

DUP?

CLASSIFICATION OF ANCESTRY GROUPS IN DECENNIAL CENSUS

Michael J. Levin

and

Nancy Sweet

U.S. Bureau of the Census

United States decennial censuses give snapshots of both the number and the characteristics of the population. The questions used to obtain the information change as the population itself changes; both the particular questions asked, and the form of those questions gives us insight into the general thinking of American society. The collection of ethnic data reflects these changes.

Ancestry data are more difficult to collect than other social and economic information because of the lack of clear-cut definitions, changing terminologies, poor reliability, and lack of knowledge of the degree of affiliation with a group or groups. Although 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 should give more useful information if criteria for inclusion in particular groups could be established with reliability. Ancestry would refer to the ethnic group, "roots", or country in which a person or person's parents or ancestors was born, regardless of the number of generations removed from their country of origin; the ancestry would reflect identification, but not the degree of attachment or association the persons had with the particular ethnic group(s).

Since the Census Bureau had never asked a question on ancestry before, a considerable amount of preparations was necessary to determine which groups would be coded, which would not, which would be subsumed into other groups, and what the general hierarchy of the named designations would be. This classification was complicated by the fact that immigration has not been constant, that some groups arrived before other groups, that geography which existed in the 18th and 19th centuries no longer exists, that some groups immigrated under circumstances which would probably make them uncomfortable about naming their sending countries, that some could not name their sending countries in any case because time since arrival of the group has been too long, or because their cultures were stripped away from slavery or other alienation.

The ancestry data from the 1980 census reflect the diverse nationality groups which have come to the United States throughout its history. Prior to the 17th century, most of the residents of this country were American Indians; however, most of the 7 million respondents who reported American Indian ancestry in the 1980 census did so in combination with other groups, such as German, Irish, and English. Persons of English, German and Irish origin - the three largest groups reported in the census - were the first to come in large numbers; the immigration of these groups peaked in the 19th century. From the early 17th century to the early 19th century, many Africans were forcibly moved to the United States. The number of "newer immigrant" groups, including the Italian and Polish, reached an apex in the early part of the 20th century. Other groups, such as Portuguese and Greek, have had significant, but relatively

smaller, streams of immigrants settling in this country. The immigrants during the last two decades include substantial numbers of West Indian, Spanish, and Asian persons; the latter two groups were also among earlier immigrants.

In order to determine the ancestry classifications, the Census Bureau prepared a preliminary list using a number of source materials, including the Harvard Encyclopedia of American Ethnic Groups. Persons knowledgeable in different aspects of ethnicity then reviewed the preliminary list for accuracy, consistency, and completeness. On the basis of a review of the comments, 1980 census results, and additional research, the Census Bureau determined a final list of ancestries for use in 1980 processing and display.

Persons knowledgeable in ethnic identification sometimes have different views on some classifications since several groups may justifiably be classified in various ways. As a result, some experts would have classified several groups in the ancestry code list differently.

Although the final list was not exhaustive, it was seen as a preliminary attempt to classify groups for census purposes. This paper, then, is a first attempt to reassess the selection criteria used in the 1980 as the Census Bureau begins to consider refinements in the list for use in the 1990 Census. In this paper, we will only be looking at European groups, partly because to look at all groups would necessitate a much wider scope, and partly because we have recently been prodded to look carefully at the European groups by Magocsi (1987) and others.

THE ANCESTRY QUESTION

The ancestry question in the 1980 Census was based on self-identification and was open-ended. For the CPS and the Census, the question was worded "What is ...'s ancestry?" (but has been changed to "What is ...'s ancestry or ethnic origin?" for testing for the 1990 census), and had no prelisted categories. Some individuals reported a single ancestry group, others reported more than one group. The 1980 Census was the first census to collect ethnic data on persons regardless of the number of generations removed from their country of origin. Ethnic information collected in previous censuses came from questions on country of birth of persons and their parents and identified ethnicity for only foreign-born or native persons of foreign or mixed parentage.

Some respondents reported a single ancestry group while others reported more than one ancestry group, i.e., a multiple response. All single- and double-ancestry responses in each survey and the census were coded. In addition, 17 triple-origin ancestries expected to be frequently reported were coded, while only the first two reported ancestries were coded for all other responses of three or more ancestries. Since persons who reported multiple ancestries were included in more than one group, the sum of persons reporting the ancestry groups was greater than the total; for example, a person reporting "German-English" was tabulated in both the "German and other group(s)" and "English and other group(s)" categories. Also, persons reporting one of the unique three-origin groups were tabulated in each of the three ancestry categories (see USBC 1983:6 for

the list and numbers of triple-ancestry responses.)

On the basis of these criteria, 16 European ancestry groups had more than one million persons in 1980:

English.....	49,598,035
German.....	49,224,146
Irish.....	40,165,702
French (excluding Basque..)	12,892,246
Italian.....	12,183,692
Scottish.....	10,048,816
Polish.....	8,228,037
Dutch.....	6,304,499
Swedish.....	4,345,392
Norwegian.....	3,453,839
Russian n.e.c.....	2,781,432
Czech.....	1,892,456
Hungarian.....	1,776,902
Welsh.....	1,664,598
Danish.....	1,518,273
Portguese.....	1,024,351

In developing the processing plans for the 1980 ancestry data, certain conventions were adopted. For example, some pairs of ancestry responses (e.g. French-Canadian, French-Basque) may appear to reflect multiple ancestries, but instead are unique ethnic groups and were treated as a single group. Persons reporting combinations of closely related ancestries, such as "German-Bavarian," were tabulated as a single ancestry (in this case, German). In addition, responses such as "Polish-American" or "Italian-American" were treated as a single entry (i.e., "Polish" or "Italian"). Also, responses such as "Irish-Catholic" and "Russian-Jewish" were treated as a single ancestry (i.e., "Irish" or "Russian"), since United States law forbids the collection of information on religious identification in a mandatory census.

In the remaining sections we will look at how the 1980 classification scheme might be modified for the 1990 census.

CLASSIFICATION OF ANCESTRY GROUPS

Paul R. Magocsi (1987) has recently published a paper critical of the Census Bureau's classification of ancestry groups in 1980. Although 1980 was the first attempt at collecting ancestry data, and thus some problems in classification were expected, many of Magocsi's criticisms are important to consider as we develop a new classification scheme for 1990. There are a number of problems with Magocsi's criticism, especially because he confused ancestry groups displayed in the report "Ancestry of the Population by State: 1980" with data collected but not necessarily presented in the report. Magocsi starts his salvo by noting 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:). He finds weakness in the strength of the Census Bureau's open-ended question, noting that the wide variety of accepted responses "seemingly overwhelmed the unsuspecting census tabulators" (1987:). Also, "having qualified its own classification system, it seems 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:)."

Much of the rest of the paper, however, is an important contribution to the study of ancestry in the United States in its direction of how the Census

Bureau might better collect ancestry data to be used as a foundation for subsequent research and planning. Here we will consider Magocsi's comments on a case by case basis to help understand what the Census Bureau can do to obtain more effective data for analysis.

Here, we will follow the coding scheme developed by the Census Bureau and followed by Magocsi in his analysis, allowing us to look at the data collected on a "country by country" basis. At the end we will consider the "new" ethnic groups proposed by Magocsi which are not presented in the country analysis.

Magocsi finds that the Census Bureau is unrealistic in its assumptions about our display of information for Western European groups, noting "with regard to those parts of Europe, the Census Bureau tabulators seem to have been laboring under the widespread but nonetheless anachronistic notion that nation-states like Great Britain, Spain, Italy, the Netherlands, and especially France are ethnically homogeneous entities that in the course of their historic development have by the twentieth century eliminated the "problem" of national or ethnic minorities within their borders (1987:)". By looking at each country in turn, we can see how the Census Bureau collected data in 1980, and how we might improve data collection procedures for 1990.

Great Britain. The following categories were included in the 1980 census:

016 English.....	45,178,537
017 Channel, Guersey, Jersey Is....	22,306
018 Cornish.....	21,906
019 Welsh.....	1,882,260
020 Scottish.....	8,342,437
021 Manx.....	9,220
022 Northern Ireland.....	16,418
096 British Isles.....	12,113
097 British.....	175,258

These data demonstrate both the strengths and the weaknesses of the system used in the 1980 decennial census. The categories for English, Scottish, and Welsh probably worked effectively, although the numbers might have been smaller than some individuals in those groups might have liked. These groups are among the long-term immigrants with peak migration during the earliest flows of migration to the United States, causing many second and third or later generation migrants selecting multiple ancestry groups. Only the first two ancestry groups were coded unless respondents selected one of 17 triples (listed in Appendix). Persons selecting one of triples were recorded three times, once in each group; all other multiple responses appeared in only two places.

The smaller groups are more problematic, and fall into three kinds of categories unlikely responses, catch-all groups, and small groups. Each has a separate kind of issue to resolve. For Northern Irelander, for example, persons must have responded "Northern Irelander" to be coded into this category. If a person responded "Irish" to the question "What is's ancestry?", he or she was coded as Irish rather than as "Northern Irish" since the coder could not know that the persons would have chosen that response if a series of responses were listed. Therefore, although the Census Bureau recorded more than 16,000 responses of Northern Irelander, we probably would have obtained even more responses in this category had it been listed separately.

The second issue has to do with the catch-all category. In this case we have two: British Isles (12,000 responses) and British (175,258). Here

the respondent chose to select a rather broad category than a specific category. There are a number of reasons for this. 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 inter-married to a significant extent, and therefore could or would chose individual "British" ancestry groups. Others may not have understood that the "smallest unit" was expected, and chose British by default. Although we can expand the instructions to respondents it is unlikely that we will be able to eliminate this kind of responses for any of the reasons given here. Hence, one of Magocsi's biggest complaints, that we group categories indiscriminately, is only partially valid, because we must group according to the actual responses we get. Although he criticizes our showing British separately, noting "such a decision is unfortunate, since it means that all Americans whose ancestry derives from the British Isles...who may decide to describe their ancestry with the name of the multinational country they left - are simply classified as English", is not really true.

NEED TO LOOK AT THIS FURTHER, I THINK THIS IS MISPLACED

The third type of issue involves small groups, and the use of the census to obtain data on these groups. Here we have three small groups - Channel, Guernsey, Jersey Is (22,000), Cornish (22,000), and Manx (9,000). Since these data are based on a sample of the whole population, the numbers are estimated, based on the responses of about 20 percent of the entire United States population. This means that the 9000 Manx are based on the responses of about 1500 to 2000 persons, a relatively small sample of the total U.S.

population. The quality of these data for these small groups is further complicated by the method of coding the data. Coders in three locations around the country after the 1980 hand-coded the ancestry responses using a long, detailed coding list. On a number of occasions the coders made mistakes, for example, interchanging digits in the three-digit codes as they entered them on the forms, errors which could not be corrected because the miscode was also legal, and therefore, not caught by the computer during computer editing. Sometimes the coder thought he or she remembered the correct code for a group, but accidentally supplied the language or birthplace code rather than the ancestry response. This procedure also added errors, which would not affect the 45 million responses for English very much, but which would affect the 9000 responses for Manx considerably.

So, although as Magocsi notes, "it is true that some American descendants of the early Cornish immigrants who began coming in large numbers during the 1830s are still proud of their Cornish ancestry (1987:)", it is also true that they may have trouble obtaining completely reliable results from the Census data.

Yet another problem develops in trying to distinguish groups which are linkages of two separate ancestry groups, for example, Scotch-Irish or French Canadian, from their separate parts recorded as two parts of a multiple. French Canadian was coded as a single entry, presumably because it was more likely that this was not intended as two multiple entries. "Since the unique single response could not be distinguished from the multiple response, Scotch-Irish was treated as a multiple origin group and tabulated in both

categories "Scotch" and "Irish"...It will never be possible, however, to determine whether respondents intended to report a single response or a multiple-ancestry response" (USBC 1983:6). With this in mind, it is unlikely that the Census Bureau would change its policy on the coding of Scotch-Irish, since other combined multiples would also have to be coded as singles.

French. The following categories were obtained for French ancestry:

028 French.....	11,218,137
029 Breton.....	8,053
030 Corsican.....	20,393
031 French Basque.....	12,237
090 Alsatian.....	41,444
091 Lorraine.....	946
098 Acadian.....	32,799

One of the problems with Magocsi's analysis is his confusion between what the Census Bureau collected and what was published in the Supplementary Report on Ancestry (Ancestry of the Population by State: 1980). For example, because the number of persons reported as Breton was so small in 1980, for the supplementary report, Breton was grouped with French, which, according to Magocsi, "defies all logic other than that based on ethnically meaningless state boundaries (1987:)". In this case, the Census Bureau must claim innocence since data were collected, and tabulated separately for Breton; however, it is not clear what validity these data have because the numbers are very small, and have the same restrictions in use as for the small "British" groups, that they are susceptible to reporting and coding problems. The Census Bureau also reported Corsican separately. However, in developing code lists for 1980 and for 1990, it is important to remember a caveat

presented by Magocsi: "Bretons and Corsicans maintain their own organizations in the United States. Yet despite their distinct origins and community life in America, neither group is given its own entry, mentioned in the French entry, or identified anywhere in the otherwise ethnically sensitive Harvard Encyclopedia (1987:)". The Census Bureau is only restricted in its use of 3 digit codes, so that up to 1000 separate codes for ancestry groups can be included; however, we are restricted to those we know about, and which will be indicated by respondents in the mail-out, mail-back form. We use the Harvard Encyclopedia of Ethnic Groups and other sources in determining the code list; it is only through these sources and papers such as Magocsi's that we can learn about the various groups to include.

Alsatian and "Lorrainers" were coded separately, with the results shown in the charts. The number of persons selecting Lorraine was extraordinarily small, with the total being extremely unreliable. Nonetheless, it is possible to disaggregate the data. The inclusion of "Occitan" or "Occitan/Provençal" is possible, but if there are no more responses than obtained for Lorraine, it is unlikely that the Census Bureau would want to include the category over the long term.

Spanish. The following categories were collected for the European Spanish responses:

200 Spaniard, Castillian.....	86,657
201 Canary Is., Majorcan.....	2,557
202 Spanish Basque.....	8,534
203 Catalan.....	2,351
204 Galisian.....	3,218

As seen here, there were 2,400 persons who responded as Catalan in 1980. Given the small numbers, it is difficult to decide whether Catalan should be shown separately in general Census Bureau publications, and even more difficult to determine whether any demographic, social, or economic characteristics could be shown for this group, even though Magocsi feels that "there seems no excuse for subordinating Catalans, with their own language, national identity, even political autonomy in the twentieth century, under the rubric 'Spaniard'" (1987:). Of course, the data were collected separately, but not shown in the Supplementary Report because of their small numbers.

Italian. The following Italian groups had more than 750 responses in 1980:

052 Italian.....	11,896,153
053 Abruzzi.....	8,022
054 Apulian.....	3,553
055 Basilicata, Lucania.....	4,244
059 Friulian.....	984
060 Laxio, Vatican City, Rome..	5,029
062 Lombardian.....	1,079
066 Puglia.....	1,629
068 Sicilian.....	35,086
070 Trentino.....	1,866
072 Valle D'Aosta.....	897

In addition to these the Census Bureau also had separate categories for Calabrian, Amalfian, Campania, Emilia-Romagna, Ligurian, Marches, Molise, Piedmontese, Sardinian, Tuscan, Umbrian, and Venetian, none of which received more than 750 responses (although some respondents reported for each category). Therefore, Friulian with its small number (about 1000) was tabulated separately; again, the results may or may not be valid,