 ACS Content Test race and Hispanic origin items allow for more comprehensive self-identification.<sup>38</sup>

When deciding about the future of the ancestry survey item on the ACS, decision makers must take into account several competing considerations, including the relative burden that the ancestry question places on the public, the relatively high nonresponse rate, the utility and necessity of the data collected, as well as the views of affected groups and community stakeholders. While this research cannot speak to all these issues, it indicates that there is some level of redundancy in the ancestry data when compared with data from the revised race and Hispanic origin questions. It also shows that the race/Hispanic origin items collectively receive a higher rate of response, and are on average more successful at eliciting detailed responses than the ancestry item. These results should be considered as the American Community Survey program strives to ensure that it is minimizing respondent burden while maintaining high data quality standards.

---

<sup>37</sup> Even with the tendency for some respondents to mark the White checkbox and only provide detailed information on the ancestry item, when compared with the ancestry item in ACS production there was overall more detailed White reporting to the race/Hispanic origin items from the 2016 ACS Content Test.

<sup>38</sup> Each write-in line from the race and Hispanic origin questions could potentially have generated 10 distinct responses while the write-in from ancestry only generated a maximum of two. To make better comparisons, we ignored race write-in codes 3 through 10 and re-assessed the prevalence of each group. For Dutch, English, Irish, German and Scottish, the new 2016 ACS Content Test rate of Dutch was the only one found to no longer be significantly different from the production ancestry rate, when compared using t-tests at the  $\alpha = 0.10$  level without multiple comparisons.

## 7. REFERENCES

- Alba, Richard, and Tariqul Islam (2009). "The Case of the Disappearing Mexican Americans: An Ethnic-Identity Mystery." *Population Research and Policy Review*, Vol. 28, No. 2: 109-121.
- Alberti, Nicholas (2006). "2005 National Census Test: Analysis of the Race and Ethnicity Questions." Decennial Statistical Studies Division, U.S. Census Bureau, October 30, 2006.
- Brittingham, Angela and G. Patricia de la Cruz (2004). "Ancestry: 2000." U.S. Census Bureau, Census 2000 Brief, C2KBR-35, Washington, DC. Retrieved August 3, 2018 from <https://www.census.gov/history/pdf/ancestry.pdf>.
- Childs, Jennifer Hunter, Rodney Terry, Nathan Jurgenson, Matthew Clifton, and George Higbie. (2010). "Iterative Cognitive Testing of the 2010 Race and Hispanic Origin Alternative Questionnaire Experiment (AQE) Reinterview. U.S. Census Bureau, Survey Methodology #2010-13. Retrieved November 1, 2017 from: <https://www.census.gov/srd/papers/pdf/ssm2010-13.pdf>.
- Compton, Elizabeth, Michael Bentley, Sharon Ennis, and Sonya Rastogi. (2012). "2010 Race and Hispanic Origin Alternative Questionnaire Experiment." Decennial Statistical Studies Division and Population Division, U.S. Census Bureau.
- Duncan, Brian and Stephen Trejo. (2008). "Ancestry versus Ethnicity: The Complexity and Selectivity of Mexican Identification in the United States." in Amelie F. Constant, Konstantinos Tatsiramos, Klaus F. Zimmermann (ed.) *Ethnicity and Labor Market Outcomes (Research in Labor Economics, Volume 29)*, Emerald Group Publishing Limited: 31-66.
- Farley, Reynolds. (1991). "The New Census Question about Ancestry: What Did It Tell Us?" *Demography*, Vol. 28, No. 3: 411-429.
- Fernandez, Leticia, Eleanor Gerber, Matthew Clifton, George Higbie, and Mikelyn Meyers. (2009). "Cognitive Pretesting of 2010 Alternative Questionnaire Experiment (AQE) Race and Hispanic Origin Treatment Panels." U.S. Census Bureau, Survey Methodology #2009-08, Washington, D.C. Retrieved November 1, 2017 from: <https://www.census.gov/srd/papers/pdf/rsm2009-08.pdf>.
- Harth, Jacquelyn, Angela Buchanan, Derek Breese, Merarys Rios, Hyon B. Shin, Sarah K. Heimel, and Lindsay Longsine. (2017). "2016 American Community Survey Content Test Evaluation Report: Race and Hispanic Origin." U.S. Census Bureau. Retrieved from:

[https://www.census.gov/content/dam/Census/library/working-papers/2017/acs/2017\\_Harth\\_01.pdf](https://www.census.gov/content/dam/Census/library/working-papers/2017/acs/2017_Harth_01.pdf).

Heimel, Sarah. (2014). "American Community Survey Fiscal Year 2014 Content Review Median County-Level Allocation Rates of ACS Data." U.S. Census Bureau, October 28, 2014. Retrieved from: [https://www2.census.gov/programs-surveys/acs/operations\\_admin/2014\\_content\\_review/methods\\_results\\_report/Allocation\\_Rates\\_Report.pdf](https://www2.census.gov/programs-surveys/acs/operations_admin/2014_content_review/methods_results_report/Allocation_Rates_Report.pdf).

Hobbs, Frank, and Angela Brittingham. (2007). "Measurement and Evaluation of Ancestry: Differences Between Specific and Generalized Responses." Presented at the 2007 American Statistical Association Conference.

Holm, Sture. (1979). "A Simple Sequentially Rejective Multiple Test Procedure." *Scandinavian Journal of Statistics*, Vol. 6, No. 2: 65-70.

Horwitz, Rachel, Jennifer Guarino Tancreto, Mary Frances Zelenak, and Mary Davis. (2013). "Using Paradata to Identify Potential Issues and Trends in the American Community Survey Internet Instrument." U.S. Census Bureau. Retrieved from: [https://www.census.gov/content/dam/Census/library/working-papers/2013/acs/2013\\_Horwitz\\_02.pdf](https://www.census.gov/content/dam/Census/library/working-papers/2013/acs/2013_Horwitz_02.pdf).

Lieberson, Stanley and Mary C. Waters. (1988). *From Many Strands: Ethnic and Racial Groups in Contemporary America*. New York: Russell Sage Foundation.

Mathews, Kelly, Jessica Phelan, Nicholas Jones, Sarah Konya, Rachel Marks, Beverly Pratt, and Michael Bentley. (2017). "2020 Research and Testing: 2015 National Content Test Race and Ethnicity Analysis Report." U.S. Census Bureau.

McKay, Ruth and Manuel de la Puente. (1996). "Cognitive Research in Designing the CPS Supplement on Race and Ethnicity." U.S. Census Bureau. 1995 Annual Research Conference Proceedings.

McKenney, Nampeo R. and Arthur R. Cresce (1990). "The Identification of Ethnicity in the United States: The Census Bureau Experience." Presented at the Population Association of America meeting in Toronto, Canada.

Pratt, Beverly, Lindsay Hixson, and Nicholas Jones. (2015). "Measuring Race and Ethnicity Across the Decades: 1790-2010." U.S. Census Bureau. Retrieved September 27, 2017 from: [https://www.census.gov/data-tools/demo/race/MREAD\\_1790\\_2010.html](https://www.census.gov/data-tools/demo/race/MREAD_1790_2010.html).

Schwede, Laurie, Theresa F. Leslie, and Deborah H. Griffin (2002). "Interviewers' Reported Behaviors in Collecting Race and Hispanic Origin Data." U.S. Census Bureau. Retrieved September 26, 2017 from: [https://www.census.gov/content/dam/Census/library/working-papers/2002/acs/2002\\_Schwede\\_01.pdf](https://www.census.gov/content/dam/Census/library/working-papers/2002/acs/2002_Schwede_01.pdf).

U.S. Census Bureau. (2002). "Measuring America: The Decennial Censuses from 1790 to 2000." by Jason G. Gauthier, POL/02-MA(RV), Washington, DC.

U.S. Census Bureau. (2007). "First, Second, and Total Responses to the Ancestry Question by Detailed Code: 2000." Census 2000 Report Number PHC-T-43. Retrieved August 7, 2017 from: <https://www.census.gov/data/tables/2000/dec/phc-t-43.html>.

U.S. Census Bureau. (2014). American Community Survey Design and Methodology (January 2014). Retrieved August 3, 2018 from: <https://www.census.gov/programs-surveys/acs/methodology/design-and-methodology.html>.

U.S. Census Bureau. (2018). Questions Planned for the 2020 Census and American Community Survey. Retrieved February 27, 2019 from: <https://www2.census.gov/library/publications/decennial/2020/operations/planned-questions-2020-acs.pdf>.

U.S. Office of Management and Budget. (1997). Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity. Retrieved August 3, 2018 from: <https://www.whitehouse.gov/wp-content/uploads/2017/11/Revisions-to-the-Standards-for-the-Classification-of-Federal-Data-on-Race-and-Ethnicity-October30-1997.pdf>.

## Appendix A. Mail Materials from the Test Treatment of the 2016 ACS Content Test

This appendix presents the test version of the race and Hispanic origin questions from the 2016 ACS Content Test, as they appeared on the paper questionnaire mailed to sampled households. Section 2.1 showed the control version, which is similar to the format that will be used on future questionnaires. The test treatment:

- Asked one question that encompassed Hispanic ethnicity and the major race groups.
- Changed the question stem to “Which categories describe person [x]?” to avoid using the word “race.”
- Provided high-level checkboxes for each OMB race group.
- Added a distinct category for Middle Eastern or North African (MENA) respondents.
- Provided examples next to each top-level identity.
- Provided write-in spaces for all groups.
- Removed detailed checkboxes for Asian and NHPI.
- Changed the “Some other race” category wording to include the words “ethnicity or origin.”

**Figure A-1. Mail Version of the Race and Ethnicity Question from the Test Treatment of the 2016 ACS Content Test**

**5** Which categories describe Person 1? Mark all boxes that apply **AND** print details in the spaces below.

**White** — Print details, for example, German, Irish, English, etc. ↴

**Hispanic, Latino, or Spanish origin** — Print details, for example, Mexican or Mexican American, Puerto Rican, Colombian, etc. ↴

**Black or African Am.** — Print details, for example, African American, Jamaican, Nigerian, etc. ↴

**Asian** — Print details, for example, Chinese, Asian Indian, Vietnamese, etc. ↴

**American Indian or Alaska Native** — Print name(s) of enrolled or principal tribe(s), for example, Navajo Nation, Mayan, Native Village of Barrow Inupiat Traditional Government, etc. ↴

**Middle Eastern or North African** — Print details, for example, Lebanese, Iranian, Egyptian, etc. ↴

**Native Hawaiian or Other Pacific Islander** — Print details, for example, Native Hawaiian, Chamorro, Fijian, etc. ↴

**Some other race, ethnicity, or origin** — Print details. ↴

Source: U.S. Census Bureau, 2016 ACS Content Test



## Appendix B. Internet Materials from the 2016 ACS Content Test

This appendix presents the internet versions of the 2016 ACS Content Test questionnaires, both control and test treatments. The control version is presented first in Figures A-2 and A-3, followed by the test version.

For the control treatment, the internet instrument contained identical checkbox and write-in categories as was presented in the paper race and Hispanic origin questions.

**Figure A-2. Internet Version of the Control Treatment Hispanic Origin Question**

**5** Is T L M M M M of Hispanic, Latino, or Spanish origin? Select all boxes that apply **AND** enter origins. Note, you may report more than one group. (Help)

- No, not of Hispanic, Latino, or Spanish origin
- Yes, Mexican, Mexican Am., Chicano
- Yes, Puerto Rican
- Yes, Cuban
- Yes, another Hispanic, Latino, or Spanish origin - Enter details, for example, Salvadoran, Dominican, Colombian, Guatemalan, Spaniard, Ecuadorian, etc.

**Figure A-3. Internet Version of the Control Treatment Race Question**

**6** What is 's race? Select all boxes that apply **AND** enter origins in the spaces below. Note, you may report more than one group. (Help)

- White - Enter details, for example, German, Irish, English, Italian, Lebanese, Egyptian, etc.
- Black or African Am. - Enter details, for example, African American, Jamaican, Haitian, Nigerian, Ethiopian, Somali, etc.
- American Indian or Alaska Native - Enter details, for example, Navajo Nation, Blackfeet Tribe, Mayan, Aztec, Native Village of Barrow Inupiat Traditional Government, Nome Eskimo Community, etc.
- Chinese
- Filipino
- Asian Indian
- Vietnamese
- Korean
- Japanese
- Other Asian - Enter details, for example, Pakistani, Cambodian, Hmong, etc.
- Native Hawaiian
- Samoan
- Chamorro
- Other Pacific Islander - Enter details, for example, Tongan, Fijian, Marshallese, etc.
- Some other race - Enter race or origin.


For the test treatment, the internet instrument contained the same checkboxes as in the paper version, but subsequent screens then presented detailed categories plus a write-in for six of the eight top-level race/ethnicity options. Figure A-9 shows the screen presented to persons who identified as AIAN, offering three write-ins associated with different AIAN identifications, and Figure A-12 shows the screen presented to persons who identified as Some Other Race, which offered a single write-in to collect more information.

**Figure A-4. Image of the Internet Version of the Test Treatment Combined Race and Hispanic Origin Question**

5 Which categories describe rrr? Select all boxes that apply. Note, you may report more than one group. [\(Help\)](#)


- White** - For example, German, Irish, English, Italian, Polish, French, etc.
- Hispanic, Latino, or Spanish origin** - For example, Mexican or Mexican American, Puerto Rican, Cuban, Salvadoran, Dominican, Colombian, etc.
- Black or African Am.** - For example, African American, Jamaican, Haitian, Nigerian, Ethiopian, Somali, etc.
- Asian** - For example, Chinese, Filipino, Asian Indian, Vietnamese, Korean, Japanese, etc.
- American Indian or Alaska Native** - For example, Navajo Nation, Blackfeet Tribe, Mayan, Aztec, Native Village of Barrow Inupiat Traditional Government, Nome Eskimo Community, etc.
- Middle Eastern or North African** - For example, Lebanese, Iranian, Egyptian, Syrian, Moroccan, Algerian, etc.
- Native Hawaiian or Other Pacific Islander** - For example, Native Hawaiian, Samoan, Chamorro, Tongan, Fijian, Marshallese, etc.
- Some other race, ethnicity, or origin**

**Figure A-5. White Detailed Question for the Test Treatment Race and Hispanic Origin Question**

 Select all boxes that apply and/or enter details in the space below.  
**WHITE** ([Help](#))


- German
- Irish
- English
- Italian
- Polish
- French
- Enter, for example, Scottish, Norwegian, Dutch, etc.

**Figure A-6. Hispanic Detailed Question for the Test Treatment Race and Hispanic Origin Question**

 Select all boxes that apply and/or enter details in the space below.  
**HISPANIC, LATINO, OR SPANISH** ([Help](#))


- Mexican or Mexican American
- Puerto Rican
- Cuban
- Salvadoran
- Dominican
- Colombian
- Enter, for example, Guatemalan, Spaniard, Ecuadorian, etc.

Figure A-7. Black Detailed Question for the Test Treatment Race and Hispanic Origin Question

 Select all boxes that apply and/or enter details in the space below.  
**BLACK OR AFRICAN AMERICAN** ([Help](#))

- African American
- Jamaican
- Haitian
- Nigerian
- Ethiopian
- Somali
- Enter, for example, Ghanaian, South African, Barbadian, etc.

Figure A-8. Asian Detailed Question for the Test Treatment Race and Hispanic Origin Question

 Select all boxes that apply and/or enter details in the space below.  
**ASIAN** ([Help](#))

- Chinese
- Filipino
- Asian Indian
- Vietnamese
- Korean
- Japanese
- Enter, for example, Pakistani, Cambodian, Hmong, etc.

Figure A-9. AIAN Detailed Question for the Test Treatment Race and Hispanic Origin Question


 Enter details in the space(s) below. ([Help](#))

**AMERICAN INDIAN**  
*Enter, for example, Navajo Nation, Blackfeet Tribe, Muscogee (Creek) Nation, etc.*

**ALASKA NATIVE**  
*Enter, for example, Native Village of Barrow Inupiat Traditional Government, Nome Eskimo Community, Orutsarmuit Native Village, etc.*


**CENTRAL OR SOUTH AMERICAN INDIAN**  
*Enter, for example, Mayan, Aztec, Taino, etc.*

Figure A-10. MENA Detailed Question for the Test Treatment Race and Hispanic Origin Question

 Select all boxes that apply and/or enter details in the space below.  
MIDDLE EASTERN OR NORTH AFRICAN [\(Help\)](#)


- Lebanese
- Iranian
- Egyptian
- Syrian
- Moroccan
- Algerian
- Enter, for example, Israeli, Iraqi, Tunisian, etc.

Figure A-11. NHPI Detailed Question for the Test Treatment Race and Hispanic Origin Question

 Select all boxes that apply and/or enter details in the space below.  
NATIVE HAWAIIAN OR OTHER PACIFIC ISLANDER [\(Help\)](#)

- Native Hawaiian
- Samoan
- Chamorro
- Tongan
- Fijian
- Marshallese
- Enter, for example, Palauan, Tahitian, Chuukese, etc.

Figure A-12. SOR Detailed Question for the Test Treatment Race and Hispanic Origin Question

 Enter other details about 's race, ethnicity, or origin. [\(Help\)](#)

Enter details

## Appendix C. CATI and CAPI Questionnaire Materials

This appendix presents the versions of the 2016 ACS Content Test questionnaires that were used in both the CATI and CAPI operations. The left side of Table A-1 presents Control version questions, while the right side presents the Test version questions.

**Table A-1. CATI and CAPI Versions of the Race and Ethnicity Questions**

Control Version	Test Version
<p><b>[HISPANIC]</b>  <b>Are you of Hispanic, Latino, or Spanish origin?</b></p> <p>_Yes – IF YES, ASK HISPANIC DETAIL            _No</p> <p><b>[HISPANIC DETAIL]</b>  <b>Are you Mexican, Mexican American, Chicano; Puerto Rican; Cuban, or of some other Hispanic, Latino, or Spanish origin?</b></p> <p>_Mexican, Mexican American, or Chicano            _Puerto Rican            _Cuban            _Another Hispanic, Latino, or Spanish origin– IF THIS CATEGORY IS SELECTED, ASK OTHER HISPANIC DETAIL</p> <p><b>[OTHER HISPANIC DETAIL]</b>  <b>What is that origin or origins?</b>  <b>(For example, Salvadoran, Dominican, Colombian, Guatemalan, Spaniard, Ecuadorian, etc.)</b></p> <hr/> <p>(INTERVIEWER TYPES IN ORIGIN OR ORIGINS)</p> <p><b>[RACE]</b>  <b>I'm going to read a list of races. You may choose one or more races. For this survey, Hispanic origin is not a race.</b></p> <p><b>What is your race?</b>  <b>Are you White; Black or African American; American Indian or Alaska Native; Asian; Native Hawaiian or Other Pacific Islander; or Some other race?</b></p> <p>_White            _Black or African American</p>	<p><b>[RACE/ETHNICITY]</b>  <b>I'm going to read a list of categories. You may choose all that apply.</b></p> <p><b>Are you White; Hispanic, Latino, or Spanish origin; Black or African American; Asian; American Indian or Alaska Native; Middle Eastern or North African; Native Hawaiian or Other Pacific Islander; or Some other race, ethnicity, or origin?</b></p> <p>_White            _Hispanic, Latino, or Spanish Origin            _Black or African American            _Asian            _American Indian or Alaska Native            _Middle Eastern or North African            _Native Hawaiian or Other Pacific Islander            _Some other race, ethnicity, or origin</p> <p><b>You said that you are (NAME THE GROUP OR GROUPS).</b></p> <p><b>Now I'm going to collect detailed information. You may give more than one response.</b></p>

Control Version	Test Version
<p data-bbox="203 233 714 363">           _American Indian or Alaska Native            _Asian            _Native Hawaiian or Other Pacific Islander            _Some other race         </p> <p data-bbox="203 415 738 447"><b>IF WHITE WAS SELECTED ASK WHITE DETAIL</b></p> <p data-bbox="203 525 397 556"><b>[WHITE DETAIL]</b></p> <p data-bbox="203 577 812 688"> <b>What are your WHITE origin or origins? For example, German, Irish, English, Italian, Lebanese, Egyptian, etc.</b> </p> <hr data-bbox="203 766 787 770"/> <p data-bbox="203 779 706 846">(INTERVIEWER TYPES IN ORIGIN OR ORIGINS)</p> <p data-bbox="203 919 795 987"><b>IF BLACK OR AFRICAN AMERICAN WAS SELECTED ASK BLACK DETAIL</b></p> <p data-bbox="203 1064 397 1096"><b>[BLACK DETAIL]</b></p> <p data-bbox="203 1117 820 1228"> <b>What are your BLACK OR AFRICAN AMERICAN origin or origins? For example, African American, Jamaican, Haitian, Nigerian, Ethiopian, Somali, etc.</b> </p> <hr data-bbox="203 1312 787 1316"/> <p data-bbox="203 1323 706 1390">(INTERVIEWER TYPES IN ORIGIN OR ORIGINS)</p>	<p data-bbox="852 415 1396 447"><b>IF WHITE WAS SELECTED, ASK WHITE DETAIL</b></p> <p data-bbox="852 525 1047 556"><b>[WHITE DETAIL]</b></p> <p data-bbox="852 577 1453 688"> <b>What are your specific categories for WHITE? For example, German, Irish, English, Italian, Polish, French, etc.</b> </p> <hr data-bbox="852 745 1437 749"/> <p data-bbox="852 756 1356 823">(INTERVIEWER TYPES IN ORIGIN OR ORIGINS)</p> <p data-bbox="852 896 1469 928"><b>IF HISPANIC WAS SELECTED, ASK HISPANIC DETAIL</b></p> <p data-bbox="852 1058 1079 1089"><b>[HISPANIC DETAIL]</b></p> <p data-bbox="852 1110 1477 1260"> <b>What are your specific categories for HISPANIC, LATINO, OR SPANISH origin? For example, Mexican or Mexican American, Puerto Rican, Cuban, Salvadoran, Dominican, Colombian, etc.</b> </p> <hr data-bbox="852 1323 1437 1327"/> <p data-bbox="852 1333 1356 1400">(INTERVIEWER TYPES IN ORIGIN OR ORIGINS)</p> <p data-bbox="852 1474 1453 1541"><b>IF BLACK OR AFRICAN AMERICAN WAS SELECTED, ASK BLACK DETAIL</b></p> <p data-bbox="852 1619 1047 1650"><b>[BLACK DETAIL]</b></p> <p data-bbox="852 1659 1453 1795"> <b>What are your specific categories for BLACK OR AFRICAN AMERICAN? For example, African American, Jamaican, Haitian, Nigerian, Ethiopian, Somali, etc.</b> </p> <hr data-bbox="852 1858 1437 1862"/> <p data-bbox="852 1869 1356 1900">(INTERVIEWER TYPES IN ORIGIN OR ORIGINS)</p>

<p><b>IF AMERICAN INDIAN OR ALASKA NATIVE WAS SELECTED ASK AIAN DETAIL</b></p> <p><b>[AIAN DETAIL]</b></p> <p><b>What are your AMERICAN INDIAN OR ALASKA NATIVE enrolled or principal tribe or tribes? For example, Navajo Nation, Blackfeet Tribe, Mayan, Aztec, Native Village of Barrow Inupiat Traditional Government, Nome Eskimo Community, etc.</b></p> <hr/> <p>(INTERVIEWER TYPES IN TRIBE OR TRIBES)</p> <p><b>IF ASIAN WAS SELECTED ASK ASIAN DETAIL</b></p> <p><b>[ASIAN DETAIL]</b></p> <p><b>You may choose one or more Asian groups.</b></p> <p><b>Are you Chinese, Filipino, Asian Indian, Vietnamese, Korean, Japanese, or of some other Asian origin?</b></p> <p><input type="checkbox"/> Chinese  <input type="checkbox"/> Filipino  <input type="checkbox"/> Asian Indian  <input type="checkbox"/> Vietnamese  <input type="checkbox"/> Korean  <input type="checkbox"/> Japanese  <input type="checkbox"/> Other Asian – IF SELECTED ASK DETAILED ASIAN</p> <p><b>[DETAILED ASIAN]</b></p> <p><b>What is that other Asian origin or origins? (For example, Pakistani, Cambodian, Hmong, etc.)</b></p> <hr/> <p>(INTERVIEWER TYPES IN ORIGIN OR ORIGINS)</p>	<p>ORIGINS)</p> <p><b>IF ASIAN WAS SELECTED, ASK ASIAN DETAIL</b></p> <p><b>[ASIAN DETAIL]</b></p> <p><b>What are your specific categories for ASIAN? For example, Chinese, Filipino, Asian Indian, Vietnamese, Korean, Japanese, etc.</b></p> <hr/> <p>(INTERVIEWER TYPES IN ORIGIN OR ORIGINS)</p> <p><b>IF AMERICAN INDIAN OR ALASKA NATIVE WAS SELECTED, ASK AIAN DETAIL</b></p> <p><b>[AIAN DETAIL]</b></p> <p><b>What are your specific categories for AMERICAN INDIAN OR ALASKA NATIVE? For example, Navajo Nation, Blackfeet Tribe, Mayan, Aztec, Native Village of Barrow Inupiat Traditional Government, Nome Eskimo Community, etc.</b></p> <hr/> <p>(INTERVIEWER TYPES IN TRIBES, VILLAGES, ETC.)</p> <p><b>IF MIDDLE EASTERN OR NORTH AFRICAN WAS SELECTED, ASK MENA DETAIL</b></p> <p><b>[MENA DETAIL]</b></p> <p><b>What are your specific categories for MIDDLE EASTERN OR NORTH AFRICAN? For example,</b></p>
-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------



Control Version	Test Version
<p><b>IF NATIVE HAWAIIAN OR PACIFIC ISLANDER WAS SELECTED ASK NHPI DETAIL:</b></p> <p><b>[NHPI DETAIL]</b>  <b>You may choose one or more Pacific Islander groups.</b></p> <p><b>Are you Native Hawaiian, Samoan, Chamorro, or of some other Pacific Islander origin?</b></p> <p>_Native Hawaiian  _Samoan  _Chamorro  _Other Pacific Islander – IF SELECTED ASK DETAILED PACIFIC ISLANDER</p> <p><b>[DETAILED PACIFIC ISLANDER]</b>  <b>What is that other Pacific Islander origin or origins?</b>  <b>(For example, Tongan, Fijian, Marshallese, etc.)</b></p> <hr/> <p>(INTERVIEWER TYPES IN ORIGIN OR ORIGINS)</p> <p><b>IF “SOME OTHER RACE OR ORIGIN” WAS SELECTED ASK SOR DETAIL</b></p> <p><b>[SOR DETAIL]</b>  <b>What is your other race group or groups?</b></p> <hr/> <p>(INTERVIEWER TYPES IN ORIGIN OR ORIGINS)</p>	<p><b>Lebanese, Iranian, Egyptian, Syrian, Moroccan, Algerian, etc.</b></p> <hr/> <p>(INTERVIEWER TYPES IN ORIGIN OR ORIGINS)</p> <p><b>IF NATIVE HAWAIIAN OR OTHER PACIFIC ISLANDER WAS SELECTED, ASK NHPI DETAIL</b></p> <p><b>[NHPI DETAIL]</b></p> <p><b>What are your specific categories for NATIVE HAWAIIAN OR OTHER PACIFIC ISLANDER? For example, Native Hawaiian, Samoan, Chamorro, Tongan, Fijian, Marshallese, etc.</b></p> <hr/> <p>(INTERVIEWER TYPES IN ORIGIN OR ORIGINS)</p> <p><b>IF SOME OTHER RACE WAS SELECTED, ASK SOR DETAIL</b></p> <p><b>[SOR DETAIL]</b>  <b>What are your specific categories for SOME OTHER RACE, ETHNICITY, OR ORIGIN?</b></p> <hr/> <p>(INTERVIEWER TYPES IN ORIGIN OR ORIGINS)</p>

## Appendix D. Table B04006: People Reporting Ancestry

The following table is published on American FactFinder, and shows official estimates using 2016 ACS 1-year data. These estimates include persons in Hawaii, Alaska and Group Quarters, etc.

**Table A-2. Reproduction of AFF Table B04006, People Reporting Ancestry**

<b>Ancestry</b>	<b>Estimate</b>	<b>Margin of Error</b>
Total:	<b>323,127,515</b>	<b>*****</b>
Afghan	123,947	(13,039)
Albanian	199,867	(16,394)
Alsatian	8,343	(1,505)
American	20,151,829	(112,100)
Arab:	2,032,892	(44,267)
Egyptian	256,070	(15,924)
Iraqi	148,993	(14,242)
Jordanian	78,108	(7,885)
Lebanese	489,396	(16,508)
Moroccan	119,461	(10,460)
Palestinian	142,308	(13,510)
Syrian	187,331	(11,608)
Arab	279,332	(15,329)
Other Arab	387,178	(21,839)
Armenian	467,890	(19,073)
Assyrian/Chaldean/Syriac	114,508	(10,002)
Australian	108,822	(7,943)
Austrian	703,315	(16,996)
Basque	65,458	(6,185)
Belgian	366,963	(10,994)
Brazilian	426,809	(19,790)
British	1,539,393	(24,447)
Bulgarian	106,084	(8,928)
Cajun	109,867	(7,041)
Canadian	665,472	(16,057)
Carpatho Rusyn	7,938	(1,988)
Celtic	46,856	(4,482)
Croatian	402,918	(11,285)

(Table continued on the next page)

<b>Ancestry</b>	<b>Estimate</b>	<b>Margin of Error</b>
Cypriot	8,661	(2,255)
Czech	1,412,051	(22,079)
Czechoslovakian	307,571	(8,361)
Danish	1,295,169	(22,160)
Dutch	4,044,507	(37,864)
Eastern European	734,203	(16,861)
English	23,835,787	(94,377)
Estonian	26,864	(2,939)
European	4,534,075	(63,139)
Finnish	665,597	(17,024)
French (except Basque)	7,962,052	(73,287)
French Canadian	2,120,016	(32,598)
German	44,754,050	(105,499)
German Russian	27,247	(3,706)
Greek	1,278,174	(27,658)
Guyanese	243,498	(13,713)
Hungarian	1,424,423	(25,031)
Icelander	54,877	(4,782)
Iranian	476,171	(18,250)
Irish	32,304,175	(112,639)
Israeli	146,570	(12,095)
Italian	16,896,518	(96,796)
Latvian	92,343	(5,372)
Lithuanian	635,409	(14,718)
Luxembourger	44,257	(3,026)
Macedonian	59,774	(6,184)
Maltese	40,820	(3,822)
New Zealander	21,661	(3,148)
Northern European	356,892	(13,291)
Norwegian	4,421,962	(36,981)
Pennsylvania German	322,836	(14,179)
Polish	9,258,128	(64,443)
Portuguese	1,375,288	(27,263)
Romanian	476,307	(17,072)

(Table continued on the next page)

<b>Ancestry</b>	<b>Estimate</b>	<b>Margin of Error</b>
Russian	2,754,205	(35,789)
Scandinavian	773,260	(21,279)
Scotch-Irish	3,212,692	(39,072)
Scottish	5,658,914	(53,093)
Serbian	181,607	(9,517)
Slavic	124,505	(7,109)
Slovak	687,243	(15,043)
Slovene	177,834	(8,179)
Soviet Union	2,552	(1,013)
Subsaharan African:	3,557,902	(64,600)
Cape Verdean	113,508	(9,551)
Ethiopian	305,809	(16,522)
Ghanaian	133,125	(11,730)
Kenyan	80,141	(9,014)
Liberian	80,253	(11,294)
Nigerian	380,785	(19,537)
Senegalese	12,503	(3,877)
Sierra Leonean	21,725	(4,454)
Somali	156,737	(13,202)
South African	69,475	(5,996)
Sudanese	48,373	(9,233)
Ugandan	20,274	(4,247)
Zimbabwean	10,167	(2,907)
African	1,931,662	(43,374)
Other Subsaharan African	247,579	(17,874)
Swedish	3,867,110	(36,970)
Swiss	957,460	(18,714)
Turkish	230,342	(13,111)
Ukrainian	1,028,492	(21,870)
Welsh	1,898,884	(29,965)

(Table continued on the next page)

<b>Ancestry</b>	<b>Estimate</b>	<b>Margin of Error</b>
West Indian (except Hispanic groups):	3,019,686	(47,108)
Bahamian	55,637	(5,576)
Barbadian	71,482	(6,166)
Belizean	62,369	(8,106)
Bermudan	5,854	(1,392)
British West Indian	103,244	(8,653)
Dutch West Indian	42,808	(4,269)
Haitian	1,049,779	(31,695)
Jamaican	1,132,460	(30,801)
Trinidadian and Tobagonian	227,523	(13,227)
U.S. Virgin Islander	19,981	(3,225)
West Indian	297,650	(14,849)
Other West Indian	8,278	(1,582)
Yugoslavian	251,961	(15,502)
Other groups	123,620,851	(168,292)
Unclassified or not reported	56,400,123	(208,945)

Source: 2016 ACS 1-year ACS Estimates

Note: Margins of error are in parentheses. An '\*\*\*\*\*' entry in the margin of error column indicates that the estimate is controlled.

## Appendix E. Expanded Version of Table 11

In Section 5.5, we presented an abbreviated version of this table, with only the groups that were statistically different between ACS Production and 2016 ACS Content Test. Below is the full table.

**Table A-3. Comparing Group Estimates Derived from Race/Hispanic Origin and Ancestry**

Ancestry or Detailed Race/Ethnicity Group	2016 ACS Production Ancestry (N=4,880,000)		2016 ACS Content Test, Race/Hispanic Origin (N=43,500)		Difference		Adjusted P-Value
Afghan	0.04	(<0.1)	0.10	(0.1)	0.06	(0.1)	1.00
Albanian	0.06	(<0.1)	0.08	(0.1)	0.01	(0.1)	1.00
Alsatian	0.00	(<0.1)	-		-		
American	6.28	(<0.1)	6.32	(0.4)	0.04	(0.4)	1.00
Arab:	0.64	(<0.1)	0.75	(0.2)	0.11	(0.2)	1.00
Egyptian	0.08	(<0.1)	0.12	(0.1)	0.04	(0.1)	1.00
Iraqi	0.05	(<0.1)	0.06	(<0.1)	0.02	(<0.1)	1.00
Jordanian	0.02	(<0.1)	0.06	(<0.1)	0.04	(<0.1)	1.00
Lebanese	0.15	(<0.1)	0.18	(0.1)	0.03	(0.1)	1.00
Moroccan	0.04	(<0.1)	0.02	(<0.1)	-0.01	(<0.1)	1.00
Palestinian	0.04	(<0.1)	0.07	(<0.1)	0.02	(<0.1)	1.00
Syrian	0.06	(<0.1)	0.05	(<0.1)	-0.01	(<0.1)	1.00
Arab	0.09	(<0.1)	-		-		
Other Arab	0.12	(<0.1)	0.14	(0.1)	0.02	(0.1)	1.00
Armenian	0.15	(<0.1)	0.16	(0.1)	0.01	(0.1)	1.00
Assyrian/Chaldean/Syriac	0.04	(<0.1)	0.03	(<0.1)	-0.01	(<0.1)	1.00
Australian	0.03	(<0.1)	0.04	(<0.1)	0.01	(<0.1)	1.00
Austrian	0.22	(<0.1)	0.23	(0.1)	0.01	(0.1)	1.00
Basque	0.02	(<0.1)	0.01	(<0.1)	-0.01	(<0.1)	1.00
Belgian	0.12	(<0.1)	0.11	(<0.1)	-0.01	(<0.1)	1.00
Brazilian	0.13	(<0.1)	0.25	(0.1)	0.11	(0.1)	1.00
British	0.48	(<0.1)	0.20	(0.1)	-0.28	(0.1)	<0.01 *
Bulgarian	0.03	(<0.1)	0.03	(<0.1)	0.00	(<0.1)	1.00
Cajun	0.03	(<0.1)	0.05	(<0.1)	0.01	(<0.1)	1.00
Canadian	0.21	(<0.1)	0.17	(<0.1)	-0.04	(0.1)	1.00
Carpatho Rusyn	0.00	(<0.1)	-		-		
Celtic	0.01	(<0.1)	-		-		
Croatian	0.13	(<0.1)	0.09	(<0.1)	-0.04	(<0.1)	1.00
Cypriot	0.00	(<0.1)	-		-		
Czech	0.44	(<0.1)	0.55	(0.1)	0.11	(0.1)	1.00
Czechoslovakian	0.10	(<0.1)	0.04	(<0.1)	-0.06	(<0.1)	<0.01 *
Danish	0.40	(<0.1)	0.37	(0.1)	-0.04	(0.1)	1.00

(Table continued on the next page)

Ancestry or Detailed Race/Ethnicity Group	2016 ACS Production Ancestry (N=4,880,000)		2016 ACS Content Test, Race/Hispanic Origin (N=43,500)		Difference		Adjusted P-Value
Dutch	1.26	(<0.1)	1.70	(0.2)	0.44	(0.2)	0.03 *
Eastern European	0.23	(<0.1)	0.17	(0.1)	-0.06	(0.1)	1.00
English	7.47	(<0.1)	11.75	(0.6)	4.28	(0.6)	<0.01 *
Estonian	0.01	(<0.1)	-		-		
European	1.41	(<0.1)	0.99	(0.2)	-0.42	(0.2)	0.17
Finnish	0.21	(<0.1)	0.22	(0.1)	0.02	(0.1)	1.00
French except Basque	2.48	(<0.1)	2.67	(0.2)	0.19	(0.2)	1.00
French Canadian	0.66	(<0.1)	0.37	(0.1)	-0.29	(0.1)	<0.01 *
German	13.96	(<0.1)	16.99	(0.7)	3.03	(0.7)	<0.01 *
German Russian	0.01	(<0.1)	-		-		
Greek	0.40	(<0.1)	0.56	(0.1)	0.16	(0.1)	1.00
Guyanese	0.08	(<0.1)	0.08	(<0.1)	0.00	(<0.1)	1.00
Hungarian	0.45	(<0.1)	0.51	(0.1)	0.07	(0.1)	1.00
Icelandic	0.02	(<0.1)	-		-		
Iranian	0.15	(<0.1)	0.24	(0.1)	0.09	(0.1)	1.00
Irish	10.05	(<0.1)	13.99	(0.7)	3.94	(0.7)	<0.01 *
Israeli	0.05	(<0.1)	0.08	(0.0)	0.03	(<0.1)	1.00
Italian	5.26	(<0.1)	5.96	(0.5)	0.70	(0.5)	0.80
Latvian	0.03	(<0.1)	0.02	(<0.1)	-0.01	(<0.1)	1.00
Lithuanian	0.20	(<0.1)	0.16	(<0.1)	-0.04	(<0.1)	1.00
Luxembourger	0.01	(<0.1)	-		-		
Macedonian	0.02	(<0.1)	-		-		
Maltese	0.01	(<0.1)	0.02	(<0.1)	0.00	(<0.1)	1.00
New Zealander	0.01	(<0.1)	-		-		
Northern European	0.11	(<0.1)	0.14	(0.1)	0.03	(0.1)	1.00
Norwegian	1.38	(<0.1)	1.23	(0.2)	-0.15	(0.2)	1.00
Pennsylvania German	0.10	(<0.1)	0.03	(<0.1)	-0.08	(<0.1)	<0.01 *
Polish	2.89	(<0.1)	3.51	(0.4)	0.62	(0.4)	0.29
Portuguese	0.42	(<0.1)	0.40	(0.1)	-0.02	(0.1)	1.00
Romanian	0.15	(<0.1)	0.19	(0.1)	0.04	(0.1)	1.00
Russian	0.86	(<0.1)	0.88	(0.2)	0.02	(0.2)	1.00
Scandinavian	0.24	(<0.1)	0.19	(0.1)	-0.04	(0.1)	1.00
Scotch-Irish	1.01	(<0.1)	0.97	(0.2)	-0.03	(0.2)	1.00
Scottish	1.77	(<0.1)	2.59	(0.3)	0.83	(0.3)	<0.01 *
Serbian	0.06	(<0.1)	0.06	(<0.1)	0.00	(<0.1)	1.00

(Table continued on the next page)

Ancestry or Detailed Race/Ethnicity Group	2016 ACS Production Ancestry (N=4,880,000)		2016 ACS Content Test, Race/Hispanic Origin (N=43,500)		Difference		Adjusted P-Value
Slavic	0.04	(<0.1)	0.06	(<0.1)	0.02	(<0.1)	1.00
Slovak	0.22	(<0.1)	0.23	(0.1)	0.01	(0.1)	1.00
Slovene	0.06	(<0.1)	0.05	(<0.1)	-0.01	(<0.1)	1.00
Soviet Union	0.00	(<0.1)	-		-		
Subsaharan African:	1.09	(<0.1)	0.85	(0.2)	-0.24	(0.2)	1.00
Cape Verdean	0.04	(<0.1)	0.03	(<0.1)	0.00	(<0.1)	1.00
Ethiopian	0.10	(<0.1)	0.04	(<0.1)	-0.05	(<0.1)	<0.01 *
Ghanaian	0.04	(<0.1)	0.11	(0.2)	0.07	(0.2)	1.00
Kenyan	0.02	(<0.1)	0.01	(<0.1)	-0.02	(<0.1)	<0.01 *
Liberian	0.03	(<0.1)	0.02	(<0.1)	-0.01	(<0.1)	1.00
Nigerian	0.12	(<0.1)	0.17	(0.1)	0.05	(0.1)	1.00
Senegalese	0.00	(<0.1)	-		-		
Sierra Leonean	0.01	(<0.1)	-		-		
Somali	0.05	(<0.1)	0.03	(<0.1)	-0.02	(<0.1)	1.00
South African	0.02	(<0.1)	0.03	(<0.1)	0.01	(<0.1)	1.00
Sudanese	0.01	(<0.1)	-		-		
Ugandan	0.01	(<0.1)	-		-		
Zimbabwean	0.00	(<0.1)	-		-		
African	0.60	(<0.1)	0.33	(0.1)	-0.27	(0.1)	0.16
Other Subsaharan African	0.06	(<0.1)	0.04	(<0.1)	-0.02	(<0.1)	1.00
Swedish	1.20	(<0.1)	1.29	(0.2)	0.08	(0.2)	1.00
Swiss	0.30	(<0.1)	0.35	(0.1)	0.05	(0.1)	1.00
Turkish	0.07	(<0.1)	0.06	(<0.1)	-0.01	(<0.1)	1.00
Ukrainian	0.32	(<0.1)	0.26	(0.1)	-0.06	(0.1)	1.00
Welsh	0.59	(<0.1)	0.74	(0.1)	0.15	(0.2)	1.00
West Indian (except Hispanic):	0.94	(<0.1)	0.85	(0.2)	-0.09	(0.2)	1.00
Bahamian	0.02	(<0.1)	0.01	(<0.1)	-0.01	(<0.1)	1.00
Barbadian	0.02	(<0.1)	0.02	(<0.1)	-0.01	(<0.1)	1.00
Belizean	0.02	(<0.1)	0.01	(<0.1)	-0.01	(<0.1)	1.00
Bermudan	0.00	(<0.1)	-		-		
Haitian	0.33	(<0.1)	0.19	(<0.1)	-0.14	(<0.1)	<0.01 *

(Table continued on the next page)



Ancestry or Detailed Race/Ethnicity Group	2016 ACS Production Ancestry (N=4,880,000)		2016 ACS Content Test, Race/Hispanic Origin (N=43,500)		Difference		Adjusted P-Value
	Estimate	Margin of Error	Estimate	Margin of Error	Estimate	Margin of Error	
Jamaican	0.35	(<0.1)	0.40	(0.1)	0.04	(0.1)	1.00
Trinidadian and Tobagonian	0.07	(<0.1)	0.08	(0.1)	0.01	(0.1)	1.00
U.S. Virgin Islander	0.01	(<0.1)	0.01	(<0.1)	0.00	(<0.1)	1.00
Other West Indian	0.14	(<0.1)	0.15	(0.1)	0.01	(0.1)	1.00
Yugoslavian	0.08	(<0.1)	0.04	(<0.1)	-0.04	(<0.1)	1.00
Unclassified	1.95	(<0.1)	2.86	(0.3)	0.91	(0.3)	<0.01 *
Not reported or missing	15.54	(0.1)	0.83	(0.2)	-14.70	(0.2)	<0.01 *

Sources: U. S. Census Bureau, 2016 American Community Survey 1-year Estimates & 2016 ACS Content Test, Control Treatment.

Note: Margins of error are shown in parentheses. Significant at  $\alpha=0.1$  level. P-values with an asterisk (\*) indicate a significant difference between the two rates. P-values have been adjusted for multiple comparisons using the Holm-Bonferroni method. An '-' entry in the estimate column indicates that either no sample observations or too few sample observations were available to compute an estimate. Data for housing units in Alaska and Hawaii were excluded from the 2016 American Community Survey 1-year estimates in this table to make them comparable to the 2016 ACS Content Test data, which did not sample in these states. DRB Approval Number CBDRB-FY19-RAGLIN-B0005.

## **Appendix F. Analysis Tables using the 2016 ACS Content Test, Test Treatment**

This appendix discusses the test treatment from the 2016 ACS Content Test. In this appendix we present the same analysis tables from the body of the report, but replace the 2016 ACS Content Test control treatment data with test treatment data. Since the control treatment design is being incorporated into production for 2020, we used the control treatment to answer the research questions. We document the test treatment results here so that this information is available if there is future interest in using the test treatment design.

The layout of the test treatment can be seen in Appendices A, B, and C. Notably, the test treatment asked Hispanic origin and race in one combined race or ethnicity question, included a separate category for Middle Eastern or North African, and provided checkboxes (in the internet instrument) for detailed groups within each major category.

### **Research Question 1**

*What is the response quality, in terms of response presence and validity, to the race/Hispanic origin and ancestry questions? How does response quality vary by mode and by race and Hispanic origin?*

Previous analysis reported the item nonresponse rates for race/Hispanic origin and Ancestry. Results showed that of the people rostered on the test treatment forms:

- 0.6 percent either completely skipped the combined ethnicity question or only provided an uncodable answer in that series, and
- 15.7 percent did not respond to the ancestry question (Harth et al., 2017).

The interactions of these two questions and the amount of valid data provided for individuals in the test treatment of the 2016 ACS Content Test is shown in Table A-4.

**Table A-4: Presence and Validity of Race/Hispanic Origin and Ancestry Responses (Test Treatment)**

Validity of Response to Race/Hispanic origin and Ancestry	Overall Universe (N=43,500)	Internet Responses (N=21,000)	Mail Responses (N=10,000)	CATI Responses (N=2,000)	CAPI Responses (N=10,500)
Valid response to both	82.5 (0.7)	80.0 (0.9)	79.3 (1.0)	87.2 (2.7)	86.5 (1.3)
One Ancestry provided	53.9 (1.0)	42.9 (1.1)	52.3 (1.5)	57.0 (3.3)	66.8 (2.1)
Two Ancestries provided	28.6 (0.9)	37.1 (0.9)	27.1 (1.5)	30.2 (3.1)	19.7 (1.7)
Valid Race/Hispanic origin, blank Ancestry	1.7 (0.2)	2.9 (0.4)	2.1 (0.5)	0.02 (0.03)	0.3 (0.1)
Valid Race/Hispanic origin, uncodable Ancestry	15.2 (0.6)	16.6 (0.8)	17.7 (0.9)	12.4 (2.7)	12.5 (1.3)
Nonresponse to both	0.5 (0.1)	0.5 (0.2)	0.5 (0.2)	0.2 (0.2)	0.6 (0.3)
Other combinations	0.1 (0.1)	0.02 (0.02)	0.4 (0.2)	0.2 (0.2)	0.2 (0.1)
<b>TOTAL</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>

Sources: U. S. Census Bureau, 2016 American Community Survey Content Test, Test Treatment.

Notes: Margin of error is provided in parentheses. Minor additive discrepancies are due to rounding. DRB Approval Number CBDRB-FY19-RAGLIN-B0005.

Overall, 82.5 percent of people provided a valid response to both the race/Hispanic origin questions and the ancestry question while 15.2 percent provided a valid race/Hispanic origin but left ancestry blank. Another 1.7 percent of people provided a valid race/Hispanic origin response but an uncodable ancestry response. Less than one percent of people did not respond to any of the race, Hispanic origin, or ancestry questions (0.5 percent).

## Research Question 2

*How many people have ancestry responses that matched exactly, half matched (one of the two ancestries matched exactly), or did not match to race/Hispanic origin responses? How did these match rates vary by response mode?*

For those people with valid data in both questions, the following possible categories were defined in order to compare race and Hispanic origin information to ancestry information.

- Exact match
- Half match
- No match

The results to the research question are shown below. Only those persons with a valid response to both ancestry and the race and Hispanic origin series are included in this analysis.

**Table A-5: Match Rates For All Persons in the Test Treatment of the 2016 ACS Content Test**

	<b>Overall (N=35,500)</b>	<b>Internet (N=16,500)</b>	<b>Mail (N=7,900)</b>	<b>CATI (N=1,700)</b>	<b>CAPI (N=9,400)</b>
Exact Match	74.0 (0.8)	73.8 (1.0)	54.1 (1.9)	77.0 (2.7)	82.7 (1.4)
One ancestry	51.2 (1.0)	42.1 (1.1)	37.5 (1.9)	53.3 (3.5)	66.6 (1.9)
Two ancestries	22.7 (0.9)	31.7 (1.1)	16.6 (1.1)	23.7 (3.0)	16.0 (1.6)
Half Match	7.9 (0.5)	11.4 (0.6)	5.0 (0.6)	7.9 (1.8)	5.4 (0.9)
No match	18.2 (0.7)	14.8 (0.8)	40.9 (1.8)	15.1 (2.7)	11.9 (1.3)
One ancestry	14.1 (0.6)	11.5 (0.7)	28.4 (1.6)	12.0 (2.2)	10.6 (1.2)
Two ancestries	4.1 (0.3)	3.3 (0.4)	12.5 (1.1)	3.1 (1.4)	1.3 (0.4)
<b>Total</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>

Sources: U. S. Census Bureau, 2016 American Community Survey Content Test, Test Treatment.

Notes: Margin of error is shown in parentheses. Minor additive discrepancies are due to rounding. DRB Approval Number CDBRB-FY19-RAGLIN-B0005.

As was reported in the 2016 ACS Content Test Report, of all persons who provided a valid response in both ancestry and race/Hispanic origin, 74.0 percent provided ancestry responses that were exactly the same as information found in race/Hispanic origin. This table shows that 7.9 percent of persons who provided a valid response in both ancestry and race/Hispanic origin had a half match and the remaining 18.2 percent had no response in ancestry that also exactly matched to their race/Hispanic origin responses.

Similar to the control treatment, individuals who were enumerated on a mail questionnaire were less likely to have exact matches between race/Hispanic origin and ancestry, compared to other modes of response. On mail questionnaires, 12.5 percent of persons provided two ancestries and neither ancestry was found to be reported in race/Hispanic origin. From internet responses, only 3.3 percent of individuals fit that description.

We also analyzed the match rates at an ancestry-level, for each individual ancestry that was coded in the 2016 ACS Content Test test treatment where a person also gave valid race and ethnicity information.

**Table A-6. Match Rates for all Reported Ancestries (Test Treatment)**

	<b>Internet</b>	<b>Mail</b>	<b>CATI</b>	<b>CAPI</b>	<b>TOTAL</b>
Ancestries Matched to Race	79.8 (0.9)	56.5 (1.8)	80.7 (2.6)	84.8 (1.3)	<b>77.6 (0.7)</b>
Ancestries Not Matched to Race	20.1 (0.8)	43.5 (1.8)	19.3 (2.6)	15.2 (1.3)	<b>22.3 (0.7)</b>
<b>TOTAL</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>

Sources: U. S. Census Bureau, 2016 American Community Survey Content Test, Test Treatment.

Notes: Margins of error are shown in parentheses.

Table A-6 shows that 77.6 percent of ancestries reported were found to be identical matches to something reported in race or Hispanic origin, while 22.3 percent were not. Ancestries reported on the paper (mail) questionnaire were the least likely to be matched to a race or Hispanic origin response (56.5 percent).

### Research Question 3

*Of the ancestry responses that did not match the reported race/Hispanic origin, how often were the ancestry responses in the same OMB race/Hispanic origin category? When in the same OMB category, how often did the reported ancestry provide more specificity, the same level of specificity, or less specificity as the reported race/Hispanic origin?*

The table shows whether the major classification of the ancestry response is the same as a major group provided for that person in their race and Hispanic origin data, even if the specific response was not the same.

**Table A-7. Match Rates, Difference in Major Category, and Specificity of Information Provided by Ancestry Response Compared With Race/Hispanic Origin Response, by Race or Hispanic Origin of the Ancestry Response (Test Treatment)**

	White (N=25,500)	Black (N=5,900)	Asian (N=2,300)	AIAN (N=950)	NHPI (N=60)	SOR (N=700)	Hispanic (N=8,200)	Total (N=47,000)
Ancestry matched	79.4 (0.9)	93.5 (1.2)	90.5 (2.6)	48.5 (5.1)	84.4 (11.1)	35.8 (6.2)	80.4 (1.7)	<b>77.6 (0.7)</b>
Ancestry did not match	20.6 (0.9)	6.4 (1.2)	9.5 (2.6)	51.6 (5.1)	15.6 (11.1)	64.2 (6.2)	19.6 (1.7)	<b>22.3 (0.7)</b>
Different major category	0.8 (0.2)	1.9 (1.0)	1.5 (0.5)	30.5 (4.3)	4.9 (5.5)	59.7 (6.0)	2.4 (0.6)	<b>2.6 (0.2)</b>
Unclassified	-	-	-	-	-	-	-	<b>3.6 (0.3)</b>
Same major category	19.8 (0.9)	4.5 (0.8)	8.0 (2.5)	21.1 (4.6)	10.7 (10.6)	4.5 (1.5)	17.2 (1.6)	<b>16.2 (0.6)</b>
Ancestry more specific	11.0 (0.8)	3.0 (0.6)	3.8 (1.4)	0.3 (0.2)	4.6 (5.9)	2.4 (0.9)	3.8 (0.7)	<b>8.0 (0.5)</b>
Same specificity	8.7 (0.6)	1.5 (0.5)	4.1 (2.0)	17.9 (3.7)	6.1 (9.4)	0.4 (0.4)	13.4 (1.5)	<b>8.1 (0.4)</b>
Ancestry less specific	<0.1 (<0.1)	0.1 (0.1)	0.1 (0.1)	2.8 (2.1)	-	1.8 (0.9)	0.1 (0.1)	<b>0.1 (0.1)</b>
<b>Total ancestry responses</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>

Source: U. S. Census Bureau, 2016 American Community Survey Content Test, Test Treatment.

Notes: Margins of error are shown in parentheses. An '-' entry in the estimate column indicates that either no sample observations or too few sample observations were available to compute an estimate. Minor additive discrepancies are due to rounding. DRB Approval Number CBDRB-FY19-RAGLIN-B0005.

The Total column shows that of all the ancestries in the 2016 ACS Content Test test treatment,

- 77.6 percent matched exactly to something reported in race and Hispanic origin, while 22.3 percent did not.
- 2.6 percent did not match exactly and provided entirely new major race information. An example is if “NAVAJO” was reported in ancestry and no indication of AIAN identification was reported in race.
- 3.6 percent did not match exactly and provided new information but of an ancestry considered to not be a major race category (i.e., ancestry was “AMERICAN”).

- 8.0 percent did not match exactly but were not only in the same major category as something in race but provided a more specific description of a major race category that had been reported. For instance, Irish in ancestry with only the White checkbox marked in race.
- 8.1 percent did not match exactly but were in the same major category as something reported in race or Hispanic origin with comparably specific descriptions of that major category. For example, someone wrote in “Native American” for ancestry after marking the AIAN checkbox for race.
- 0.1 percent did not match exactly but were in the same major category as something reported in race and Hispanic origin with less specific descriptions than was reported in race and Hispanic origin. For example, a person had marked the Chinese checkbox in race but their ancestry response was “Asian.”

From the control treatment in Table 8, 18.1 percent of all White ancestries did not match exactly to race information and provided more specific information. However in the test treatment, the comparable result was 11.1 percent of White ancestries. This is likely due to the detailed race checkboxes that were in the internet instrument of the test treatment. However, Hispanic ancestry and Asian ancestry that were unmatched to race were more likely than their control treatment counterparts to be more specific in ancestry than in Race.

## Research Question 4

*Of the people who identified with a particular group in either the race/Hispanic origin question or ancestry question, what percent identified with the group (1) in both responses, (2) in only the detailed race/Hispanic origin response, or (3) in only the ancestry response?*

Tables A-8 and A-9 show findings from the test treatment relating to Research Question 4. In that analysis, we showed that when reported in only one item, detailed Hispanic and Asian groups had higher overall reporting in the race/Hispanic origin item, while White detailed groups had higher reporting in the ancestry item. It was also shown that some respondents who provided a detailed White or Black group in ancestry only checked the White or Black checkbox in the race survey item.

Table A-9 shows that in the test treatment, detailed groups from every major category except SOR had higher reporting in the race/Hispanic origin item.<sup>39</sup> While there were still some respondents who only checked the White or Black checkbox in the race item and provided a detailed White or Black group in ancestry, that number seemed to be reduced in the test treatment, which had detailed checkbox options in the internet mode.

Also in the control treatment, 11 of the detailed groups analyzed had higher reporting in in the race/Hispanic origin items, and no groups had higher reporting in the ancestry survey item. Table A-9 shows that in the test treatment, 15 detailed groups had higher reporting in the in the race/Hispanic origin items when only reported in one location, while no groups had more detailed reporting on the ancestry item.

**Table A-8. Location of Detailed Responses Across Survey Items, by Major Category**

Major Category	N	Reported in				Difference		Adjusted P-value	
		Reported in Both	Race/Hispanic Origin Only	Reported in Ancestry Only	(Race/HO only – Ancestry only)				
White	26,000	70.3 (1.0)	28.8 (1.0)	0.9 (0.2)	27.9 (1.0)	<0.01	*		
Black	7,400	80.1 (1.8)	18.2 (1.7)	1.6 (0.9)	16.6 (1.9)	<0.01	*		
Asian	2,900	76.2 (2.5)	22.5 (2.5)	1.2 (0.4)	21.3 (2.6)	<0.01	*		
AIAN	1,700	39.2 (4.0)	43.5 (3.7)	17.3 (2.4)	26.2 (4.8)	<0.01	*		
NHPI	100	42.1 (13.9)	54.5 (14.8)	3.4 (3.3)	51.0 (16.3)	<0.01	*		
SOR	1,400	20.6 (4.4)	49.2 (5.1)	30.1 (3.3)	19.1 (7.3)	<0.01	*		
Hispanic	9,600	81.2 (1.8)	16.7 (1.8)	2.1 (0.5)	14.6 (2.0)	<0.01	*		

Source: U. S. Census Bureau, 2016 American Community Survey Content Test, Test Treatment.

Note: Margins of error are shown in parentheses. Significant at  $\alpha=0.1$  level. P-values with an asterisk (\*) indicate a significant difference between the two rates. P-values have been adjusted for multiple comparisons using the Holm-Bonferroni method. DRB Approval Number CBDRB-FY19-RAGLIN-B0005.

<sup>39</sup> Again, the Some Other Race category in ancestry likely has elevated response due to the write-in “Indian” being coded distinctly from “Asian Indian.”

**Table A-9. Location of Detailed Responses Across Survey Items, by Major Category and Detailed Group**

Major Category or Detailed Group	N	Reported in		Reported in		Reported in		Difference		Adjusted P-value
		Both	(0.8)	Race/Hispanic Origin Only	(0.8)	Ancestry Only	(0.5)	(Race/HO only – Ancestry only)	(1.1)	
Any detailed	38,000	75.4	(0.8)	16.6	(0.8)	8.0	(0.5)	8.6	(1.1)	<0.01
White, detailed	21,500	72.4	(0.9)	17.9	(0.9)	9.7	(0.7)	8.1	(1.3)	<0.01 *
German	8,200	60.8	(1.7)	28.6	(1.6)	10.6	(0.9)	18.0	(1.9)	<0.01 *
Irish	7,300	52.2	(1.9)	38.2	(2.0)	9.6	(1.0)	28.6	(2.5)	<0.01 *
English	8,700	44.4	(1.6)	49.1	(1.7)	6.5	(0.8)	42.5	(2.1)	<0.01 *
Italian	2,700	62.9	(2.8)	25.7	(2.4)	11.4	(1.8)	14.3	(3.2)	<0.01 *
Polish	1,900	51.8	(2.9)	36.4	(2.8)	11.8	(2.0)	24.6	(4.0)	<0.01 *
French	2,300	39.1	(2.8)	51.1	(2.7)	9.8	(1.3)	41.2	(3.2)	<0.01 *
Black, detailed	6,500	72.6	(2.1)	16.4	(1.7)	11.0	(1.6)	5.4	(2.5)	<0.01 *
African American	5,500	68.6	(2.4)	20.0	(2.0)	11.5	(1.8)	8.5	(2.9)	<0.01 *
Jamaican	300	60.4	(13.2)	21.4	(9.8)	18.2	(8.2)	3.1	(12.4)	1.00
Haitian	200	72.6	(12.0)	11.4	(5.9)	16.0	(11.4)	-4.5	(13.6)	1.00
Nigerian	90	88.8	(9.9)	6.7	(7.0)	4.5	(5.2)	2.2	(7.4)	1.00
Ethiopian	80	67.3	(23.4)	10.5	(9.5)	22.3	(20.5)	-11.8	(21.7)	1.00
Somali	30	66.7	(38.7)	-		-		-		
Asian, detailed	2,700	71.6	(2.6)	24.8	(2.6)	3.6	(1.1)	21.2	(3.0)	<0.01 *
Chinese	750	73.9	(4.8)	22.3	(4.6)	3.8	(1.7)	18.6	(5.0)	<0.01 *
Filipino	550	77.4	(6.6)	19.4	(6.2)	3.2	(2.1)	16.2	(6.5)	<0.01 *
Asian Indian	500	38.0	(9.6)	59.0	(9.9)	2.9	(3.1)	56.1	(11.2)	<0.01 *
Vietnamese	250	73.8	(8.0)	24.0	(8.0)	2.2	(1.6)	21.8	(8.4)	<0.01 *
Korean	250	73.0	(10.3)	23.6	(9.9)	3.5	(2.8)	20.1	(10.3)	0.02 *
Japanese	150	73.6	(9.5)	16.5	(7.3)	9.8	(5.9)	6.7	(9.2)	1.00
NHPI, detailed	100	40.9	(14.3)	53.6	(15.0)	5.5	(4.5)	48.1	(16.9)	<0.01 *
Native Hawaiian	40	-		70.0	(26.0)	-		-		-
Samoan	-	-		-		-		-		-
Chamorro	20	65.0	(34.8)	35.0	(34.8)	-		35.0	(34.8)	1.00
Tongan	-	-		-		-		-		-
Fijian	-	-		-		-		-		-
Marshallese	-	-		-		-		-		-

(Table continued on next page)



Major Category or Detailed Group	N	Reported in Both		Reported in REO Only		Reported in Ancestry Only		Difference (Race/HO only – Ancestry only)		Adjusted P-value
SOR, detailed	1,000	25.0	(5.4)	34.1	(6.0)	40.9	(4.2)	-6.8	(8.9)	1.00
Brazilian	100	57.6	(21.0)	34.8	(21.8)	7.6	(6.4)	27.1	(24.4)	0.74
Belizean	10	-		84.4	(36.6)	-		-		-
Cabo Verdean	30	79.8	(20.5)	-		-		-		-
Guyanese	70	46.1	(18.1)	11.7	(9.9)	42.3	(17.4)	-30.6	(21.7)	0.26
Creole	20	-		-		88.9	(21.9)	-		-
North American	10	-		55.8	(54.9)	-		-		-
Hispanic, detailed	8,900	74.8	(2.0)	20.6	(2.0)	4.6	(0.8)	16.0	(2.3)	<0.01 *
Mexican	5,800	76.8	(2.7)	19.3	(2.7)	3.9	(0.8)	15.5	(2.9)	<0.01 *
Puerto Rican	850	61.7	(6.2)	32.8	(5.8)	5.5	(2.9)	27.3	(6.8)	<0.01 *
Cuban	400	73.5	(6.5)	13.7	(4.0)	12.8	(6.5)	0.9	(8.7)	1.00
Salvadoran	450	86.1	(4.8)	10.4	(4.1)	3.4	(2.2)	7.0	(4.5)	0.16
Dominican	450	71.8	(9.8)	23.7	(8.6)	4.5	(3.5)	19.3	(8.8)	0.01 *
Colombian	200	53.5	(14.4)	36.1	(17.0)	10.4	(5.8)	25.7	(20.9)	0.52

Source: U. S. Census Bureau, 2016 American Community Survey Content Test, Test Treatment.

Note: Margins of error are shown in parentheses. Significant at  $\alpha=0.1$  level. P-values with an asterisk (\*) indicate a significant difference between the two rates. P-values have been adjusted for multiple comparisons using the Holm-Bonferroni method. An '-' entry in the estimate column indicates that either no sample observations or too few sample observations were available to compute an estimate. DRB Approval Number CBDRB-FY19-RAGLIN-B0005.

## Research Question 5

*How do group estimates derived from the 2016 ACS Content Test race and Hispanic origin survey items compare to those derived from the production ancestry survey item?*

Table A-10 shows the test treatment version of Table A-3, which is the expanded version of the first table from Research Question 5, comparing detailed group estimates derived from race/Hispanic origin data in the 2016 ACS Content Test with production ancestry group estimates. Findings from the test treatment echo many of those in the control treatment. The changes in the white detailed groups appear larger in the test treatment than in the control treatment, particularly the groups that had detailed checkboxes in the internet instrument. In the Test treatment there was the additional evidence of generic, regional groups having lower estimates, such as European, Eastern European, Norther European, and African. Reporting of American was also lower. The biggest statistical difference was the percent not reported or missing, which was 15.54 percent in production ancestry, while only 0.63 percent in the race/Hispanic Origin item.

**Table A-10. Comparing Estimates Derived from Race/Hispanic Origin and Ancestry, by Detailed Group**

Ancestry or Detailed Race/Ethnicity Group	2016 ACS Production Ancestry (N=4,880,000)		2016 ACS Content Test, Race/Hispanic Origin (N=43,500)		Difference		Adjusted P-Value
	Estimate	CI	Estimate	CI	Estimate	CI	
Afghan	0.04	(<0.1)	0.03	(<0.1)	-0.01	(<0.1)	1.00
Albanian	0.06	(<0.1)	0.07	(0.1)	0.01	(0.1)	1.00
Alsatian	0.00	(<0.1)	-		-		-
American	6.28	(<0.1)	4.36	(0.4)	-1.92	(0.4)	<0.01 *
Arab:	0.64	(<0.1)	0.86	(0.2)	0.22	(0.2)	1.00
Egyptian	0.08	(<0.1)	0.19	(0.1)	0.11	(0.1)	1.00
Iraqi	0.05	(<0.1)	0.09	(0.1)	0.05	(0.1)	1.00
Jordanian	0.02	(<0.1)	0.16	(0.1)	0.14	(0.1)	1.00
Lebanese	0.15	(<0.1)	0.08	(<0.1)	-0.07	(<0.1)	0.02 *
Moroccan	0.04	(<0.1)	0.04	(<0.1)	0.00	(<0.1)	1.00
Palestinian	0.04	(<0.1)	0.10	(<0.1)	0.06	(<0.1)	1.00
Syrian	0.06	(<0.1)	0.07	(<0.1)	0.02	(<0.1)	1.00
Arab	0.09	(<0.1)	-		-		-
Other Arab	0.12	(<0.1)	0.14	(0.1)	0.02	(0.1)	1.00
Armenian	0.15	(<0.1)	0.11	(0.1)	-0.04	(0.1)	1.00
Assyrian/Chaldean/Syriac	0.04	(<0.1)	0.04	(<0.1)	0.00	(<0.1)	1.00
Australian	0.03	(<0.1)	0.07	(0.1)	0.04	(0.1)	1.00
Austrian	0.22	(<0.1)	0.24	(0.1)	0.02	(0.1)	1.00
Basque	0.02	(<0.1)	0.01	(<0.1)	-0.01	(<0.1)	1.00
Belgian	0.12	(<0.1)	0.22	(0.1)	0.10	(0.1)	1.00
Brazilian	0.13	(<0.1)	0.24	(0.1)	0.11	(0.1)	1.00
British	0.48	(<0.1)	0.14	(0.1)	-0.35	(0.1)	<0.01 *
Bulgarian	0.03	(<0.1)	0.01	(<0.1)	-0.03	(<0.1)	<0.01 *
Cajun	0.03	(<0.1)	0.05	(<0.1)	0.01	(<0.1)	1.00
Canadian	0.21	(<0.1)	0.21	(0.1)	0.01	(0.1)	1.00
Carpatho Rusyn	0.00	(<0.1)	-		-		-
Celtic	0.01	(<0.1)	-		-		-
Croatian	0.13	(<0.1)	0.18	(0.1)	0.06	(0.1)	1.00
Cypriot	0.00	(<0.1)	-		-		-
Czech	0.44	(<0.1)	0.47	(0.1)	0.03	(0.1)	1.00
Czechoslovakian	0.10	(<0.1)	0.10	(<0.1)	0.01	(<0.1)	1.00
Danish	0.40	(<0.1)	0.64	(0.1)	0.24	(0.1)	0.30
Dutch	1.26	(<0.1)	1.83	(0.2)	0.56	(0.2)	<0.01 *
Eastern European	0.23	(<0.1)	0.06	(<0.1)	-0.17	(<0.1)	<0.01 *
English	7.47	(<0.1)	19.56	(0.6)	12.09	(0.6)	<0.01 *
Estonian	0.01	(<0.1)	-		-		-
European	1.41	(<0.1)	0.31	(0.1)	-1.10	(0.1)	<0.01 *

(Table continued on the next page)

Ancestry or Detailed Race/Ethnicity Group	2016 ACS Production Ancestry (N=4,880,000)		2016 ACS Content Test, Race/Hispanic Origin (N=43,500)		Difference		Adjusted P-Value
Finnish	0.21	(<0.1)	0.25	(0.1)	0.04	(0.1)	1.00
French except Basque	2.48	(<0.1)	5.17	(0.3)	2.68	(0.3)	<0.01 *
French Canadian	0.66	(<0.1)	0.26	(0.1)	-0.40	(0.1)	<0.01 *
German	13.96	(<0.1)	19.93	(0.7)	5.98	(0.7)	<0.01 *
German Russian	0.01	(<0.1)	-		-		-
Greek	0.40	(<0.1)	0.41	(0.1)	0.01	(0.1)	1.00
Guyanese	0.08	(<0.1)	0.05	(<0.1)	-0.03	(<0.1)	1.00
Hungarian	0.45	(<0.1)	0.55	(0.1)	0.10	(0.1)	1.00
Icelander	0.02	(<0.1)	-		-		-
Iranian	0.15	(<0.1)	0.31	(0.1)	0.17	(0.1)	1.00
Irish	10.05	(<0.1)	17.40	(0.8)	7.35	(0.8)	<0.01 *
Israeli	0.05	(<0.1)	0.07	(<0.1)	0.02	(<0.1)	1.00
Italian	5.26	(<0.1)	6.64	(0.4)	1.38	(0.4)	<0.01 *
Latvian	0.03	(<0.1)	0.02	(<0.1)	-0.01	(0.0)	1.00
Lithuanian	0.20	(<0.1)	0.17	(<0.1)	-0.02	(0.0)	1.00
Luxembourger	0.01	(<0.1)	-		-		-
Macedonian	0.02	(<0.1)	-		-		-
Maltese	0.01	(<0.1)	0.02	(<0.1)	0.00	(<0.1)	1.00
New Zealander	0.01	(<0.1)	-		-		-
Northern European	0.11	(<0.1)	0.03	(<0.1)	-0.08	(<0.1)	<0.01 *
Norwegian	1.38	(<0.1)	1.66	(0.2)	0.29	(0.2)	0.85
Pennsylvania German	0.10	(<0.1)	0.05	(<0.1)	-0.05	(<0.1)	0.82
Polish	2.89	(<0.1)	4.49	(0.4)	1.60	(0.4)	<0.01 *
Portuguese	0.42	(<0.1)	0.52	(0.1)	0.10	(0.1)	1.00
Romanian	0.15	(<0.1)	0.10	(<0.1)	-0.04	(<0.1)	1.00
Russian	0.86	(<0.1)	0.91	(0.1)	0.05	(0.1)	1.00
Scandinavian	0.24	(<0.1)	0.22	(0.1)	-0.02	(0.1)	1.00
Scotch-Irish	1.01	(<0.1)	0.28	(0.1)	-0.72	(0.1)	<0.01 *
Scottish	1.77	(<0.1)	3.71	(0.3)	1.94	(0.3)	<0.01 *
Serbian	0.06	(<0.1)	0.07	(<0.1)	0.02	(<0.1)	1.00
Slavic	0.04	(<0.1)	0.04	(<0.1)	0.00	(<0.1)	1.00
Slovak	0.22	(<0.1)	0.24	(0.1)	0.03	(0.1)	1.00
Slovene	0.06	(<0.1)	0.04	(<0.1)	-0.02	(<0.1)	1.00
Soviet Union	0.00	(<0.1)	-		-		-
Subsaharan African:	1.09	(<0.1)	0.96	(0.2)	-0.13	(0.2)	1.00
Cape Verdean	0.04	(<0.1)	0.06	(<0.1)	0.03	(<0.1)	1.00
Ethiopian	0.10	(<0.1)	0.18	(0.1)	0.09	(0.1)	1.00
Ghanaian	0.04	(<0.1)	0.06	(<0.1)	0.02	(<0.1)	1.00

(Table continued on the next page)

Ancestry or Detailed Race/Ethnicity Group	2016 ACS Production Ancestry (N=4,880,000)		2016 ACS Content Test, Race/Hispanic Origin (N=43,500)		Difference		Adjusted P-Value
	Estimate	Margin of Error	Estimate	Margin of Error	Estimate	Margin of Error	
Kenyan	0.02	(<0.1)	0.02	(<0.1)	-0.01	(<0.1)	1.00
Liberian	0.03	(<0.1)	0.02	(<0.1)	-0.01	(<0.1)	1.00
Nigerian	0.12	(<0.1)	0.18	(0.1)	0.06	(0.1)	1.00
Senegalese	0.00	(<0.1)	-		-		-
Sierra Leonean	0.01	(<0.1)	-		-		-
Somali	0.05	(<0.1)	0.06	(0.1)	0.01	(0.1)	1.00
South African	0.02	(<0.1)	0.02	(<0.1)	0.00	(<0.1)	1.00
Sudanese	0.01	(<0.1)	-		-		-
Ugandan	0.01	(<0.1)	-		-		-
Zimbabwean	0.00	(<0.1)	-		-		-
African	0.60	(<0.1)	0.27	(0.1)	-0.33	(0.1)	<0.01 *
Other Subsaharan African	0.06	(<0.1)	0.09	(<0.1)	0.03	(<0.1)	1.00
Swedish	1.20	(<0.1)	1.36	(0.2)	0.16	(0.2)	1.00
Swiss	0.30	(<0.1)	0.31	(0.1)	0.01	(0.1)	1.00
Turkish	0.07	(<0.1)	0.05	(<0.1)	-0.02	(<0.1)	1.00
Ukrainian	0.32	(<0.1)	0.25	(0.1)	-0.07	(0.1)	1.00
Welsh	0.59	(<0.1)	0.58	(0.1)	-0.01	(0.1)	1.00
West Indian (except Hispanic):	0.94	(<0.1)	0.84	(0.2)	-0.10	(0.2)	1.00
Bahamian	0.02	(<0.1)	0.03	(<0.1)	0.01	(<0.1)	1.00
Barbadian	0.02	(<0.1)	0.03	(<0.1)	0.01	(<0.1)	1.00
Belizean	0.02	(<0.1)	0.02	(<0.1)	0.00	(<0.1)	1.00
Bermudan	0.00	(<0.1)	-		-		-
Haitian	0.33	(<0.1)	0.28	(0.1)	-0.05	(0.1)	1.00
Jamaican	0.35	(<0.1)	0.36	(0.1)	0.01	(0.1)	1.00
Trinidadian and Tobagonian	0.07	(<0.1)	0.07	(<0.1)	0.00	(<0.1)	1.00
U.S. Virgin Islander	0.01	(<0.1)	0.01	(<0.1)	0.00	(<0.1)	1.00
Other West Indian	0.14	(<0.1)	0.05	(<0.1)	-0.09	(<0.1)	<0.01 *
Yugoslavian	0.08	(<0.1)	0.09	(<0.1)	0.01	(<0.1)	1.00
Unclassified	1.95	(<0.1)	2.27	(0.3)	0.32	(0.3)	1.00
Not reported or missing	15.54	(0.1)	0.63	(0.2)	-14.91	(0.2)	<0.01 *

Sources: U. S. Census Bureau, 2016 American Community Survey 1-year Estimates and 2016 American Community Survey Content Test, Test Treatment.

Note: Margins of error are shown in parentheses. Significant at  $\alpha=0.1$  level. P-values with an asterisk (\*) indicate a significant difference between the two rates. P-values have been adjusted for multiple comparisons using the Holm-Bonferroni method. An '-' entry in the estimate column indicates that either no sample observations or too few sample observations were available to compute an estimate. Data for housing units in Alaska and Hawaii were excluded from the 2016 American Community Survey 1-year estimates in this table to make them comparable to the 2016 ACS Content Test data, which did not sample in these states. DRB Approval Number CBDRB-FY19-RAGLIN-B0005.

Table A-11 reflects many of the findings from Table 12 from Research Question 5. The test treatment of the 2016 ACS Content Test had more White, Hispanic, and Asian detailed reporting than ancestry in production. There was also more overall detailed reporting. Also similar to the control treatment, no larger OMB race or Hispanic origin group in the test treatment had less detailed reporting than in production.

**Table A-11. Comparing Estimates Derived from Race/Hispanic Origin and Ancestry, by Major Category and Detailed Group**

Major Category or Detailed Group	2016 ACS Production Ancestry (N=4,880,000)		2016 ACS Content Test, Race/Hispanic Origin (N=43,500)		Difference		Adjusted P-Value	
Detailed, overall	70.82	(0.1)	80.02	(0.8)	9.22	(0.8)	<0.01	*
White	45.57	(0.1)	67.70	(1.0)	22.13	(1.0)	<0.01	*
White, detailed	42.12	(0.1)	50.53	(1.2)	8.39	(1.2)	<0.01	*
English	7.47	(<0.1)	19.56	(0.6)	12.09	(0.6)	<0.01	*
French (except Basque)	2.48	(<0.1)	5.17	(0.3)	2.68	(0.3)	<0.01	*
German	13.96	(<0.1)	19.93	(0.7)	5.98	(0.7)	<0.01	*
Irish	10.05	(<0.1)	17.40	(0.8)	7.35	(0.8)	<0.01	*
Italian	5.26	(<0.1)	6.64	(0.4)	1.38	(0.4)	<0.01	*
Polish	2.89	(<0.1)	4.49	(0.4)	1.60	(0.4)	<0.01	*
Black or African American	11.11	(<0.1)	11.95	(0.7)	0.84	(0.7)	1.00	
Black or African American, detailed	10.04	(<0.1)	9.75	(0.7)	-0.29	(0.7)	1.00	
African American	8.01	(<0.1)	8.15	(0.6)	0.13	(0.6)	1.00	
Ethiopian	0.08	(<0.1)	0.15	(0.1)	0.07	(0.1)	1.00	
Haitian	0.33	(<0.1)	0.28	(0.1)	-0.05	(0.1)	1.00	
Jamaican	0.35	(<0.1)	0.36	(0.1)	0.01	(0.1)	1.00	
Nigerian	0.12	(<0.1)	0.18	(0.1)	0.07	(0.1)	1.00	
Somali	0.05	(<0.1)	0.06	(0.1)	0.01	(0.1)	1.00	
Asian	5.56	(<0.1)	6.61	(0.4)	1.04	(0.4)	<0.01	*
Asian, detailed	5.35	(<0.1)	6.27	(0.4)	0.92	(0.4)	<0.01	*
Chinese	1.25	(<0.1)	1.62	(0.2)	0.38	(0.2)	0.25	
Filipino	0.93	(<0.1)	1.27	(0.2)	0.34	(0.2)	0.08	*
Asian Indian	1.06	(<0.1)	1.19	(0.2)	0.13	(0.2)	1.00	
Japanese	0.32	(<0.1)	0.34	(0.1)	0.01	(0.1)	1.00	
Korean	0.49	(<0.1)	0.52	(0.1)	0.03	(0.1)	1.00	
Vietnamese	0.53	(<0.1)	0.61	(0.1)	0.07	(0.1)	1.00	
AIAN	2.86	(<0.1)	2.72	(0.3)	-0.14	(0.3)	1.00	
NHPI	0.19	(<0.1)	0.29	(0.1)	0.09	(0.1)	1.00	
NHPI, detailed	0.17	(<0.1)	0.27	(0.1)	0.10	(0.1)	1.00	
Chamorro	0.01	(<0.1)	0.04	(<0.1)	0.03	(<0.1)	1.00	
Fijian	0.01	(<0.1)	-		-		-	
Marshallese	0.01	(<0.1)	-		-		-	
Native Hawaiian	0.06	(<0.1)	0.10	(<0.1)	0.04	(<0.1)	1.00	
Samoan	0.04	(<0.1)	0.03	(<0.1)	0.00	(<0.1)	1.00	
Tongan	0.01	(<0.1)	-		-		-	

(Table continued on the next page)

Major Category or Detailed Group	2016 ACS Production Ancestry (N=4,880,000)		2016 ACS Content Test, Race/Hispanic Origin (N=43,500)		Difference		Adjusted P-Value
	Estimate	Margin of Error	Estimate	Margin of Error	Estimate	Margin of Error	
SOR	1.27	(<0.1)	2.28	(0.3)	1.01	(0.3)	<0.01 *
SOR, detailed	1.27	(<0.1)	1.49	(0.2)	0.21	(0.2)	1.00
Belizean	0.02	(<0.1)	0.02	(<0.1)	0.00	(<0.1)	1.00
Brazilian	0.13	(<0.1)	0.24	(0.1)	0.11	(0.1)	1.00
Cabo Verdean	0.04	(<0.1)	0.06	(<0.1)	0.03	(<0.1)	1.00
Creole	0.02	(<0.1)	-		-		-
Guyanese	0.08	(<0.1)	0.05	(<0.1)	-0.03	(<0.1)	1.00
North American	0.01	(<0.1)	0.02	(<0.1)	0.02	(<0.1)	1.00
Hispanic	16.13	(<0.1)	17.66	(0.8)	1.53	(0.8)	0.09 *
Hispanic, detailed	14.28	(<0.1)	15.85	(0.8)	1.57	(0.8)	0.06 *
Cuban	0.61	(<0.1)	0.72	(0.1)	0.10	(0.1)	1.00
Columbian	0.32	(<0.1)	0.38	(0.1)	0.07	(0.1)	1.00
Dominican	0.57	(<0.1)	0.56	(0.1)	-0.01	(0.1)	1.00
Mexican	9.35	(<0.1)	10.36	(0.8)	1.01	(0.8)	1.00
Puerto Rican	1.33	(<0.1)	1.57	(0.2)	0.23	(0.2)	1.00
Salvadoran	0.66	(<0.1)	0.95	(0.2)	0.30	(0.2)	1.00

Sources: U. S. Census Bureau, 2016 American Community Survey 1-year Estimates and 2016 American Community Survey Content Test, Test Treatment.

Note: Margins of error are shown in parentheses. Significant at  $\alpha=0.1$  level. P-values with an asterisk (\*) indicate a significant difference between the two rates. P-values have been adjusted for multiple comparisons using the Holm-Bonferroni method. An '-' entry in the estimate column indicates that either no sample observations or too few sample observations were available to compute an estimate. Data for housing units in Alaska and Hawaii were excluded from the 2016 American Community Survey 1-year estimates in this table to make them comparable to the 2016 ACS Content Test data, which did not sample in these states. AIAN = American Indian and Alaskan Native, SOR = Some other race. DRB Approval Number CBDRB-FY19-RAGLIN-B0005.

Table A-12 mirrors Table 13 in Research Question 5, and compares the world regional distribution of detailed groups using test treatment data from the 2016 ACS Content Test. Findings are similar to those in the control treatment, with estimates of Western Europe; East, Southeast, and South Asia higher in the 2016 ACS Content Test than ACS production, and North American lower in the 2016 ACS Content Test. Additionally in the test treatment specifically, estimates in Eastern Europe and Central Asia, as well as Middle East and North Africa, were significantly different with groups having higher estimates in the test treatment of the 2016 ACS Content Test compared to ACS production data.



**Table A-12. Comparing Group Estimates Derived from Race/Hispanic Origin and Ancestry, by World Region**

Ancestry or Detailed Race/Ethnicity Group	2016 ACS Production Ancestry (N=4,880,000)		2016 ACS Content Test, Race/Hispanic Origin (N=43,500)		Difference		Adjusted P-Value	
	Rate	Margin of Error	Rate	Margin of Error	Rate	Margin of Error	Value	Significance
Western Europe	36.43	(0.1)	46.50	(1.1)	10.08	(1.1)	<0.01	*
Eastern Europe and Central Asia	5.97	(<0.1)	7.29	(0.5)	1.32	(0.5)	<0.01	*
Mexico, Central America, and South America	12.76	(<0.1)	13.75	(0.9)	0.99	(0.9)	0.19	
Caribbean	3.36	(<0.1)	3.53	(0.3)	0.17	(0.3)	0.55	
Middle East and North Africa	1.07	(<0.1)	1.54	(0.3)	0.47	(0.3)	0.03	*
Sub-Saharan Africa	1.09	(<0.1)	0.96	(0.2)	-0.13	(0.2)	0.55	
East, Southeast, and South Asia	5.34	(<0.1)	6.22	(0.4)	0.87	(0.4)	<0.01	*
Pacific Islands	0.23	(<0.1)	0.36	(0.1)	0.13	(0.1)	0.19	
North America (except the U.S., Mexico, and Caribbean)	0.91	(<0.1)	0.55	(0.1)	-0.36	(0.1)	<0.01	*

Sources: U. S. Census Bureau, 2016 American Community Survey 1-year Estimates and 2016 American Community Survey Content Test, Test Treatment.

Notes: Margins of error are shown in parentheses. Significant at  $\alpha=0.1$  level. P-values with an asterisk (\*) indicate a significant difference between the two rates. P-values have been adjusted for multiple comparisons using the Holm-Bonferroni method. Data for housing units in Alaska and Hawaii were excluded from the 2016 American Community Survey 1-year estimates in this table to make them comparable to the 2016 ACS Content Test data, which did not sample in these states. DRB Approval Number CBDRB-FY19-RAGLIN-B0005.

## Appendix G. Limitations in Identifying Matches Between Race and Ancestry

This research examined how often a person’s ancestry data was an exact match with what they had reported as their race and Hispanic origin. The following tables demonstrate some of the imperfections around categorizing data as a match for this analysis.

To understand how ancestry data is compared to race data, we have to understand how write-in responses are handled for each question. Each write-in response, whether from race or from ancestry, is assigned a three-character code during data processing. Different write-ins can result in the same code; for instance, a response under Race of “ITALIAN,” “BOLOGNESE,” “ROMAN,” or “UMBRIAN” will result in the same three-character code (in this case, 079). However, if those same responses were provided under the Ancestry question (and if four ancestry codes could be assigned instead of the current maximum of two), they would result in four different codes (in this case, 051, 058, 059 and 071) to more specifically identify a person’s heritage. Note that not all countries get nuanced ancestry codes for sub-geographies. For instance, all Mexican states are assigned the same ancestry code but most Indian states are assigned a distinct code.

For this analysis, each ancestry code was associated with one and only one race/ethnicity code that was considered to be the corresponding match (though an ancestry code could also match to a race/Hispanic origin checkbox). Only the codes were compared, based on the mapping, without regard to the original written text. An ancestry code of 051 would match to a reported race code of 079, even if the actual ancestry response was “ROMAN” and the race response was “ITALIAN”.

The 1-to-1 mapping rule has some limitations. For instance, the table below shows what happens when respondents answer each question (Race or Ancestry) with either “BIRACIAL”, “MIXED”, or “MULTIRACIAL”. Since those descriptions are not critical to delineate from an ancestry perspective, they all receive the same ancestry code (995). However, under the Race question, they result in three different codes. Again, a given ancestry code can only match to one race code; here, ancestry code 995 only matches to race code 723. Thus, as shown in Table A-14 below, one result of the 1:1 mapping rule is that if a person wrote “MULTIRACIAL” under both Race and Ancestry, that ancestry response would not be considered to have a match in the race data.

**Table A-13. Matching Results for Select Multiracial Responses**

<b>Ancestry write-in (respondent-provided)</b>	<b>Ancestry code (Census assigned)</b>	<b>Race Response (respondent-provided)</b>	<b>Race code (Census assigned)</b>	<b>Consistent Match between Ancestry and Race?</b>
MIXED	995	Write-in of MIXED	723	Yes
MULTIRACIAL	995	Write-in of MULTIRACIAL	725	No
BIRACIAL	995	Write-in of BIRACIAL	719	No
MULTIRACIAL	995	Write-in of MIXED	723	Yes

This example shows the imperfect nature of the mapping rule used in this research. The mapping rule does provide a baseline, conservative assessment of consistency. The majority of ancestry codes and race codes are compatible to a 1:1 matching but additional instances of unexpected inconsistency are discussed below.

There are over a thousand distinct Race codes to delineate the many American Indian tribes. However, when written under the Ancestry questions, the majority of those tribes get coded under the same singular generic Ancestry code (920), which is only considered to be a match to a singular Race code (M41).

**Table A-14. Matching Results for Select AIAN Responses**

<b>Ancestry write-in (respondent- provided)</b>	<b>Ancestry code (Census assigned)</b>	<b>Race Response (respondent-provided)</b>	<b>Race code (Census assigned)</b>	<b>Consistent Match between Ancestry and Race?</b>
NATIVE AMERICAN	917	NATIVE AMERICAN write-in	M41	Yes
CHEROKEE	919	CHEROKEE write-in	B21	Yes
ALASKA NATIVE	920	ALASKA NATIVE write-in	M47	No
CHOCTAW	920	CHOCTAW write-in	C08	No
SIOUX	920	SIOUX write-in	K36	No
SIOUX	920	AMERICAN INDIAN write-in	M41	Yes
AMERICAN INDIAN	920	AMERICAN INDIAN write-in	M41	Yes
AMERICAN INDIAN	920	American Indian or Alaska Native checkbox		Yes
NATIVE AMERICAN	917	American Indian or Alaska Native checkbox		No

The process to compare Race and Ancestry data also underestimates the consistency of Hispanic origin responses in certain cases as shown in Table A-3. An Ancestry response of “MEXICAN AMERICAN” yields a single, distinct code which is different from a response of “MEXICAN”; as a result, if a person marks the Mexican checkbox in Hispanic origin, then writes-in “AMERICAN” under Some Other Race, then writes “MEXICAN AMERICAN” in Ancestry, the algorithm would conclude their responses were not consistent. However, if they had written “AMERICAN MEXICAN” under Ancestry, then their responses would be consistent. Additionally, an Ancestry response of “MEXICO” results in a different ancestry code (218) than a response of “MEXICAN” (210) and only “MEXICAN” is crosswalked to correspond to the Mexican checkbox in the Hispanic origin question.

**Table A-15. Matching Results for Select Hispanic Responses**

Ancestry write-in (respondent-provided)	Ancestry code (Census assigned)	Hispanic Origin (HO) response (respondent-provided)	Race/Hispanic code (Census assigned)	Consistent Match between Ancestry and HO?
AMERICAN	939	Yes, Mexican, Mexican Am., Chicano checkbox + Write-in of AMERICAN	995	Yes
MEXICAN	210	Yes, Mexican, Mexican Am., Chicano checkbox + Write-in of AMERICAN	995	Yes
AMERICAN MEXICAN	939 210	Yes, Mexican, Mexican Am., Chicano checkbox + Write-in of AMERICAN	995	Yes
MEXICAN AMERICAN	211	Yes, Mexican, Mexican Am., Chicano checkbox + Write-in of AMERICAN	995	No
MEXICAN	210	Yes, Mexican, Mexican Am., Chicano checkbox		Yes
MEXICANO	212	Yes, Mexican, Mexican Am., Chicano checkbox		No
MEXICO	218	Yes, Mexican, Mexican Am., Chicano checkbox		No
JALISCO MEXICO	218 218	Yes, Mexican, Mexican Am., Chicano checkbox		No
MEXICAN OAXACAN	210 218	Yes, Mexican, Mexican Am., Chicano checkbox		Half match
HISPANIC	290	Yes, another Hispanic, Latino, or Spanish origin checkbox		No
HISPANIC	290	Yes, another Hispanic, Latino, or Spanish origin checkbox + Write-in of HISPANIC	273	Yes
LATINO	252	Yes, Mexican, Mexican Am., Chicano checkbox		No

The following tables show limitations of the matching algorithm for ancestries that are classified as White or Black. These ancestries are the ones currently published in data products such as AFF Table B04006.

**Table A-16. Matching Results for Select White Responses**

<b>Ancestry write-in (respondent-provided)</b>	<b>Ancestry code (Census assigned, max of 2)</b>	<b>Race Response (respondent-provided)</b>	<b>Race code (Census assigned)</b>	<b>Consistent Match of Ancestry to Race?</b>
IRISH GERMAN SCOTTISH	050 032	IRISH GERMAN SCOTTISH	078 065 119	YES
IRISH SCOTTISH	094	SCOTTISH IRISH	118	NO
IRISH SCOTTISH	094	IRISH SCOTTISH	078	YES
SCOTTISH IRISH	087	IRISH SCOTTISH	078	NO
SCOTTISH IRISH	087	SCOTTISH IRISH	118	YES
ENGLISH FRENCH CANADIAN	022 935	ENGLISH FRENCH CANADIAN	045 059 183	Only English
FRENCH ENGLISH CANADIAN	026 022	ENGLISH FRENCH CANADIAN	059 045 183	French & English
CANADIAN AND FRENCH	935	CANADIAN AND FRENCH	183 059	NO
FRENCH CANADIAN	935	FRENCH CANADIAN	184	YES
CAJUN FRENCH	935	CAJUN FRENCH	182 059	NO
CAJUN	937	CAJUN	182	YES
SCANDINAVIAN	098	SCANDINAVIAN	115	NO
SCANDINAVIAN	098	NORDIC	116	YES
NORDIC	098	NORDIC	116	YES
EASTERN EUROPEAN	190	EASTERN EUROPEAN	044	NO
EASTERN EUROPEAN	190	EUROPEAN	047	YES
WESTERN EUROPEAN	187	WESTERN EUROPEAN	140	YES
WESTERN EUROPEAN	187	EUROPEAN	047	NO
EUROPEAN	195	EUROPEAN	047	YES
CAUCASIAN	924	CAUCASIAN + did not mark WHITE checkbox	188	NO
WHITE	924	WHITE	008	YES
CAUCASIAN	924	WHITE	008	YES
CAUCASIAN	924	CAUCASIAN + marked White checkbox	188	YES

Table A-16 shows the following nuances in the coding process:

- The order in which a person writes their response can impact the way it is coded, and thus the matching results. “IRISH SCOTTISH,” “SCOTTISH IRISH,” and “IRISH – other ancestry – SCOTTISH” all lead to different interpretations of a person’s ancestry.
- The identity of “FRENCH CANADIAN” is coded more often from Ancestry data than race data (see Table 11). Correctly interpreting a two-word response as one ancestry that is made of two words (for instance, French Canadian or Scotch Irish) is one difficulty in the coding process.
- “SCANDINAVIAN” and “NORDIC” as Ancestry write-ins receive the same code (098) but as Race write-ins receive different codes (115 and 116, respectively); the ancestry code of 098 is only crosswalked to 116.
- “EASTERN EUROPEAN,” “WESTERN EUROPEAN,” and “EUROPEAN” as Ancestry write-ins receive different codes (190, 187, and 195 respectively) and as Race write-ins receive different codes, but they are crosswalked in an unexpected way. An Ancestry write-in of “EASTERN EUROPEAN” will only crosswalk to a Race write-in of “EUROPEAN” not to a Race write-in of “EASTERN EUROPEAN.” However, an Ancestry write-in of “WESTERN EUROPEAN” will crosswalk to a Race write-in of “WESTERN EUROPEAN.” The Race code of 044 (Eastern European) does not have an Ancestry code mapped to it at all.

An Ancestry write-in of “CAUCASIAN” will crosswalk either to the White checkbox or to a Race write-in of “WHITE” but will not crosswalk to a race write-in of “CAUCASIAN.”

As with White ancestries, Black ancestries also have some unexpected mappings with Race write-ins. For instance, the responses of “AFRICAN”, “WEST AFRICAN”, and “EAST AFRICAN” receive distinct codes from the ancestry question (599, 598, 597 respectively) but share the same code from the race question (320). However, occasionally the race codes are more nuanced than ancestry; this is seen in the coding of West Indies responses such as “CARIBBEAN”, “TAINO INDIAN” and “WEST INDIAN”.

**Table A-17. Matching Results for Select Black Responses**

<b>Ancestry write-in (respondent-provided)</b>	<b>Ancestry code (Census assigned)</b>	<b>Race Response (respondent-provided)</b>	<b>Race code (Census assigned)</b>	<b>Consistent Match between Ancestry and Race?</b>
EAST AFRICAN	597	AFRICAN	320	YES
EAST AFRICAN	597	EAST AFRICAN	320	YES
EAST AFRICAN	597	WEST AFRICAN	320	YES
WEST AFRICAN	598	EAST AFRICAN	320	YES
WEST AFRICAN	598	AFRICAN AMERICAN	311	NO
AFRICAN	599	AFRICAN	320	YES
AFRICAN	599	AFRICAN AMERICAN	311	NO
AA	996 (uncodable)	AFRICAN AMERICAN + Black checkbox	311	NO
BRITISH WEST INDIES	011 335	BRITISH....WEST INDIAN	025 395	YES, 335:395 & 011:025
BRITISH WEST INDIES	011 335	DOMINICAN	244	NO
BWI	322	DOMINICAN	244	NO
BWI	322	BRITISH WEST INDIAN	395	YES, 322:395
BLACK DUTCH	310	AFRICAN AMERICAN	311	NO
WEST INDIAN	335	WEST INDIAN	395	YES, 335:395
CARIBBEAN	335	CARIBBEAN	728	NO
TAINO INDIAN	335	TAINO INDIAN	242	NO

As with the White ancestries, we provide these examples to illustrate how our results are a conservative estimate of the number of times that an ancestry response was found to also have been provided in race and Hispanic origin.