

COLLECTION OF ANCESTRY DATA IN CENSUS SURVEYS

INTRODUCTION

The Census Bureau will include a question on ancestry on the 1990 census questionnaire. The Census Bureau has received strong support from various individuals and organizations to include a question on ancestry and parental birthplace in the 1990 census. Census staff reviewed analyses of recent test results and data needs for both questions, as well as consulted with a variety of data users to decide which question to ask in 1990. The Census Bureau had to tell Congress which topics would be included in the 1990 census by April 1, 1987.

Overall, the 1980 ancestry question performed satisfactorily. However, criticism was received on several aspects of the coding, classification, and tabulation schemes developed for the ancestry data. A few individuals or groups consider some aspects of the data as problems, while others find them acceptable. We will discuss the 1980 census ancestry data faults and merits in this paper.

This is the first attempt to assess the collection, tabulation, and presentation methods used for the 1980 ancestry information examines several issues on ancestry. Also, background information on census experiences with ethnic data prior to 1980, experiences with 1980 census ancestry data, including the processing of the data and planning for the 1990 census ancestry data will be discussed in this paper. The analysis of the procedures used in 1980 guides the Census Bureau as it begins to consider refinements to the ancestry item for the 1990 census and future surveys.

EXPERIENCE WITH ETHNIC DATA PRIOR TO 1980

Survey Work

Decennial Censuses - The Census Bureau first collected information on place of birth of individuals in the 1850 census and parental birthplace in the 1870 census, providing information on the first and second generations. The time series of questions on person's and parents' birthplace in Decennial Censuses ended in 1980.

Current Population Survey - The November 1969 Current Population Survey (CPS) included an ethnic question for the first time. Interviewers used flashcards with a number of specific origins to obtain the data. Responses were based on self-identification. The CPS has annually included a question on ethnic origin ever since.

Self-identification approach is subject to misinterpretation(s) and or inconsistency in reporting. For instance, an evaluation of the CPS ethnic origin responses using prelisted categories in two consecutive annual surveys showed that only about two-thirds of the population were consistent in reporting the same origin. Also, inconsistency was most likely to occur among certain groups such as English, French, and Irish.¹ The level of inconsistency was one of the factors which lead the Bureau to test and implement the open-ended ancestry question. Also, ethnic studies have shown that the affiliation with an ethnic group may not be as strong among some Northern and Western European groups as it is among other groups in the U.S.

Also about one-third of the respondents in the CPS did not report any ethnic group listed on the flashcard. Of these, two-thirds reported multiple origin and about one-tenth reported "American."

Current Population Survey (cont.) - A question on ancestry, identical to that being used in the 1980 census, was asked in the November 1979 CPS. The question was worded "What is's ancestry?" Again, a flashcard was used to select ethnic origin group. The November 1979 CPS was used to prepare for the 1980 census and to serve as a bridge between the CPS and 1980 census.

(add findings on the CPS)

1980 Pretests and Dress Rehearsal - Mike has part

Mike has this page

1980 Pretests and Dress Rehearsal - The first 1980 pretest was held in Oakland, California in

The National Content Test -- allowed persons to report 1 or 2 origins.

It was found that allowing persons to report more than one origin groups does not have an impact on the nonresponse rate or on those who mark an other category.

The 1978 Dress Rehearsal was conducted in Richmond, Virginia

48 percent of Blacks did not report an ancestry

23 percent of Whites did not report an ancestry

Lower Manhattan

Nonresponse rate fell for Blacks to 27 percent and Whites rate was 22 percent

1980 Census Meetings on Ethnicity

1976 White House Meeting - In June 1976, the White House Office of Public Liaison in cooperation with the "Ethnic Millions Political Action Committee" sponsored a meeting on ethnicity and the 1980 census. About 60 ethnic leaders were in attendance. The leaders represented diverse groups such as Armenian, Hungarian, Asian, Ukrainian, Slovene, Danish, Bulgarian, French, Lithuanian, Italian, Russian, Greek, German, Byelorussian, Nationality Services, etc.

The participants recommended that a question on ethnicity based on self-identification be included in the 1980 census. Also, since the proportion of the population which is of foreign birth or parentage is decreasing, more attention is being focused on the third or later generations who retain strong identification; therefore, it was recommended that information on the ethnicity of all generations be collected in the 1980 census. The meeting attendees proposed collection of data for on all generations and that data on broad ethnic groups such as Eastern European be accepted in the 1980 census.

1976 Census Meeting on Ethnicity - A meeting at the Census Bureau Washington Headquarters obtained advice on the question wording, groups to include as examples and methods to obtain better reporting from respondents in the 1980 census. About 12 ethnic representatives attended the meeting.

The discussion focused on the difficulty in measuring ethnicity, respondent burden, and questionnaire space constraints. In general, it was agreed that ethnic heritage is difficult to measure. The ethnic experts exchanged views on the method of collecting the data and the terminology for the ethnic question.

1976 Census Meeting on Ethnicity (cont.) - The participants suggested using an open-ended ancestry question with examples provided. Also, the preferred terminology for the question or key words was "origin or ancestry." In addition, the advisors asked that the question permit multiple origin reporting.

1980 Meeting on Ancestry - The Census Bureau held a meeting with ethnic advisors to discuss issues regarding the edit and allocation and publication of ancestry information in the 1980 census.

The risks involved in using related information such as birthplace, language, and relationship to allocate ancestry data outweighed the 1980 ancestry. The ancestry data should be edited for entries which are 2 origin groups but classified as a single group; That is, for an entry of "French Canadian" the code should be the the single group "French Canadian" and not a multiple group of "French" and "Canadian."

Other Meetings - In addition to the specific meetings noted above, the Census Bureau held several meetings with persons knowledgeable in the field of ethnicity to discuss the publication plan for the 1980 ancestry data. Advisors suggested that data in standard reports, such as Volume I, Census Tract, and Congressional District reports show data on 15 single ancestry groups and 6 selected multiple groups which would be open-ended; that is, "English and other groups," "Polish and other groups," "French and other groups," etc.

1980 CENSUS EXPERIENCE WITH ETHNICITY

The Ancestry Question - A general question on ancestry (ethnicity) was asked for the first time in a decennial census in 1980. The data on ancestry were derived from answers to question 14 on the long-form sample questionnaire. (See facsimile of question and respondents instructions below.) The question was based on self-identification and was open-ended with space for a write-in entry. Ancestry referred to a person's "nationality group, lineage, or the country in which the person or person's parents or ancestors were born before their arrival in the United States." Thus, persons reported their ancestry group regardless of the number of generations removed from their country of origin. Responses to the question reflected the ethnic groups with which persons identified and not necessarily the degree of attachment or association the person had with the particular ethnic group(s).

Insert here

The Ancestry Question (cont.) - Respondents were instructed to be as specific as possible in reporting ancestry. Nevertheless, some persons provided categories such as "Slavic," "European," and "African" which encompass different ancestry groups.

The general classification of ancestry groups and coding scheme was determined for the first time in 1980. The classification scheme was complicated by the fact that some persons reported specific regions in a country while others simply named a country. Also, changes in political definitions of geographical areas have occurred overtime and some migration groups affected responses. Some ethnic groups traced their ethnic roots to several different nations, sometimes because of the geographical changes in boundaries or migration to different countries over time. Finally, some groups could not name a specific country since the period of arrival was too long ago or because their cultures were stripped away by slavery or because the circumstances under which they migrated made them uncomfortable about naming their country of origin.

Ancestry data were more difficult to collect than other social and economic information because of the lack of clear-cut definitions, changing terminologies, uncertain reliability, and lack of knowledge of the degree of affiliation with a group. Indirect measures, such as own birthplace, parental birthplace, and mother tongue help estimate ethnicity because they are less susceptible to changes in reporting between censuses. A direct question on ancestry is subject more to changes in interpretation but yields relatively consistent information if the criteria for inclusion in a particular group were established and the intent of the question is clear. However, it is impossible to establish a criteria

for inclusion in a group since ethnicity may be defined in many ways -- for example, some people may define ethnicity by geographic or ancestral origin; or language; religion; intermarriage; shared physical characteristics; race; cultural beliefs; behaviors, etc.

Coding Ancestry Data - An ancestry code list was created for the processing of the 1980 census ancestry responses. The Census Bureau developed a preliminary 1980 ancestry code list from several resources including the Harvard Encyclopedia of American Ethnic Groups, Outline of World Cultures, by George P. Murdock; Webster's New Collegiate, Geographical, and New International Dictionaries, and Encyclopedia Britannica.

Persons knowledgeable in different aspects of ethnicity reviewed the draft list for accuracy, consistency, and completeness. On the basis of the review of comments received from the experts, preliminary 1980 census results, and additional research, a final list of ancestries was used for the 1980 census coding.

The 1980 census collected information on single and multiple-ancestry groups. All reported single and double ancestry groups was coded. In addition, 17 previously identified triple-ancestry groups were given unique codes, but were coded regardless of the order in which they were reported. Only the first and second reported ancestry groups of all other multiple origin were coded. Since a large proportion of multiple origins were reported in the previous census surveys, the decision was made to code the first 2 groups reported, except for the cases of the 17 triple groups.

Coding Ancestry Data (cont.) Responses which occurred during the coding which were not on the ancestry code list were referred to the Ethnic and Spanish Statistics Branch members in Washington, D.C. Research and consultation determined appropriate codes for these responses. These responses were added for the general coding operation.

Edit and Allocation of Ancestry Data - A panel of advisors brought together to consider allocating ancestry recommended that the ancestry responses not be allocated because other characteristics -- race, birthplace, language, and ancestry of other individuals -- would not provide accurate information for assigning ancestry. But the panel of advisors recommended that some ancestry data be edited so that when entries such as "French Canadian" were coded as two origin groups, the appropriate single ancestry group of "French Canadian" was recorded.

Tabulation of Ancestry Data - The tabulation of ancestry data was complicated, because persons reporting multiple ancestry groups were included in more than one category in statistical tables. For example, a person reported as "English-Irish" is counted in both the multiple categories "English" and "Irish." Persons reporting one of the unique triple origin groups such as English-Irish-German were tabulated in each of the three ancestry categories. Since persons who reported more than one ancestry were included in more than one group in the tables, the sum of persons reporting at least one specific ancestry or a multiple ancestry was greater than the total population.

In the tabulation of ancestry data, some responses were coded as 2 groups but were tabulated as single ancestry groups. For example, Irish American was coded as a 2 origin category but tallied as a single response of Irish in data products. Other entries when combined with another ancestry group were also tallied as single entries. For instance, entries of White, U.S., Christian, etc. were tallied as the other single ancestry response reported with that response (i.e., French - U.S., White English, etc.).

1980 CENSUS CLASSIFICATION OF ANCESTRY GROUPS

The Classification Scheme Since the 1980 census was the first attempt at collecting ancestry data, problems in the classification of groups were expected, especially because advisors sometimes did not agree on the assignment of groups. Data on 140 ancestry groups were shown in the Supplementary Report, "Ancestry of the Population by State: 1980," PC80-S1-10. The Summary Tape File 4 contained data on 188 ancestry categories. Estimates on each code used in the coding operation for the 1980 census, which is about 600 codes, are available to users by request for a nominal fee. Paul R. Magocsi analyzed the Census Bureau's classification of ancestry groups in a recent article "Are the Armenians Really Russians? -- Or How the U.S. Census Bureau Classifies America's Ethnic Groups." He found that the "classification of the responses must be seriously reconsidered and revised, since many distinct ethnic groups were subordinated to other ancestry group classifications and therefore statistically ceased to exist" (1987:133). And "Having qualified its own classification system, it seems most appropriate at this mid-decade juncture that the Census Bureau be informed as to how its decisions seem to reflect less questions of interpretation than a lack of sensitivity or even unawareness of the existence of certain ethnic groups" (1987:137).

Paul Magocsi's paper is a contribution to the study of ancestry in the United States and of how the Census Bureau might better collect ancestry data. However, as Magocsi noted, "while it is easy to criticize the ancestry group classifications arrived by the Census Bureau, it is more difficult to provide concrete suggestions for emendations. Emendations, however do seem particularly necessary for ancestries deriving from European ethnic groups which form by far the largest proportion of the American population" (1987:137). Here we will discuss several of Magocsi's concerns and provide background information on the Census Bureau's action on

these concerns and possible solutions to collecting more accurate ancestry data in the 1990 census. Given these caveats, in the remaining sections we will look at how the 1980 classification scheme may be modified for the 1980 census.

As noted earlier, the Census Bureau consulted with persons knowledgeable in the field of ethnicity in preparing the coding, classification, and tabulation of ancestry data for the 1980 census. However, ancestry data seem to be a controversial issue on how groups should be classified as experts have differing views on the categorization of groups.

Armenians The 212,621 Armenians were coded within the Russian coding scheme (code number 159), but were tabulated in the category for Northern African and Middle Easterner in the Supplementary Report on ancestry, "reflecting the fact that most ancestors of America's Armenian population came from what is now Turkey rather than from traditional Armenian lands in the Soviet Union" (USCB:1987:9).

needs
work
on

British - Some respondents probably felt that British was an appropriate response to the question "What is's ancestry?" Others may have been third or later generation individuals whose ancestors had intermarried to a significant extent, and therefore could not or would not chose a specific "British" ancestry group(s). Others may not have understood that the most specific identification was expected. Expanded instructions to respondents are unlikely to eliminate this kind of response. The Census Bureau must group according to the actual responses. The general categories will continue to collect this type of response.

French.

The following categories were obtained for French ancestry:

<u>Code</u>	<u>Category</u>	<u>Number</u>
028	French.....	11,178,943
029	Breton.....	7,953
030	Corsican.....	19,824
031	French Basque.....	12,318
090	Alsatian.....	41,444
091	Lorraine.....	946
098	Acadian.....	32,799

Here we have the problem of relatively small numbers. Only a few persons selected Lorraine, so the total is relatively unreliable. Also, although Magocsi notes that "the Census Bureau, however, has singled out the Alsatians as a distinct ancestry group classification, even though in the United States they have shown little initiative in creating their own organizations and community life, preferring to associate with French or German communities" (1987:140), the Harvard Encyclopedia of American Ethnic Groups classifies Alsatian separately; the Census Bureau used this source for its selection in 1980.

Georgian - Magocsi feels very strongly about inclusion of Georgian "Not only should their ancestry group classification be restored, but the sub-heading Georgia should be dropped. Americans of Georgian ethnic background might use the adjective Georgian or the native name of their country Gruziia (or gruzinets - a Georgian person), but they are not likely to use the Anglicized form of their homeland, Georgia, to describe themselves" (1987:144). The 1980 census estimated 1,505 Georgians in the Census (on the basis of the sample tabulations). Of these, a large majority were reported as living in the U.S. state of Georgia. The question then, is how the Census Bureau can differentiate persons who feel a fondness for the U.S. state of Georgia and report themselves as Georgian, from those who report their ancestry as Georgia in the Soviet Union. Perhaps if the Census Bureau used "Gruziia" rather than Georgia, appropriate data will be obtained. The Census Bureau is aware of this problem, and will try to resolve the reporting of Georgians in Georgia during data capture.

Czechoslovakian The Census Bureau coded two groups for Czechoslovakia in 1980; Czech with an estimated 1,311,856 persons and Bohemian, Moravian an estimated 589,906 persons. In 1980, persons responding with Czech were combined with persons selecting Czechoslovakian, whereas Slovak was a separate category. Czech was included with Czechoslovakian erroneously because "Czech" is often reported as the abbreviation for Czechoslovakian. However, Czech and Czechoslovakian should be separate categories. In the 1990 census, data on these groups will be shown separately. Bohemian and Moravian will be coded separately since they form a sufficiently large group together.

Jews - Persons responding Jewish, Moslem, Protestant, or Catholic were tabulated in a category with other unclassifiable responses (although with a separate general code for all other residual entries not listed on the code list). Jewish cannot be tabulated separately because of the law 13 U.S.C. 221(c) which forbids the collection of information on religion in a mandatory census. In the ancestry item, religious groups and non-classifiable responses were assigned the same code during processing.

Magocsi argues for inclusion of Jewish as an ancestry category; however, inclusion is not possible. The Census Bureau included the statement "A religious group should not be reported as a person's ancestry" in its instructions for the ancestry question. Also, there is an argument that Judaism may not be merely a religious designation, but, in fact, may reflect "a cultural framework that in the old country even included one's own literary language" (Magocsi 1987:147).

Although "Jewish" responses could be coded since the information voluntarily offered need not necessarily be refused, the data would be skewed since

Jews (cont.) - instructions tell respondents not to include a religious response. Jewish still refers primarily to a religious group and therefore cannot be tabulated separately in the census. The appropriateness of including a question on religion in 1990 may jeopardize the census, and believe it paramount to maintain the position on religion as done in previous censuses.

Scotch-Irish - Scotch-Irish is an ethnic group from Northern Ireland. The response "Scotch-Irish" may refer to the unique single ancestry group from Northern Ireland or to the multiple ancestry group (both Scotch and Irish ancestry). Since the unique single ancestry response could not be distinguished from the multiple response, Scotch-Irish was treated as a multiple origin group and tabulated in both "Scotch" and "Irish." The Census Bureau had planned to publish data on the unique group of "Scotch-Irish" in the PC80-2 subject reports. However, this report was not published due to budgetary restraints. The estimate was recently tabulated however, with 3,005,500 persons reported as Scotch-Irish ancestry (i.e., Scotch was the first ancestry and Irish was the second ancestry reported). Nevertheless, no one knows if respondents intended to report a single response or a multiple-ancestry response.

Scotch-Irish (cont.) - For the 1990 census, the Bureau will review the policy on collecting data on responses of "Scotch-Irish." If the 1990 coding system can identify the unique single entry of Scotch-Irish without posing a problem in the data processing, the Census Bureau will collect data on this unique group in the 1990 census.

Russians - The Census Bureau felt that particular attention was required regarding the ethnic groups which originated from the areas now comprising the Union of the Soviet Socialist Republics or which were once part of the Russian Empire and the Austro-Hungarian Empire. Some of the ethnic groups are still distinct, while others seem to have merged. "For example, a number of respondents reported the category Ruthenian and although some advisors still consider them a distinct group, others feel Ruthenians should be combined with Ukrainians" (USBC 1983:7).

"The response Russian presents another difficulty because it can have several meanings. Russian may mean Great Russian, or for some groups such as Russian Jews, Russian Moslems, Belorussians, and in some cases Ukrainians, it may simply refer to their country of origin. Hence, persons reporting Russian in the census may include those who identified with the specific ethnic group "Russian" as well as those who may have reported the term in a different context" (USBC 1983:7). Magocsi feels that "this is not only a understatement but a euphemistic formulation that ... barely camouflages the fact that the Census Bureau's designation rendered in the published data as "Russian, not elsewhere classified" is meaningless" (1987:141).

For 1990, the Census Bureau will review the Russian designation and separate some groups into distinct categories. In general, the grouping of responses and the terminology of Russian ancestries is complex and controversial.

Summary - The ancestry groups which were specifically questioned by Paul Magocsi and others, such as, Portuguese and Basque and those mentioned above will be reviewed again for the 1990 census.

DATA PRODUCTS ON ANCESTRY

Standard 1980 Census Reports - Since this was the first time in a decennial census information on ancestry groups was published, the Census Bureau sought advice on the data presentation. The advisors recommended showing data on specific single ancestry groups and also multiple groups such as "English and other groups," which is an open-ended multiple category.

In the standard decennial census publications, data on 15 single ancestry groups and 6 multiple ancestry groups were displayed. Information on these ancestry groups was published for states, metropolitan areas, places of 2,500 or more, and counties, tracts, and congressional districts. Social, economic, and demographic information was provided in the Volume I, Chapter C reports for the 15 single and 6 multiple origin groups.

Special Reports - The list of groups for the Supplementary Report, "Ancestry of the Population by State: 1980," was developed before the 1980 census data were seen. This was a difficult task; the groups which were selected for the tabulation were those which were expected to be larger. The special tabulation which provided the information for the Supplementary Report on ancestry included 209 ancestry categories (including specific 3-origin groups). In the report, estimates for 149 ancestry groups are shown. Only groups with less than 5,000 persons and a few groups where the estimates seemed unreasonable were not published in the report.

Special Reports (cont.) - the series. Interest in the report was voiced by many data users, such as Senator Moynihan, the National Center for Urban Affairs, and many ethnic organizations. The publication of this report was given high priority to expedite the release of the data; in fact, the report was published within 6 months of delivery time of the initial data.

Preliminary table outlines were sent to ethnic advisors for review. Comments from the reviewers were examined and reconciled when differences occurred and staff performed additional research in areas of concern. The final table outlines, including the geographical/cultural designation headings and specific ancestry groups, were discussed with the reviewers to resolve any concern regarding the data presentation that census would be using for the report.

The eight geographical/cultural designations used in the Supplementary Report on ancestry were (1) European (excluding Spaniard), (2) North African and Middle Easterner, (3) Sub Sahara African, (4) Asian (excluding Middle Easterner), (5) Non-Spanish Caribbean, Central and South American, (6) Spanish, (7) Pacific, and (8) North American.

During the preparation of the analytical text and definitions and explanations descriptions in the report, advice was sought from persons knowledgeable in the study of ethnicity.

The estimates for ancestry groups were reviewed for reasonableness. Census staff used the 1970 data on native of foreign or mixed parentage and foreign born, ethnic estimates provided in the Harvard Encyclopedia of American Ethnic Groups, data on ethnic origin from the National Opinion Research Center, ancestry data

Special Reports (cont.) - from the November 1979 Current Population Survey (CPS), and data on immigrants from the Immigration and Naturalization Service. Also, staff conferred with ethnic experts on some 1980 census estimates on ancestry groups. Consultations were made with specialists in each cultural groups (Asian, African, Caribbean, European, etc.) regarding the classification and designations to use for ancestry categories in reports. The staff performed a detailed review and evaluation of the 1980 ancestry data before it were published.

The findings of the review were that the majority of the 1980 ancestry information were accurate and reasonable. However, when evaluation of the ancestry data from the 1980 census was compared with that of the November 1979 CPS, differences for some ancestry groups were greater than expected. Both these sources used essentially the same open-ended ancestry question, although data collection, sample size, population coverage, and processing procedures differed. However, the differences in some of the results between the sources were much greater than one might expect, even allowing for the factors noted above.

The review of 1980 census data showed that reporting and/or processing problems may have affected the data for a few ancestry groups. For instance, the number of Dutch West Indians was unexpectedly high in Oklahoma and Texas, as were the Georgians (Russian origin) in Georgia and Aleuts in Hawaii. Overall, the ancestry data proved to be reliable and reasonable. See the "Explantions" section of the Supplementary Report, "Ancestry of the Population by State: 1980," PC80-S1-10, for more information on the differences between ancestry groups in the 1980 census and the November 1979 CPS and other data problems in 1980.

Computer Tape Files - Ancestry data may be found on the 1980 census Summary Tape Files 3 and 4. Also, information can be obtained from the Public Use Microdata File and the Sample Edited Detail File. Summary Tape File (STF) 3 provides estimates on 19 ancestry categories for the U.S. states, metropolitan areas, counties, places of 10,000 or more persons, and congressional districts. The STF 4 provides estimates on 188 ancestry categories for similar areas available from STF 3. From the Microdata File, users can obtain information on any ancestry category even individual codes; information can be cross tabulated for social, economic, and demographic characteristics. The Microdata File is a sample of the long form questionnaires. There are 3 subfiles in the Microdata File, two 1-percent files and one 5-percent file. From the Sample Edited Detailed File, information on any ancestry group for all characteristics on the long form census questionnaire can be obtained.

Editals - A tally of the individual ancestry codes which occurred in each state during the 1980 census has been tabulated. Information was summed for all states for each code and national estimates are available from the Census Bureau for a nominal fee. Users interested in obtaining these data should contact the Ethnic and Spanish Statistics Branch, Population Division, Census Bureau Room 2340, Building 3, Washington, D.C. 20233 or telephone (301) 763-7571.