DIRECT AND INDIRECT MEASURES OF ETHNICITY: HOW DIFFERENT DEFINITIONS AFFECT THE SIZE AND CHARACTERISTICS OF VARIOUS ETHNIC GROUPS

Nampeo R. McKenney, Reynolds Farley, and Michael J. Levin[1]

INTRODUCTION

The growth of the United States population has been sustained by immigration. The flows from the different areas of the world--Africa, Europe, Latin America, and Asia--have led to a population that is diverse in race, language, national origin, and ethnicity. Throughout our Nation's history, there has been a great deal of interest in the composition of the population, but there has been little consensus about defining ethnic groups in censuses or surveys.

The determination of size and characteristics of ancestry[2] groups in the United States has been complicated by problems not innerent in other types of data. Variables such as occupation and marital status have their own definitional problems, but respondents are guided by societal standards and by previous census and survey experiences. Use of ancestry data is more difficult 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 can provide more complete information about the size and characteristics of a group.

This paper is one of a series of reports investigating the quality and completeness of data gathered by the new ancestry question that was asked in the Census Bureau November 1979 Current Population Survey (CPS) and the 1980 decennial census. The paper has two aims: first, data from the November 1979 CPS are analyzed to illustrate how the definition of an ancestry group influences estimates of its size and characteristics. Second, the ancestries reported in the November 1979 CPS are compared with those in the decennial census taken 5 months later to test for reliability of ancestry reporting.

Ethnic Identification: Previous Censuses

Until 1980, U.S. censuses asked three types of questions to describe the cultural or geographic origins of the population. First, data about race have been obtained in all decennial censuses. Second, a question about birthplace has been included in the decennial censuses since 1850, allowing determination of immigrants' countries of birth. Beginning in 1870, the decennial censuses asked respondents about both own birthplace and parental birthplaces. Third, censuses have included a question about mother tongue, ability to speak English, or use of languages other than English. Questions of this type were first asked in 1890, but tabulations concerning foreign mother tongue date from 1910.

Depending on the criteria selected, these variables can give indirect estimates of the size of ancestry groups. Because the number of first- and second-generation Americans declined between 1930 and 1970, questions about place of birth of parents and mother tongue were dropped from the 1980 census, so the foreign stock (first- and second-generation) population cannot be calculated. In 1980, a more direct measure was needed for estimating the size and characteristics of ancestry groups.

Ethnic Identification: 1980 Census and November 1979 CPS

In the 1980 census, respondents were asked a direct ethnic question, that is, "What is ...'s ancestry?" The ancestry item was based on self-identification. To determine how the results from the 1980 census questions compared with those used in previous censuses, both direct and indirect responses to ethnic identification were obtained by the Bureau of the Census in a special ethnic supplement to the monthly Current Population Survey (CPS) in November 1979.

The special ethnic supplement was designed to provide users with a basic set of data on ethnicity, which would serve as a bridge between the 1970 and 1980 censuses. Items in the survey included ancestry, country of birth of the individual and parents, citizenship, year of immigration, mother tongue, current language spoken in the home, and ability to speak English.

The ancestry question in the November 1979 survey was worded the same as in the 1980 census. Some individuals in the survey reported a single ancestry group; others reported more than one group. All single- and double-ancestry responses 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 ancestries reported was greater than the total population. Persons reporting "German-English," for example, were tabulated in both the German and English categories.

Although ancestry in 1979 was based on self-identification, selection was not completely independent since enumerators were instructed to prod respondents for a specific ancestry if the response was a religion, "American," or general, such as "European." When such responses were given, interviewers were instructed to explain that ancestry refers to the specific origin or nationality of the person or his or her ancestors before their arrival in the United States. However, if the respondent still reported "American" or a religion, then that response was accepted. About 14 million persons (approximately 6 percent of the total) provided a response of "American" or "United States," while 0.1 percent specified a religion or unclassifiable response and 11 percent did not report any ancestry. In the 1979 CPS, persons who called themselves "hyphenated" Americans such as "Swedish-American" were tabulated according to the ancestry that was not American. Also, those

who specified both an ancestry and a religion were tabulated by the ancestry only.

DIRECT AND INDIRECT DEFINITIONS OF ETHNICITY

The ancestry question, a direct measure of ethnicity, appears to be a substantial improvement over previous indirect measures (place of birth of persons and parents and language) for identifying ethnic groups. First, a major advantage of the new ancestry question is that it allows identification of the ethnicity of all persons, not just persons who would be classified on the basis of their own birthplace or birthplaces of their parents. Second, unlike birthplace, ethnicity provides data on the individual's own perception of ethnic identification. Third, the ancestry question is much more informative than the language inquiries.

The November 1979 CPS also permits the division of a particular ancestry group into first-generation immigrants, second-generation immigrants, and higher-order generations--that is, people who were born in the United States and who had both parents born here.

To analyze information provided by the ancestry question, correspondence between ancestries and countries was established. (See Levin and Farley, 1982 for a more complete description of the procedure.) For example, Korea was the appropriate country of birth for Korean ancestry. Puerto Rico was the appropriate place of birth for Puerto Rican ancestry. (For the purposes of this paper only, Puerto Rico birthplace was considered foreign.) Some ancestries were associated with two or more countries; for example, Belgian and the Netherlands both correspond to Dutch ancestry. Other ancestries, such as West Indian or Slavic, were associated with a number of countries of birth.

All persons who reported a given ancestry were classified by place of birth. If they were born in a "corresponding" country, they were first-generation immigrants to the United States. For instance, persons who claimed Jamaican ancestry and who were born on that island were identified as first-generation Jamaicans.

Persons who were not born in a country associated with their ancestry but had one or both parents born there, were classified as second-generation immigrants. For example, an American-born person who reported as Italian and whose father was born in Italy became a second-generation Italian. Those persons whose birthplace or parents' birthplace did not agree with their ancestry were classified as third or later generations. For example, persons who claimed Chinese ancestry and who were born in the United States and had parents born in Canada, were classified as third or later generation. One limitation of the classification system is that second-generation immigrants whose parents were not born in a country associated with the ancestry reported were included in the third and subsequent generations.

Table 1 presents a list of selected ancestry groups (single and multiple ancestries combined) with more than 100,000 persons. The first column shows the total number of people who reported a particular ancestry in the 1979 survey. Also, this table displays the distribution of the ancestry groups by generation; data refer to all persons who reported an ancestry whether single or multiple. The last column in Table 1 indicates the proportion of third and later generation persons who gave multiple responses to the ancestry question. The groups in Table 1 are arranged in descending order by the proportion who were in the third or later generations.

Table 1: Number of Persons, Proportions by Generation, and Proportion Multiple in Third and Later Generations for Selected Ancestry Groups: November 1979

Ancestry group	Total (000s)	Total	Generation			Percent of third or later generations reporting a multiple ancestry
			First	Second	Third or later	
Total[1]	241,302	100.0	4.4	8.9	86.7	73.5
African	2,926	100.0	0.6	0.2	99.2	9.8
Irish	43,752	100.0	0.6	3.3	96.1	79.5
Dutch	8,121	100.0	1.0	3.2	95.8	88.4
Scottish	14,205	100.0	1.1	3.4	95.5	90.4
Welsh	2,559	100.0	1.1	3.8	95.1	93.3
English	40,304	100.0	1.6	4.0	94.4	77.2
Scandinavian	340	100.0	-	6.3	93.7	70.3
German	51,649	100.0	1.5	4.9	93.6	69.1
French	14,047	100.0	1.5	5.5	93.1	80.8
Norwegian	4,120	100.0	2.1	10.1	87.8	76.4
Danish	1,673	100.0	1.8	11.0	87.3	78.8
Swedish	4,446	100.0	1.5	11.7	86.8	82.4
Swiss	1,229	100.0	3.4	12.4	84.3	81.3
Belgian	448	100.0	2.7	16.6	80.8	81.8
Finnish	616	100.0	7.3	18.0	74.8	72.4
Slavic	722	100.0	3.0	22.1	74.8	72.2
Polish	8,421	100.0	4.3	21.0	74.7	71.6
Czechoslovakian	1,695	100.0	4.8	21.6	73.5	64.3
Lithuanian	832	100.0	5.3	23.2	71.4	80.2
Ukrainian	525	100.0	9.0	20.1	70.9	77.0
Italian	11,751	100.0	6.5	25.4	68.1	46.5
Russian	3,466	100.0	6.2	27.5	66.3	73.3
French-Canadian	1,253	100.0	8.7	25.1	66.2	57.7
Hungarian	1,592	100.0	6.5	27.3	66.2	82.4
Other Spanish	3,566	100.0	19.2	15.3	65.5	49.9
Portuguese	446	100.0	13.6	21.2	65.4	64.8
Lebanese	322	100.0	9.5	27.1	63.4	64.8
Rumanian	336	100.0	11.9	30.6	57.5	78.2
Austrian	1,070	100.0	8.4	37.2	54.4	88.7
Greek	990	100.0	43.4	5.1	51.5	64.9
Yugoslavian	466	100.0	19.5	31.9	48.6	63.4
West Indian	193	100.0	28.4	27.2	44.4	47.7
Mexican	6,682	100.0	28.1	29.4	42.5	25.7
Canadian	610	100.0	21.4	39.1	39.5	76.6
Japanese	580	100.0	37.7	22.9	39.5	15.2
Chinese	734	100.0	43.7	29.3	27.0	87.3
Dominican	119	100.0	51.5	[illegible]	25.6	[illegible]
Asian Indian	142	100.0	57.5	[illegible]	24.1	44.5
Filipino	764	100.0	50.8	30.3	18.9	80.6
Puerto Rican	1,333	100.0	43.1	44.7	12.2	51.3
Jamaican	136	100.0	68.9	21.7	9.3	22.1
Colombian	118	100.0	68.1	24.6	7.3	22.1
Vietnamese	198	100.0	63.2	[illegible]	7.1	57.4
Cuban	676	100.0	69.6	26.3	4.6	66.8
Iranian	118	100.0	64.5	[illegible]	4.1	39.2
Korean	266	100.0	76.4	21.3	2.3	29.3

- Represents zero or rounds to zero.

[1] Does not include American, Afro-American, or American Indian since birthplace is overwhelmingly United States. Includes persons who reported a multiple ancestry; since those persons are tabulated in more than one category, the total is more than the total population of 215,613,000 in November 1979.

The first panel of data in Table 1 shows that about 40 million persons reported English as their ancestry. Only 640,000 of these persons (1.6 percent of the total) were actually born in England, while another 1.6 million persons (4.0 percent of the total) had one or both parents born in England. About 94 percent of the English ancestry persons were third or subsequent generations. The groups listed close to English in this table are very similar in terms of their generational distribution.

By the fourth block of data, groups are shown that have less than one-half in the third or subsequent generations. Less than one-half the Mexicans, for example, were born in Mexico or had one or both parents born in Mexico. Other groups, such as Jamaican, Iranian, and Korean, have very few persons in the third and subsequent generations.

The data in Table 1 illustrate the advantages of the ancestry question. Information about a person's birthplace and parental birthplaces would only identify part of the major ethnic groups. Seven percent or less of the people who actually reported African, Irish, Dutch, Scottish, Welsh, English, French, Scandinavian, and German would be classified in those groups in 1979 if

only birthplace questions were asked. Thus, the ancestry question provides a more complete measurement for identifying these groups, especially for those whose forebears came to the United States in the nineteenth or early twentieth centuries. On the other hand, this question is no more advantageous than the birthplace questions for identifying some of the groups that are recent arrivals, such as Iranian, Cuban, Vietnamese, or Korean. In these instances, ancestry measures little more than the first and second generation.

The last column of Table 1, showing the proportion of third and subsequent generations who claimed multiple ancestries also may illustrate the relative levels of intermarriage across ethnic groups. In order to claim multiple ancestry, in theory, a person's parents or earlier ancestors must have been of different ancestries. In other words, in order to have a multiple ancestry, a marriage of two single ancestries should have occurred at some point. For some groups--Asian Indian, for example--about half of the persons born in the U.S. and whose parents were both born in the U.S. claimed multiple ancestries, that is, Asian Indian and some other ancestry. On the other hand, about 90 percent of the native parentage Scottish persons claimed multiple ancestries, while about 6 percent of the Dominicans fell into this category. Many Scottish persons have been marrying persons of other ancestries for generations; apparently, few third and later generation Dominicans have parents who married outside the Dominican community, so most who claimed to be Dominican reported a single ancestry. Data files from the November 1979 CPS and the 1980 census will provide more information needed to analyze patterns of ethnic intermarriage.

SIZE AND CHARACTERISTICS OF ETHNIC GROUPS

The generational classification within groups was used to analyze the size and characteristics of the diverse ancestry groups. As stated in the previous section, the generations were defined as (1) first generation of a particular ancestry being those whose own birthplace agreed with their ancestry, (2) second generation being those whose own birthplace did not agree, but who had one or both parents with a birthplace which did agree, and, (3) third or later generations being those whose own birthplace and whose parents' birthplaces did not agree with their chosen ancestry.[3]

Table 2 shows the different rankings of selected ancestry groups by generation. For total ancestry (single and multiple combined), all groups are European except for Mexican, and the order roughly reflects the peak periods of immigration. German, Irish, and English, among the earliest arrivals in the U.S., lead the list and are much larger than any of the other groups. When looking only at third and subsequent generation persons, the order is somewhat similiar to the total. However, the major differences are that Mexican drops out of the top 10, and Norwegian which is the 11th largest total ancestry group is included. Hence, for third and subsequent generations, all of the largest groups are European.

Table 2. Rank of Ancestry Groups by Generation: November 1979[1]

Rank	Total ancestry		First generation		Second generation		Third or later generations	
1	German	51,849	Mexican	1,879	Italian	2,388	German	48,32[illegible]
2	Irish	43,752	Italian	767	German	2,555	Irish	42,20[illegible]
3	English	40,204	German	705	Mexican	1,363	English	37,77[illegible]
4	Scottish	14,025	English	538	Polish	1,270	Scottish	13,57[illegible]
5	French	14,347	Puerto Rican	575	English	1,180	French	13,27[illegible]
6	Italian	11,751	Cuban	469	Irish	1,137	Italian	7,59[illegible]
7	Polish	8,421	Greek	430	Russian	953	Dutch	7,18[illegible]
8	Dutch	8,121	Filipino	388	French	760	Polish	6,29[illegible]
9	Mexican	6,662	Polish	361	Puerto Rican	595	Swedish	6,24[illegible]
10	Swedish	4,686	Chinese	308	Swedish	572	Norwegian	5,61[illegible]

[1] Ranking based on groups shown in Table 1.

If the traditional questions on birthplace were used, only 5 of the birthplaces--Italy, Germany, England, Poland, and Mexico--corresponding to the top 10 total ancestry groups appear in the list for first or second generation. Furthermore, Mexican, which is ranked ninth among total ancestry groups, tops the list for first generation with almost 2 million persons born in Mexico who reported Mexican ancestry. Greeks, Cubans, Chinese, and Filipino who have a substantial number of more recent immigrants, are in the list of first generation immigrants, but not in the top 10 in the second generation.

If the second generation were used as the measure of ancestry, all groups except Dutch and Scottish of the top 10 total ancestry groups would be included, and Russian (12th on the total ancestry list) and Puerto Rican (18th) would also be added. Several groups--Irish, Russian, French, and Swedish--in the top 10 of the second generation, were not among the first generation list.

The numbers for both first- and second-generation ancestries shown in Table 2 are very small compared to those for the third and subsequent generations; thus, if the birthplace definitions were used instead of the ancestry designation, the selective ranking of the groups by population size would be very different.

Tables 3 and 4 show a non-random sample of 14 ancestry groups. These tables are used to determine how differentiating by generation affects the characteristics of various groups. In Table 3, for example, Polish had a median age of 59 years for persons in the second generation; the median for all persons reporting Polish ancestry was 28 years. Some groups, such as Russian and Swedish, had extraordinarily high median ages in the first generation, reflecting the decline of immigration after the turn of the century. Other groups, such as Mexican, had lower median ages, reflecting their continued immigration.

Table 3. Median Age of Selected Ancestry Groups by Generation: November 1979

Ancestry group	Total	First generation	Second generation	Ancestry only[1]
Russian	33	77	60	32
Swedish	29	76	66	33
Polish	25	66	59	19
Italian	27	59	44	19
Irish	29	56	53	31
German	27	58	58	25
French	27	60	51	26
English	31	49	51	28
Cuban	28	41	11	22
Lebanese	28	38	49	17
Filipino	26	37	12	11
Jamaican	23	37	11	28
Chinese	27	36	19	14
Mexican	21	26	17	10

[1] Figures differ from "Third or later generations" since persons reporting with another tongue or current language are not included.

Table 4. Characteristics of Selected Ancestry Groups by Generation: November 1979
(Numbers in thousands)

Characteristic	Ancestry group						
	English	French	German	Irish	Italian	Polish	Russian
Total Persons (Number)							
Total	49,006	14,367	51,849	43,752	11,751	8,421	3,446
First generation	638	324	765	253	687	361	215
Second generation	1,356	559	2,555	1,437	2,984	1,773	742
Ancestry only	29	11,342	45,572	163	7,301	5,262	2,229
In Central Cities (Percent)							
Total	44	43	45	43	55	52	56
First generation	52	45	47	44	49	37	31
Second generation	51	43	44	46	53	49	48
Ancestry only	42	44	45	48	57	56	59
High School Graduates (Percent)							
Total	73	67	71	68	66	67	68
First generation	72	63	68	66	55	52	51
Second generation	75	61	59	76	60	56	50
Ancestry only	75	58	72	60	71	71	80
Median Income (Number)							
Total	18.2	17.1	17.9	17.1	17.1	17.9	21.2
First generation	17.0	18.3	16.4	17.6	13.3	15.3	12.1
Second generation	17.4	16.8	16.8	17.9	16.8	19.0	22.8
Ancestry only	19.2	17.6	18.3	16.2	17.7	19.7	23.3

Table 4. Characteristics of Selected Ancestry Groups by Generation: November 1979 (Continued)
(Numbers in thousands)

Characteristic	Ancestry group						
	Swedish	Chinese	Cuban	Filipino	Jamaican	Lebanese	Mexican
Total Persons (Number)							
Total	4,996	706	676	764	196	322	6,682
First generation	76	308	469	388	127	31	1,879
Second generation	572	208	178	232	45	87	1,963
Ancestry only	3,998	157	12	126	2	168	1,182
In Central Cities (Percent)							
Total	44	62	58	64	66	37	38
First generation	(B)	61	59	61	51	(B)	36
Second generation	41	68	61	68	(B)	44	39
Ancestry only	44	68	(B)	67	(B)	36	40
High School Graduates (Number)							
Total	76	76	53	84	68	77	38
First generation	(B)	74	53	69	36	(B)	21
Second generation	69	80	48	92	(B)	84	46
Ancestry only	76	78	(B)	93	(B)	84	57
Median Income (Percent)							
Total	18.3	19.0	16.2	17.5	15.2	20.0	12.6
First generation	21	17.4	14.3	16.8	15.8	(B)	11.3
Second generation	13.5	24.8	13.6	19.8	(B)	22.4	13.9
Ancestry only	19.9	15.8	(B)	17.8	(B)	20.6	16.4

- Represents zero or rounds to zero

B Base less than 75,000.

However, the median age for those in the first generation was generally higher than that for total ancestry groups. By the second generation, there was less consistency. The median ages of the second generation were above the median for total ancestry groups for the older immigrant groups; on the other hand, the median is very low for some second-generation groups--specifically, Cuban, Filipino, and Jamaican.

Table 4 shows the distribution of three variables--proportion in central cities, proportion high school graduates, and median household income--for each of the 14 selected ancestry groups, divided by generation.

The immigration trends to this country can be seen in going from left to right across the table; the older immigrant groups are on the left, groups with large numbers of new immigrants are right. Given the low median ages of some of the groups, it seems likely that the proportions of the groups in the chart will change over time.

The data show considerable variation by generation. For some groups, such as Italian, Swedish, and Mexican, the proportion of high school graduates and the median income of households were lower among the first generation than for all persons in the ancestry group. On the other hand, this pattern was not evident for groups such as Chinese, Cuban, and Lebanese, which have large numbers of recent immigrants.

A previous paper (see Levin and Farley 1982) discussed a classification scheme for determining ethnic groups using different criteria, namely, respondent's birthplace, father's birthplace, mother's birthplace, current language, and mother tongue. From these 5 variables, 31 combinations were formed, e.g., birthplace only, birthplace and mother's birthplace, etc. A 32nd category was used when none of the variables agreed with the ancestry a person reported. Of the 32 different ways of identifying ethnic groups, combinations with more than 49,000 persons were used in this paper to do a more detailed analysis of the variations in characteristics.

Tables 5a and 5b show selected characteristics for persons who reported German or Mexican as their ancestry (either single or in combination with another response). About 2 million persons of German ancestry reported German as mother tongue (language spoken as a child); however, none reported German as current language, or Germany as their own or their parents' birthplace. They constituted about 5 percent of all Germans in the CPS. Their median age was 52 years, much higher than the 27 years for all persons reporting German ancestry. About 44 percent of those with German ancestry and mother tongue lived in central cities; 14 percent were college graduates; and their median family income was about $17,400.[4]

Table 5a. Direct and Indirect Responses of German Ancestry for Selected Characteristics: November 1979

Characteristic	Total		Median age	Central cities (percent)	College graduates (percent)	Median household income [1]
	Number (000's)	Percent				
Total ancestry	51,849	100.0	27.3	44.7	15	$ 17,900
MT only	2,426	4.7	52.1	44.1	14	17,400
CL only	68	0.1	33.3	27.2	33	22,300
MT CL	262	0.5	50.2	37.6	7	13,400
MB only	428	0.8	38.5	48.6	14	16,700
MT MB	218	0.4	52.7	46.5	17	18,000
MT CL MB	79	0.2	18.4	56.7	10	21,400
FB only	860	1.2	54.6	46.8	13	16,900
MT FB	327	0.6	66.2	40.3	9	14,200
MB FB	192	0.4	69.6	48.3	8	13,800
MT MB FB	697	1.3	67.6	41.5	9	14,600
MT CL MB FB	115	0.2	61.4	28.5	17	14,400
S only	63	0.1	16.6	43.8	16	11,300
MT MB FB S	252	0.5	58.2	46.9	13	15,600
MT CL MB FB S	301	0.6	52.8	52.3	12	16,600
Ancestry only	45,572	88.2	24.8	44.9	15	18,300
Birthplace	798	1.5	58.3	47.2	13	15,600
Father's BP	2,426	4.7	60.1	44.6	12	16,700
Mother's BP	2,209	4.3	55.8	45.4	12	15,100
Second generation	2,555	4.9	58.1	46.6	12	16,800
1st and 2nd generation	3,329	6.4	55.9	46.2	13	16,900
Current language	938	1.8	48.2	41.9	14	16,600
Mother tongue	4,997	9.6	54.2	43.6	13	16,403

Notes: MT is Mother Tongue, CL is Current Language, MB is Mother's Birthplace, FB is Father's Birthplace, S is own Birthplace. Only combinations with more than 50,000 respondents are shown.

1 The household is classified according to the ancestry of the householder. Numbers are rounded to hundreds.

About 3.3 million persons of German ancestry were included in the first- and second-generation category. The median age of 56 years was also much higher than that for all persons of German ancestry (27 years). The proportion of college graduates (12 percent) and median income ($14,900) was lower than that for the total ancestry group (15 percent and $17,900, respectively).

For Germans, in fact, most variables, alone or in combination with the others, were not very representative of the demographic and socioeconomic variables of the total German ancestry group. Of course, only 12 percent of the persons agreed with any of the 5 criteria. Conversely, the majority of Germans (88 percent) claimed this ancestry but did not identify as German by any language or birthplace questions. The analysis

of the data clearly demonstrates that birthplace and nativity do not substitute for ancestry as far as Germans are concerned.

Table 5b. Direct and Indirect Measures of Mexican Ancestry for Selected Characteristics: November 1979

Characteristic	Total		Median age	Central cities (percent)	College graduates (percent)	Median household income
	Number (000's)	Percent				
Total ancestry	6,668	100.0	20.6	44.0	4.8	$ 12,60
MI only	292	4.4	25.4	51.0	8.2	18,70
MI CL	1,338	20.0	22.4	27.3	3.9	13,30
MB only	72	1.1	7.9	48.8	-	-
MI CL MB	283	4.2	18.4	16.4	1.6	14,70
FB only	135	2.0	10.0	48.6	6.4	20,23
MI FB	60	0.9	25.2	57.7	10.7	24,30
MI CL FB	286	4.3	29.2	46.4	4.1	19,16
MB FB	266	4.0	7.6	25.2	-	12,33
MI MB FB	73	1.1	43.6	40.1	6.1	20,56
MI CL MB FB	737	11.0	22.4	36.0	3.0	13,23
MI CL MB B	49	0.7	26.6	36.7	6.6	13,20
MI CL FB B	44	0.7	32.7	43.2	13.3	11,40
MB FB B	56	0.8	7.7	34.5	-	15,60
MI CL MB FB B	1,560	24.6	28.6	35.0	3.9	11,15
Ancestry only	1,162	17.7	9.9	50.8	7.1	16,66
Birthplace	1,879	28.1	28.6	41.8	2.6	11,38
Father's BP	3,354	50.2	25.5	43.2	3.2	12,20
Mother's BP	3,283	49.1	25.0	42.2	2.6	12,22
Second generation	1,903	28.6	17.3	46.0	4.4	13,35
1st and 2nd generation	3,883	57.5	24.7	44.0	3.1	12,33
Current language	4,496	66.7	26.1	41.8	3.0	12,30
Mother tongue	4,885	73.1	26.3	43.8	3.7	12,50

- Represents zero or rounds to zero

Notes: MI is Mother Tongue, CL is Current Language, MB is Mother's Birthplace, FB is Father's Birthplace, B is own Birthplace. Only combinations with more than 50,000 responses are shown.

1 The household is classified according to the ancestry of the householder. Numbers are rounded to hundreds.

The other panel of data shows that Mexicans differ somewhat from Germans. Only about 18 percent of persons of Mexican ancestry are classified as Mexican solely because of their response to the ancestry question compared to 88 percent for German. In 1979, about one-fourth of those who said they were Mexican ancestry were born in that country, had both parents born in that country and reported Spanish as both a mother tongue and current language (i.e., identified on all five different variables).

For Mexicans, as for Germans, the social and economic levels also vary substantially by the criteria. Among persons of Mexican ancestry, the average age was lower among those whose only identification with the group was through their ancestry claim, compared to the total ancestry group. For Germans, the first and second generations had a median age above and an income level lower than all persons of the ancestry group. Among Mexicans, however, only first-generation immigrants were similar to the Germans, whereas among the second generation, the median age was lower and the median income higher than those of all Mexican ancestry persons.

The data as discussed from Tables 3, 4, 5a, and 5b show that the different definitions of ethnicity result in varying size and characteristics of the ethnic groups. Groups based on either the generational classification or the various criteria based on birthplace and language are generally not representative of the total ancestry group, as derived from the direct ancestry question.

ANCESTRY DATA IN THE 1979 CPS AND THE 1980 CENSUS

Since both the November 1979 CPS and the 1980 census used essentially the same open-ended type of ancestry question, the two data sources should permit a test of the reliability of ancestry reporting over time; that is, will people respond consistently for ancestry from survey to survey or censuses? In a preliminary evaluation of ancestry reporting, data from the 1980 census were compared with estimates from the November 1979 CPS.

Comparison of the 1980 census and the CPS ancestry data showed that the census had a substantially lower proportion (31 percent) of the total population reporting a multiple ancestry than in the CPS (38 percent). However, the proportions of persons (1) providing at least one specific ancestry, (2) responding "American," or (3) not responding to the ancestry question were consistent between the sources (Table 6).

Table 6. Type of Ancestry Response: 1980 Census and November 1979 CPS
(Numbers in Thousands)

Type of ancestry response	1980 census		1979 CPS	
	Number	Percent	Number	Percent
Total persons	226,546	100.0	216,813	100.0
Reported at least one specific ancestry	188,322	83.1	179,878	82.7
Single ancestry	118,565	52.3	98,496	44.5
Multiple ancestry	69,738	30.8	82,382	38.1
Ancestry not specified	38,243	16.9	37,535	17.3
American or United States	13,299	5.9	13,592	6.3
Other	1,763	0.8	195	0.1
Not reported	23,182	10.2	23,748	11.0

1 Includes responses indicating religious groups and unclassifiable responses.

Some of the factors affecting differences in ancestry estimates between the census and the CPS were:

1. The CPS estimates were based on population controls from the 1970 census that were updated for changes since then.. The 1979 CPS assumed a civilian non-institutional population of 216.6 million, while the 1980 census actually counted a total population of 226.5 million.

2. The CPS and the 1980 census used different procedures for collecting and processing the data: The information in the CPS was obtained through telephone or personal interview, while the 1980 census was conducted primarily through self-enumeration in which questionnaires were mailed to householders who were asked to fill in the required information. The personal interview of the CPS may have aided respondents in reporting ancestries more fully and may account for the greater frequency of multiple ancestry reports.

3. Fewer codes in 1979 than in 1980 were used because the the sample was fairly small and costs were very high for coding the large number of possible ancestry responses. For example, Croatian, Armenian, Hawaiian, Slovene, Syrian, and Serbian did not have unique codes in 1979, so did not appear separately.

Table 7 compares 1980 census and 1979 CPS estimates for the ancestry groups (single and multiple combined) with more than 100,000 persons in 1980.

The hypothesis that these different estimates were congruent was tested; congruence occurred when similar proportions of the total population gave the same specific ancestry response in 1979 and 1980. Specifically, the hypothesis that the 1979 and 1980 estimates were two samples drawn from one population with a common proportion giving a designated ancestry report was tested.[5] If the two sample estimates were close or identical, the hypothesis was rejected using the .05 level of significance. In Table 7, those cases

in which there was no evidence that the two sample estimates were drawn from different populations are marked with an asterisk. In all other cases, sampling cannot account for the differences between 1979 and 1980 data. For more details on sampling variability for 1979 and 1980 data, see reports cited in footnote 5.

Table 7. Selected Ancestry Groups by Type of Response: 1980 Census and November 1979

(Persons based on total population)

Ancestry group	Total		Single		Multiple	
	1979	1980	1979	1980	1979	1980
German	23.8	21.73	7.9%	7.89*	15.9%	13.81
Irish	28.22	17.73	6.51	4.56*	15.68	13.17
English	18.47	21.89	5.31	10.48	13.16	11.41
Scottish	6.56	4.44	.75	.52	5.81	3.92
French	6.48	5.69	1.41	1.35*	5.08	4.34
American, U.S.	6.11	5.87	6.27	5.87	NA	NA
Afro-American	6.12	9.25	6.74	9.06	.38	.19
Italian	5.42	5.18*	2.82	3.06	2.60	2.24
American Indian	4.57	2.98	.95	.86	3.62	2.12
Polish	3.88	3.63	1.81	1.86	2.27	1.95
Dutch	3.76	2.78	.53	.42*	3.22	2.36
Mexican	2.59	3.61	2.72	3.09	.37	.51
Swedish	2.26	1.22	.56	.37*	1.69	1.25
Norwegian	1.88	1.52	.57	.56*	1.33	.97
Spanish	1.66	1.19	.92	.76*	.72	.44
Russian	1.68	1.23	.69	.61	.61	.62
African	1.58	.39	1.22	.07	.13	.32
Welsh	1.19	.71	.21	.16	.98	.63
Czechoslovakian	.78	.84*	.37	.35*	.42	.49
Danish	.77	.67	.29	.19*	.57	.48
Hungarian	.73	.78*	.35	.32	.49	.46*
Puerto Rican	.92	.66*	.51	.56*	.38	.09*
Dutch	.57	.43	.14	.12	.42	.32
Armenian	.49	.42	.18	.15*	.32	.27*
French-Canadian	.48	.34	.27	.18	.21	.15
Greek	.46	.42*	.26	.27*	.19	.15
Portuguese	.44	.45*	.23	.27	.21	.18*
Lithuanian	.28	.23	.15	.15*	.13	.08
Filipino	.26	.35*	.24	.28*	.01	.07
Slovak	.23	.18	.14	.11	.10	.06
Chinese	.23	.19	.15	.12	.08	.06*
Japanese	.22	.15*	.16	.10	.07	.05*
Cuban	.21	.26*	.16	.22*	.05	.04*
Finnish	.20	.17*	.12	.12*	.07	.05*
Canadian	.19	.13	.11	.10*	.08	.03
Ukrainian	.19	.12	.11	.11	.08	.05*
Yugoslavian	.18	.16	.12	.09	.06	.07*
Belgian	.21	.16	.16	.15*	.05	.01
Scandinavian	.16	.21	.15	.11	.01	.10*
Rumanian	.15	.14	.10	.06	.05	.08
Lebanese	.15	.13*	.08	.09*	.07	.05*
Korean	.12	.17	.11	.16	.02	.01*

Only 15 of the 50 ancestry groups in Table 7 showed no statistical difference between the 1979 and 1980 estimates. Most of the 15 groups, for example, Greek, Filipino, Cuban, and Dominican, are those which have been experiencing large influxes of immigrants. Since large numbers of these people recently arrived in the United States, they may be very cognizant of their ethnicity and consistent in their reporting.

A higher degree of congruence between the 1979 and 1980 estimates is found among the single ancestry groups than among the multiple. That is, for 24 of the 50 single ancestries, the sample estimates from the CPS and the census do not appear to be drawn from different populations; older groups such as Germans, Irish, and French, as well as some Latin American and Asian groups, having increased immigration flows after the 1965 changes in admission laws, show consistent reporting by single response.

The proportion with multiple reports was lower in the census than in the CPS for 30 of the groups in Table 7. However, for 19 of the groups, the proportion reporting a multiple ancestry was not signficantly different in the two samples.[6]

The statistical analysis of Table 7 suggests there is a relatively higher degree of correspondence between the 1979 and 1980 sample for some groups (15 of 50 ancestry groups showed no statistical difference). The degree of correspondence was higher among single reporting than multiple reporting. As noted previously, differences between the CPS and census were expected due to differences in the censuses, processing, etc. However, for a few groups, the differences are larger than one might expect, allowing for the the factors noted above. Since replicability between censuses is needed if ancestry is to be used as a reliable indicator of ethnic background, it is important to understand why the census-CPS correspondence was so low for these few groups.

The group that was affected most in absolute numbers by the changes in procedures was English, which gained 10 million respondents, gaining about 12 million in the single reports, while losing about 2.7 million in the multiple category. The proportion reporting English rose from about 18 percent in November 1979, to 22 percent in April of the following year. Other parts of the British Isles were affected in the opposite direction, with Irish losing 3 million (about 2 percent), Scottish losing 4 million and Welsh 1 million. The sum of these losses almost completely compensated for the increase in English. Several aspects of the results of these changes can be seen in the ancestry reporting. For example, it is becoming clear that respondents in the census did not respond in the same way when not conversing with and receiving proding from an enumerator. Multiple ancestry categories, particularly other parts of Great Britain and Ireland, seem to have dropped when self-enumeration was used with the ancestry question in the census.

The 1980 census questionnaire design may have contributed to the differences in ancestry reporting for English. The prominence of the term "English" in the census question on language (which immediately preceded the ancestry question), and the listing of "English" as the second example in the ancestry question may have influenced respondents to report a single entry of English.

It is very difficult to interpret these results in the aggregate, partly because individual responses cannot be verified, but also because persons who respond with a multiple ancestry were counted more than once. Persons responding with triple ancestries in 1979 and with a single response in 1980 affected the 1980 count considerably.

The real test of the reliability of ancestry reporting will come when the same procedures are used again, either in another CPS, or in the 1990 census.

CONCLUSION

The 1980 census ancestry question appears to be a substantial improvement over previous inquiries. First, it provides more complete information about all individuals--not just those who were born abroad or those who had foreign-born parents. This paper has shown that the traditional birthplace questions would identify no no more than a small fraction of those persons who claimed English, Irish, or German ancestry.

Second, this new question minimizes confusion between birthplace and ancestry. That is, if an Italian origin family lived in Argentina for two generations prior to moving to the United States, a question about birthplace would be answered Argentina, while the same individual might reply Italian to the ancestry question.

Third, the ancestry question is much more informative than the language inquiries. Ancestry,

for example, distinguishes Dominicans from Cubans, although both ethnic groups have the same mother tongue.

In this paper, we have shown that the size and characteristics of a group depend on the procedure used to identify the group. For most ancestries, the estimates developed from the questions about birthplace of respondent or respondent's parents are much smaller than those developed from the ancestry question. Additionally, we have shown that with a few exceptions, the ancestry reports obtained in the November 1979 CPS are generally consistent to those gathered in the 1980 census, especially for persons who reported a single ancestry.

ACKNOWLEDGMENTS

The authors thank the following Census Bureau staff: Norman Kaplan and Milton Schuman for continued excellence in dealing with the complex concepts in the computer programming of these data, and Nancy Sweet for her professional assistance in preparing this paper. All are in the Population Division, Bureau of the Census.

FOOTNOTES

1 Nampeo McKenney is Assistant Division Chief for Ethnic and Racial Statistics Programs, Population Division, Bureau of the Census; Reynolds Farley, Population Studies Center, University of Michigan; Michael J. Levin, Population Division, Bureau of the Census.

2 In this paper, the terms "ancestry" and "origin" (and ancestry group and ethnic group) are used interchangeably. Ancestry refers to a person's nationality group, lineage, or the country in which the person or the 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 ancestry question reflected the ethnic group(s) with which persons identified and not necessarily the degree of attachment or association the person had with the particular ethnic group(s).

3 Lieberson and Santi (1983) have assessed the use of nativity data to estimate ethnic characteristics and patterns.

4 The income data refer to the median income of all households with a householder who satisfies the ancestry criteria; the household is classified according to the ancestry category of the householder. These median income figures were calculated from grouped data. Income figures gathered in the November 1979 CPS tend to be smaller than income figures for the same household gathered in the March CPS, since the March survey includes many additional questions about sources and amounts of income.

5 Estimates were obtained from Current Population Reports, "Ancestry and Language in the United States: November 1979," Series P-23, No. 116, and Supplementary Report, "Ancestry of the Population by State: 1980," PC80-S1-10. Standard errors were calculated using the formula on page 24 of the 1979 report, and page 83 of the 1980 report. A "T" test was used for the hypothesis that the 1979 and 1980 estimates of the proportion of the population in a given ethnic group were drawn from one population.

6 For an investigation of the consistency of reporting of ethnic origin in the CPS between 1971 and 1973, see Johnson, 1974.

BIBLIOGRAPHY

Johnson, Charles E., Jr.
1974 "Consistency of reporting of ethnic origin in the Current Population Survey," Washington, U.S. Bureau of the Census; Technical Paper 31 (February).

Levin, Michael J. and Reynolds Farley
1982 Historical comparability of ethnic designations in the United States, Social Statistics Section, Proceedings of the American Statistical Association.

Lieberson, Stanley and Lawrence Santi
1983 The use of nativity data to estimate ethnic characteristics and patterns, paper presented at the annual meetings of the Population Association of America, Pittsburgh, Pa.

U.S. Bureau of the Census
1982 Current Population Reports, "Ancestry and Language in the United States: November 1979," series P-23, No. 116. Washington, D.C.: U.S. Government Printing Office.

U.S. Bureau of the Census
1983 Supplementary Report, "Ancestry of the Population by State: 1980," series PC80-S1-10. Washington, D.C.: U.S. Government Printing Office.