

## HISTORICAL COMPARABILITY OF ETHNIC DESIGNATIONS IN THE UNITED STATES

Michael J. Levin and Reynolds Farley, Bureau of the Census

### INTRODUCTION

Historical comparability of ancestry<sup>1</sup> data within and between United States censuses has been complicated by problems not inherent 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 experience. Individual perceptions of what these variables mean and the proper way to respond have not changed considerably over time. Collection 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 might give more useful information if criteria for inclusion in particular groups could be established with reliability. Ancestry would refer to the group, lineage, descent, 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).

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 origin. Eleven decennial censuses, beginning in 1870 and ending a century later, asked respondents about not only their own birthplace, but also about parental birthplaces. Third, censuses have included a question about mother tongue, the ability to speak English, or the use of languages other than English; questions of this type were first asked in 1890 but tabulations concerning foreign mother tongue date from 1910.

Historical comparability of racial and national origin groups from these data can be used to give an indirect measure of the changes in ancestry of the population. Because the number of second-generation immigrants declined in recent decades, the questions about place of birth of parents and mother tongue were dropped from the 1980 census; therefore, the foreign-stock population cannot be determined. Another more direct measure of ancestry was now needed.

A question on self-identified ancestry was included in the 1980 census; respondents reported ancestry directly. Since the questions on parental birthplace and mother tongue had been eliminated, however, historical comparability was made more difficult. To determine how the results from questions used in the previous censuses compare with those used in the 1980 census, both direct and indirect responses to ancestry were

needed from the same set of respondents. Fortunately, this has been done. The initial Bureau of the Census report on self-identified ancestry was obtained from data collected in November 1979 in a special supplement to the monthly Current Population Survey (CPS)<sup>2</sup>.

The ancestry 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. Questions on ability to read and write any language, which were not included in either census, were also asked in this November 1979 CPS supplement.

This paper explores the issue of historical comparability by describing the ancestry information obtained in that survey, and how it can be used to look at generational changes within a census and changes between censuses. Since the ancestry question is an innovation and may be the most appropriate question to identify a large segment of the population, it will be considered in detail. Responses to the ancestry question will be compared to responses to similar questions in previous decennial censuses. In latter sections of this paper, data from the November 1979 CPS will be analyzed with a view toward determining historical comparability between censuses using the ancestry question.

### CHANGES IN THE COMPOSITION OF THE POPULATION

Trends in Nativity. Until the 1980 census, questions about race, language, and birthplace were asked to describe the geographic origins or ethnicity of the population. A very brief review of long-term trends in population composition is useful to determine whether responses to the ancestry question agree with data gathered in previous censuses. The 1870 census was the first census to distinguish first- from second-generation immigrants. Procedures for data collection and tabulation changed over the years; for example, most compilations of the native population born of either foreign (both parents foreign born) or mixed (one parent foreign born) parentage between 1870 and 1950 pertain to the White population only. This does not seriously affect trends since there were relatively few Black or Asian immigrants during this period. Censuses have also differed in their procedures for including or tabulating the non-White population which is not Black.

When the question about place of birth was first asked in 1850, about 10 percent of the population were born abroad. There was a high volume of immigration to the U.S. following the Civil War, causing the proportion foreign born to increase. Between 1890 and World War I, the proportion of foreign born exceeded 15 percent; this was the peak in the proportion foreign born in the population. The drop in immigration brought about by restrictive migration laws

reduced the proportion of foreign born to less than 5 percent by 1970.

Fertility rates of the native population turned down in the late 1950s and, in the next decade, new laws permitted a greater volume of migration. As a result, the foreign-born population grew at a faster rate than the native-born population during the 1970s, the first time this occurred since the years preceding World War I. The proportion foreign born increased to about 6 percent in the November 1979 CPS.

The proportion of the population who are second-generation immigrants, that is, natives born of foreign or mixed parentage, lags several decades behind the proportion foreign born. The proportion who were second-generation immigrants increased in the late nineteenth century reaching a peak of about 21 percent in 1920. Since that time, the proportion has fallen continuously. The 1979 survey found that about 11 percent of the population were born in the United States but had one or both parents born abroad. If the foreign-born population continues to grow more rapidly than the native, the long-term decline in second-generation migrants may be reversed (depending upon the fertility rates of the native and foreign-born populations).

Estimates of the native population of native parentage (third and subsequent generations) are available since 1870 when just under three-fourths of those enumerated were in this category. As a result of the high volume of immigration, the proportion native of native parents declined for four decades, falling to less than two-thirds of the population. For half a century thereafter (1920 to 1970), this proportion increased because of declining immigration, reaching 84 percent in 1970.

Concepts of race and ancestry differ but some of their categories overlap. The racial composition has changed substantially. At the time of the American Revolution, Blacks made up about one-fifth of the population of the colonies. Most of the subsequent foreign immigrants were White, consequently reducing the proportion Black. By 1940, the proportional representation of Blacks was only half what it was at the time of the War of Independence. Since then, rates of natural increase have been higher for Blacks than for Whites and the proportion Black has grown to 12 percent.

Because of changes in census procedures, it is extremely difficult to confidently describe trends in the growth of the Asian population or the American Indian population. Apparently the representation of these "other races" held just about constant at about 3 to 5 per thousand for most of the era between the Civil War and World War II. Since that time, the statistics suggest that the American Indian and Asian populations have grown more rapidly than either the White or Black populations. In 1979, there were about 20 persons of "other races" per thousand.

Trends in the Origin of the Foreign-Born Population. Responses to the ancestry question can be compared to statistics from the national origin questions which were asked in previous censuses. The decennial censuses and immigration statistics show change in the countries of origin of United States immigrants.

The migration streams are indicated in the data about the foreign-born population shown in Table 1. For selected dates throughout the last century, we have recorded the total foreign-born population and the 10 most commonly reported places of birth. For all dates, Germany, Canada, and the United Kingdom have been frequently reported. Other nations appeared or disappeared from the list as the international flows changed. The Scandinavian countries, France, and Switzerland fell from the top ranks and were replaced by Italy, Russia, and Eastern European nations. The changes of recent years are evident in the data for the interval from 1960 to 1979. Ireland, Austria, and Hungary disappeared and were replaced by Cuba, the Philippines, and China. Mexico, which has been a source of migrants throughout this nation's history, has become more frequently reported as a birthplace for the foreign born. By 1979, the number born in Mexico was more than twice the number born in Germany (the second largest contributor of foreign born).

Table 1.--Foreign-Born Population in the United States and Largest Foreign-Born Groups: 1880 to 1979

(Numbers in thousands)

1880			1910 <sup>a</sup>		
Country	Number	Percent	Country	Number	Percent
Total population	50155	100.0	Total population	91972	100.0
Total Foreign Born	6680	13.3	Total Foreign Born	13346	14.5
1. Germany	1967	3.9	1. Germany <sup>b</sup>	2501	2.7
2. Ireland	1855	3.7	2. Russia <sup>c</sup>	1603	1.7
3. England <sup>d</sup>	917	1.6	3. Ireland	1352	1.5
4. Canada <sup>b</sup>	716	1.4	4. Italy	1343	1.5
5. Sweden	194	.4	5. United Kingdom	1219	1.3
6. Norway	182	.4	6. Canada	1201	1.3
7. France	107	.2	7. Austria <sup>d</sup>	1175	1.3
8. China	105	.2	8. Sweden	665	.7
9. Switzerland	89	.2	9. Hungary	496	.5
10. Bohemia	85	.2	10. Norway	404	.4

  

1940 <sup>e</sup>			1960		
Country	Number	Percent	Country	Number	Percent
Total population	132165	100.0	Total population	179326	100.0
Total Foreign Born	11594	8.8	Total Foreign Born	9738	5.4
1. Italy	1624	1.2	1. Italy	1257	.7
2. Germany	1238	.9	2. Germany	990	.6
3. Canada	1065	.8	3. Canada	953	.5
4. United Kingdom	1042	.8	4. United Kingdom	832	.5
5. Russia	1041	.8	5. Poland	748	.4
6. Poland	993	.8	6. Russia	691	.4
7. Ireland	572	.4	7. Mexico	576	.3
8. Austria	480	.4	8. Ireland	339	.2
9. Sweden	445	.3	9. Austria	305	.2
10. Mexico	377	.3	10. Hungary	245	.1

  

1970			1979		
Country	Number	Percent	Country	Number	Percent
Total population	203194	100.0	Total population	216613	100.0
Total Foreign Born	9519	4.7	Total Foreign Born	11730	5.4
1. Italy	1009	.5	1. Mexico	2092	1.0
2. Germany	833	.4	2. Germany	954	.4
3. Canada	812	.4	3. Italy	836	.4
4. Mexico	760	.4	4. Canada	828	.4
5. United Kingdom	686	.3	5. United Kingdom	673	.3
6. Poland	548	.3	6. Cuba	599	.3
7. Russia	463	.2	7. Philippines	474	.2
8. Cuba	439	.2	8. Poland	401	.2
9. Ireland	251	.1	9. Russia	326	.2
10. Austria	214	.1	10. China	309	.1

<sup>a</sup> Includes Scotland and Wales. <sup>b</sup> Includes Newfoundland. <sup>c</sup> Data for specific groups in 1910 and 1940 refer to the White population only. <sup>d</sup> Poland was not accepted as a country of origin in 1910. Persons reporting Poland were assigned to Germany, Russia or Austria.

Sources: U.S. Bureau of the Census reports for 1880, 1910, 1940, 1960, 1970; Current Population Reports, Series P-23, No. 116.

Trends in Mother Tongue and Current Language. Responses to the ancestry question may reflect not only national origin, but also a person's mother tongue or current language. The 1890 census was the first to include a language question, but it only sought to determine how many

people were unable to speak English. In 1900, if a person could not speak English the language spoken was recorded but the data were not tabulated. Censuses in 1910 and 1940 asked about the mother tongue of the foreign-stock population; that is, people who were born abroad and those native persons of foreign or mixed parentage; tabulations were published for only the White population. The 1930 and 1960 censuses, asked the mother tongue question only of foreign-born persons, but the 1970 census and the November 1979 CPS asked all respondents if a language other than English was spoken at home when the person was a child.

Table 2 presents information about mother tongue. The changes over time in the list of the 10 most common non-English languages are surprisingly small. At all dates, German, Italian, Polish, Yiddish, and French have been the leading mother tongues. Several Eastern European languages--Czechoslovakian, Hungarian, and Slovak--disappeared from the list, but Swedish and Norwegian are still commonly reported mother tongues even though large scale immigration from those nations ceased in the 1920s. The changes in the recent period are indicated by the rise of Spanish and the appearance of Chinese as the eighth most common mother tongue in 1979.

Table 2.--Persons With Mother Tongues Other Than English: 1910 to 1979

(Numbers in thousands)					
1910 <sup>a/</sup>		1940 <sup>b/</sup>		1960 <sup>b/</sup>	
Mother tongue	Number	Mother tongue	Number	Mother tongue	Number
1. German	8817	German	4950	German	1279
2. Italian	2151	Italian	3767	Italian	1226
3. Polish	1708	Polish	2416	Spanish	767
4. Yiddish	1677	Spanish	1816	Polish	582
5. Swedish	1446	Yiddish	1751	Yiddish	504
6. French	1357	French	1412	French	330
7. Norwegian	1010	Swedish	831	Russian	277
8. Bohemian	539	Norwegian	658	Hungarian	213
9. Spanish	448	Russian	585	Swedish	212
10. Danish	446	Czech	520	Greek	173

1970		1979	
Mother tongue	Number	Mother tongue	Number
1. Spanish	7824	Spanish	9422
2. German	6093	German	5206
3. Italian	4144	Italian	4201
4. French	2598	French	2498
5. Polish	2438	Polish	2472
6. Yiddish	1594	Yiddish	1200
7. Swedish	626	Norwegian	595
8. Norwegian	613	Chinese	579
9. Slovak	510	Swedish	550
10. Greek	459	Greek	536

<sup>a</sup> Data refer to the mother tongue of the foreign-stock White population only.  
<sup>b</sup> Data refer to the foreign-born population only.

Sources: U.S. Bureau of the Census reports for 1910, 1940, 1960, 1970; Current Population Reports, Series P-23, No. 116.

The 23 languages reported as a mother tongue by more than 250,000 persons in either the 1970 census or 1979 CPS are listed in Table 3. The languages currently spoken at home by more than

100,000 persons in 1979 (in addition to the languages listed in Table 3) are listed in Table 4.

Table 3.--Languages Which Were Mother Tongues Other Than English Reported by More Than 250,000 Persons in 1970 or 1979

Language	Mother tongue		Percent change
	1970	1979 <sup>a/</sup>	
Total, mother tongue not English	33175 <sup>b/</sup>	34124	3%
Arabic	193	270	40
Chinese	345	579	68
Czechoslovakian	453	511	13
Dutch	351	316	-10
French	2598	2498	4
German	6093	5206	-15
Greek	459	536	17
Hungarian	447	421	-6
Italian	4144	4201	1
Japanese	409	470	15
Korean	54	217	302
Lithuanian	293	215	-27
Norwegian	613	595	-3
Philippine languages	218 <sup>b/</sup>	487	123
Polish	2438	2472	1
Portuguese	365	448	23
Russian	335	249	-26
Serbo-Croatian	239	306	28
Slovak	510	359	-30
Spanish	7824	9422	20
Swedish	626	550	-12
Vietnamese	NA	162	NA
Yiddish	1594	1200	-25

<sup>a</sup> Data refer to persons aged 5 and over.  
<sup>b</sup> Tagalog only. <sup>c</sup> Excludes not reported.  
 NA Not available.

Source: U.S. Bureau of the Census, Census of Population: 1970, PC(2)-1A, National Origin and Language, Table 19. Ancestry and Language in the United States: November 1979 and unpublished data.

Changes between 1970 and 1979 in the representation of mother tongues reflect the aging of the population and recent immigration patterns. Some languages, including Lithuanian, Swedish, Russian, and Yiddish, were the mother tongues of a rather elderly population in 1970 causing their representation to decline by 1979. Other mother tongues, among them French, Italian, and Polish, held constant since immigrants speaking these languages replaced those who died. Finally, the number reporting languages such as Arabic, Chinese, Tagalog, or Illocano, rose rapidly because of recent immigration.

Data in Table 4 illustrate the dominant position of Spanish. Just about one-half of the people who currently use a language in their homes other than English speak Spanish; Spanish speakers comprise about 4 percent of the total population. Two other languages--German and Italian--each have more than one million current speakers while there are just about one million who speak French.

The number who spoke a particular language in 1979 can be compared to the number who reported it as mother tongue. These are not necessarily the same people but in almost every case those who currently speak a language other than English also reported it as their mother tongue. The languages of immigrants who came to the United

Table 4.--Languages Which Were Currently Spoken at Home by More Than 100,000 Persons in 1979

(Numbers in thousands)

Language	Current language	
	1979*	As percent of mother tongue
Total, current language not English	17985	53%
Arabic	177	66
Chinese	514	89
Czechoslovakian	116	23
Dutch	97	31
French	987	40
German	1261	24
Greek	365	68
Hungarian	106	25
Italian	1354	32
Japanese	265	56
Korean	191	88
Lithuanian	62	29
Norwegian	77	13
Philippine languages	419	86
Polish	731	30
Portuguese	245	55
Russian	65	26
Serbo-Croatian	119	39
Slovak	82	23
Spanish	8768	93
Swedish	64	12
Vietnamese	157	97
Yiddish	234	20

\* Data refer to persons aged 5 and over.

Source: Ancestry and Language in the United States: November 1979 and unpublished data.

States long ago are often reported as mother tongues but seldom as current languages. For instance, those who claim they currently use a Scandinavian language are only about one-eighth as numerous as those who report a Scandinavian mother tongue. The number who currently speak Yiddish is only one-fifth as great as the number who report Yiddish as mother tongue.

For those groups arriving recently, the ratio of current users of a language to those with that language as mother tongue is very much greater. The number who use Chinese or one of the Philippine languages is more than 80 percent as great as the number who reported these mother tongues. For Spanish, the proportion is greater than 90 percent.

#### ANCESTRY FROM THE NOVEMBER 1979 CURRENT POPULATION SURVEY

The 1980 census is similar to all previous censuses in providing information about the racial composition of the population. It is also similar to all censuses since 1850 in that it enumerated the population born outside the United States. The 1980 census differs from all other censuses since 1870 because it will not provide information about the sons and daughters of the foreign born, but will provide ancestry information.

The ancestry question in the November 1979 CPS supplement was based on self-identification and was open-ended. The question was worded "What is ...'s ancestry?", and had no prelisted categories. 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 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.

Although ancestry was based on self-identification, selection was not completely independent since enumerators were instructed to prod respondents for a specific ancestry if a religion, the category "American," or an unclassifiable response was given. Self-enumeration was used in the 1980 census. The census questionnaire provided a number of examples, including Hungarian, Irish, Italian, Korean, and Afro-American, to aid respondents.

If a person gave an initial response of "American," "United States," or a religious response to the ancestry question in the 1979 CPS, interviewers were instructed to explain that ancestry refers to the specific foreign nationality of the person or his or her ancestors. However, if the respondent still reported "American" or a religion then that response was accepted. About 14 million persons provided a response of "American" or "United States," while fewer than 1 percent specified a religion. In both the 1980 census and the 1979 CPS, persons who called themselves "hyphenated" Americans such as "Swedish-Americans" were tabulated according to the ancestry which was not American. Also, those who specified both an ancestry and a religion were tabulated by the ancestry only.

Table 5 lists the 50 ancestries which were reported as single ancestries by at least 100,000 persons in 1979; the group "Other Spanish" has been excluded. Each of the largest groups--German, Irish, and English--was more than twice as great as the number in each of the 5 other ancestries reported by 10 million or more persons: Scottish, French, Afro-American, American, and Italian.

The ancestry survey reflects the diverse nationality groups which have come to this country throughout its history. Prior to the seventeenth century, most of the residents of this country were American Indians. Persons of German, Irish, and English origins were the first Europeans to arrive in large numbers; the immigration of these groups peaked in the late nineteenth century. From the early seventeenth century to the early nineteenth century, many Africans were forcibly moved to the United States. The number of "newer immigrants," including the Italian and Polish, reached an apex in the early part of the twentieth century. The most recent immigrants (during the last two decades) included substantial numbers of both Spanish and Asians.

A large number of people reported Afro-American, American, or American Indian ancestry. These were the fifth, seventh, and ninth most common responses. Since ancestry was reported separately from race, a person indicating a

Table 5.--Ancestry of the U.S. Population: November 1979

(Numbers in thousands. Civilian noninstitutional population)					
Ancestry	Number reported	Percent of total reported ancestry	Reported single ancestry	Reported multiple ancestry	Proportion single ancestry
German	51,649	26.8	17,160	34,489	33
Irish	43,752	22.7	9,760	33,992	22
English	40,004	20.7	11,501	28,503	29
Scottish	14,205	7.4	1,615	12,590	11
French	14,047	7.3	3,047	11,000	22
Afro-American	13,267	6.9	12,424	844	94
American	13,183	6.8	13,183	-	100
Italian	11,751	6.1	6,110	5,641	52
American Indian	9,900	5.1	2,053	7,847	21
Polish	8,421	4.4	3,498	4,923	42
Dutch	8,121	4.2	1,362	6,759	17
Mexican	6,682	3.5	5,889	793	88
Swedish	4,886	2.5	1,216	3,670	25
Norwegian	4,120	2.1	1,232	2,888	30
Russian	3,466	1.8	1,496	1,970	43
African	2,926	1.5	2,633	292	90
Welsh	2,568	1.3	455	2,113	18
Czechoslovakian	1,695	0.9	794	901	47
Danish	1,672	0.9	438	1,234	26
Hungarian	1,592	0.8	534	1,058	34
Puerto Rican	1,333	0.7	1,107	226	83
Swiss	1,228	0.6	312	916	25
Austrian	1,070	0.6	385	685	36
French-Canadian	1,053	0.5	582	471	55
Greek	990	0.5	567	423	57
Portuguese	946	0.5	493	453	52
Lithuanian	832	0.4	317	515	38
Filipino	764	0.4	525	239	69
Slavic	722	0.4	300	422	42
Chinese	705	0.4	540	165	77
Japanese	680	0.4	529	151	78
Cuban	675	0.3	558	117	83
Finnish	616	0.3	255	361	41
Canadian	609	0.3	228	381	37
Ukrainian	525	0.3	231	294	44
Yugoslavian	468	0.3	283	184	61
Belgian	448	0.3	113	335	25
United States	408	0.2	408	-	100
Scandinavian	340	0.2	110	230	32
Rumanian	335	0.2	132	203	39
Lebanese	322	0.2	179	143	56
Korean	265	0.2	230	35	87
Vietnamese	198	0.1	177	21	89
West Indian	193	0.1	129	64	67
Jamaican	184	0.1	158	26	86
Asian Indian	182	0.1	156	26	86
Anglo-Saxon	140	0.1	77	63	55
Dominican	119	0.1	107	12	90
Iranian	118	0.1	103	15	87
Colombian	117	0.1	101	17	86

particular ancestry could be of any race. The race question on the 1980 census enumerated about 26.5 million Blacks while the ancestry question in 1979 counted 16.2 million Afro-Americans and Africans. However, the race question on the 1980 Decennial Census counted 1.4 million American Indians while 9.9 million reported this ancestry in 1979, suggesting that many people apparently identify as American Indian by ancestry but not by race. Counts from the 1980 census agreed with counts from the ancestry survey for the Asian racial groups. The 1980 census enumerated 806,000 Chinese, 775,000 Filipinos, 701,000 Japanese, and 335,000 Koreans. The 1979 CPS reported 705,000 Chinese, 764,000 Filipinos, 680,000 Japanese, and 265,000 Koreans.

ASSESSING THE SIZE OF VARIOUS GROUPS

The place of birth and language questions asked in previous censuses sought to provide information about the size of various ethnic

groups. The ancestry question will provide a very different perspective since it will indicate how many people selected a given ancestry. In this sense, it may provide more complete information about the size of a group. It will not, however, unambiguously define the size of any group. Table 6 presents data from the November 1979 survey about the 10 most frequently reported ancestries excluding Afro-American, American, and American Indian.

Table 6.--Alternative Definitions of Ancestry: November 1979

(Numbers in thousands. Civilian noninstitutional population)							
Ancestry	Total reported	Reported single	Place of birth			Language	
			Individual	Father	Mother	Mother-tongue	Current language
German	51,649	17,160	954	2,469	2,281	5,419 <sup>b</sup>	1,298 <sup>b</sup>
Irish	43,752	9,760	183 <sup>c</sup>	1,150 <sup>c</sup>	1,223 <sup>c</sup>	101 <sup>d</sup>	9 <sup>d</sup>
English	40,004	11,501	623 <sup>e</sup>	1,397 <sup>e</sup>	1,344 <sup>e</sup>	159,743	190,428 <sup>f</sup>
Scottish	14,205	1,615	48	169	155	159,743	190,428
French	14,047	3,047	119	208	241	2,498	987
Italian	11,751	6,110	836	3,711	3,170	4,201	1,354
Polish	8,421	3,498	401	2,012	1,868	2,472	731
Dutch	8,121	1,362	82	289	247	385	109
Mexican	6,682	5,889	2,092	3,688	3,608	9,422	8,768 <sup>g</sup>
Swedish	4,886	1,216	80	579	457	550	64

<sup>a</sup> Data refer to the population aged 5 years and over. <sup>b</sup> Language data include approximately 212,000 who reported Pennsylvania Dutch as mother tongue and 37,000 who reported this as current language. <sup>c</sup> Includes Ireland and Northern Ireland. <sup>d</sup> Language data refer to Gaelic. <sup>e</sup> Includes England and Wales. <sup>f</sup> Includes approximately 14 million people who reported a current language other than English but who also reported speaking English well or very well. <sup>g</sup> Spanish language.

Source: U.S. Bureau of the Census, Ancestry and Language in the United States: November 1979, Series Y-23, Number 116.

First, there is the question of whether to focus upon reported single ancestry or multiple ancestry. For the largest groups, the number giving an ancestry as one of several ancestries was three to four times as great as the number who reported that ancestry individually. For example, there were fewer than 2 million persons claiming single Scottish ancestry, but more than 14 million persons reported Scottish in combination with other ancestries. Mexican was the ninth largest group on the basis of multiple reports, but fifth largest on the basis of single reports.

Second, one of the traditional indicators of ancestry has been the count of the foreign born or the foreign-stock population. If this approach were used, the Mexican group would have been the largest and Italians would have been more numerous than Germans. Some groups, such as Dutch and French, which were large when estimated by the ancestry question, were quite small when estimated by the foreign stock concept.

Third, Table 6 shows that language does not always relate directly to ancestry. The number who reported Gaelic as mother tongue or current language was a small fraction of those who claimed Irish ancestry. Of course, many who were born of Irish foreign stock have had generations of ancestors whose mother tongue was English. Numerous people who have English as both mother tongue and current language were not born in the United Kingdom and did not claim English ancestry. On the other hand, 73 percent of those claiming Mexican ancestry reported Spanish as mother tongue, and 67 percent currently spoke Spanish in their homes. Many other Spanish speakers claimed other ancestries, such as Puerto Rican, Cuban, or Dominican.

**Trends in Ethnic Identification.** There was a similar CPS supplement on ethnicity and language in November 1969 but historical comparability between the two surveys is very difficult. Although the statistics on nativity, parentage, and mother tongue are comparable, the data for ancestry and current language are not. In 1969, respondents were read a list of ten non-Spanish and five Spanish ethnic origins, and were asked to report only one ethnic origin. Respondents who provided multiple origin responses were included in the "other" category. Only single-origin responses were tabulated separately. In 1979, no prelisted categories were provided for the ancestry question, and respondents could report single- or multiple-ancestry groups.

For purposes of comparison, Table 7 shows the national origin responses which were obtained in the November 1969 CPS and the 1979 CPS. The numbers claiming a national origin in 1969 were generally larger than the numbers of single responses for that same ancestry in 1979 but were smaller than the total number who claimed that ancestry in 1979. For example, the 1969 survey enumerated about 20 million Germans compared to 17 million who gave that single response in 1979. Another 34 million people gave German as one of several ancestries in 1979. In the 1969 CPS, 91 percent of the respondents gave a national origin; in 1979 about 89 percent of the respondents listed at least one ancestry, including those who said they were Americans or gave a religious or other nonclassifiable response.

Table 7.--Comparison of 1979 and 1969 Ethnic Data

Ethnic group	1979	
	1969	Total   Single
Total, U.S.	198,214	216,613   96,496
English	19,060	40,004   11,501
German	19,961	51,649   17,160
Irish	13,282	43,752   9,760
Italian	7,239	11,751   6,110
Polish	4,021	8,421   3,498
Russian	2,152	3,466   1,496
Spanish	9,230	12,493   9,762
Mexican	5,073	6,682   5,889
Puerto Rican	1,454	1,333   1,107
Cuban	565	675   558
Central and South American	556	
Other Spanish	1,582	3,803 <sup>a</sup>   2,208 <sup>b</sup>
Other	105,633	(X)   37,209
Not reported	17,635	23,748   (Y)

<sup>a</sup> Combines Central and South American and Other Spanish. (X) Cannot be determined due to multiple responses. (Y) Does not apply.

Data from the 1969 and 1979 CPS surveys do not provide precise information about the changing size or representation of various groups, but they do illustrate that our knowledge of the size of a group is highly dependent upon the techniques used to obtain and code the data. The 1969 study permitted respondents to select no more than 1 of 15 origins; in the 1979 CPS and the 1980 census, single and multiple responses were coded.

**WHAT THE ANCESTRY QUESTION MEASURES**

The ancestry question in the 1980 census will be used to identify ethnic or national origin groups. However, the 1980 data are not comparable to previous censuses. Data from the November 1979 CPS can be used to explore what the ancestry question measures.

The ancestry question differs from other socio-demographic queries since responses may not be "correct" in the same manner as responses to questions about age or income. Although the accuracy of a reported ancestry cannot be determined exactly, those factors which are related to the choice of an ancestry can be studied. The 1979 ancestry survey gathered information about five variables which influence a person's selection of an ancestry. These are birthplace, mother's birthplace, father's birthplace, mother tongue, and current language.

The ancestry question provides a self-identification of a person. The individual, using whatever information he or she preferred, selected an ancestry which was recorded in the November 1979 survey. The same survey, however, gathered the more objective information about where a respondent or respondent's parents were born, and what languages he or she currently spoke, or spoke as a child. Presumably, the accuracy of these reports could be ascertained, while only the consistency of ancestry reporting could be investigated. It should be possible to see how ancestry is related to these other variables.

A preliminary attempt at ascertaining correspondences between ancestry and countries of birth and languages was made by identifying relationships for each of 48 frequently reported ancestries; the ancestry "Scandinavian" has been excluded, and "United States" has been combined with "American." In some situations, correspondences were obtained easily. For Korean ancestry, for example, Korea was the appropriate country of birth, and Korean was the language. Some ancestries were associated with two or more languages, such as both French and Dutch for persons reporting Belgian ancestry. Other ancestries such as West Indian or Slavic involved many countries of birth. It was difficult to specify countries and languages appropriate for the Central and Eastern European ancestries since there have been numerous border changes and because the same language is spoken in several countries. A broad range of countries and languages was accepted for Eastern European ancestries. For example, Yiddish, the sixth most common mother tongue, was accepted not only for German and Russian ancestries but also for all other Eastern European ancestries.

Table 8 shows the correspondence between the report of an ancestry and reports of country of birth or language. Some information pertaining to historical comparability is provided since responses based on the new ancestry question can be compared to the previously used questions about place of birth and language. The line of data for Belgian ancestry, for example, shows 3 percent born in Belgium, 14 percent having had a mother born in that country, and 14 percent (very often the same respondents) having had a father born there. Three percent of those claiming

Belgian ancestry reported that French or Dutch was the language they currently spoke in their home while 16 percent said that French or Dutch was their mother tongue. The final column, showing a figure of 77 percent for Belgian, indicates the proportion with place of birth, parents' place of birth, current language and mother tongue all being inconsistent with ancestry. Presumably, these were people who retained an ethnic identity with Belgium, but whose recent ancestors had been born in the United States and who no longer spoke the language of their forebears.

Table 8.--Percent Proportions of Variables Measuring Ethnic Origin Agreeing with Self-Identified Ancestry: November 1979

Ancestry	Birth-place	Father's birth-place	Mother's birth-place	Current language	Mother tongue	Agreement with none of these
African	0.6	0.6	0.8	1.7	1.7	96.9
Afro-American	99.4	98.7	98.6	99.1	98.5	0.1
American <sup>a</sup>	99.6	99.0	99.0	99.1	98.8	0.1
American Indian	99.8	99.0	99.1	97.9	96.1	-
Anglo-Saxon	96.8	94.9	96.1	97.2	97.4	-
Asian Indian	57.5	75.0	72.3	60.0	53.7	5.4
Austrian	8.4	37.5	35.0	7.6	26.0	49.4
Belgian	2.7	14.3	14.3	3.4	16.4	77.0
Canadian	21.4	39.6	41.3	99.2	95.7	0.1
Chinese, Taiwan	43.7	63.3	65.0	61.0	67.2	22.3
Colombian	68.0	81.5	83.3	78.5	80.3	-
Cuban	69.4	81.2	82.8	88.4	89.4	1.8
Czechoslovakian	4.8	22.7	20.1	6.2	28.0	61.7
Danish <sup>b</sup>	1.8	9.8	7.8	1.3	8.1	84.9
Dominican	51.5	66.9	61.5	80.8	82.0	5.6
Dutch	1.0	3.3	2.8	1.0	3.4	94.7
English	1.6	3.7	3.7	99.0	97.4	0.4
Filipino	50.8	71.3	71.8	47.9	55.5	17.6
Finnish <sup>b</sup>	7.3	20.7	20.2	12.0	35.1	59.9
French	1.5	4.7	4.3	3.6	11.6	85.0
French-Canadian	8.7	24.2	23.9	10.9	28.0	56.2
German	1.5	4.7	4.3	1.8	8.9	88.2
Greek	43.4	43.4	37.9	33.1	49.4	40.0
Hungarian	6.5	27.8	26.6	8.4	32.9	57.0
Iranian	64.5	91.1	84.2	58.8	70.1	2.9
Irish	.6	2.9	3.0	99.3	97.5	0.4
Italian, Sicilian	6.5	29.6	25.1	10.5	33.2	59.6
Jamaican	68.9	85.2	74.0	98.0	98.0	1.1
Japanese	27.7	43.5	56.7	31.2	58.4	27.7
Korean	76.4	78.9	90.0	62.6	72.7	1.2
Lebanese	9.5	31.5	26.8	14.9	36.7	49.6
Lithuanian	5.4	25.0	23.6	7.8	31.2	60.9
Mexican	28.1	50.2	49.1	66.7	73.1	17.7
Norwegian <sup>b</sup>	2.1	9.3	7.8	1.8	13.7	81.3
Polish	4.3	22.1	20.5	8.8	32.1	62.5
Portuguese	13.4	32.7	26.1	21.1	38.5	55.4
Puerto Rican	43.1	81.1	79.4	72.5	77.7	7.8
Rumanian	11.9	33.8	30.3	13.3	20.0	49.1
Russian	6.2	29.8	24.3	5.3	28.3	58.6
Scottish	1.1	3.3	3.0	99.0	97.6	0.4
Slavic	3.0	22.2	21.9	8.2	33.8	59.9
Swedish <sup>b</sup>	1.5	11.3	8.9	1.7	11.9	81.8
Swiss	3.4	11.1	9.0	6.1	16.5	77.9
Ukrainian	9.0	26.7	24.6	14.7	39.5	51.2
Vietnamese	83.2	74.6	85.5	74.4	76.9	5.9
Welsh	1.1	3.7	3.1	99.4	97.2	0.2
West Indian	28.4	53.5	40.6	98.2	99.0	-
Yugoslavian	19.5	46.4	37.1	18.4	41.5	39.8

<sup>a</sup> Includes United States ancestry. <sup>b</sup> The general group Scandinavian is excluded. - Represents zero or rounds to zero.

Data for Belgian ancestry might be compared to that for Iranian ancestry. About two-thirds of those who claimed Iranian ancestry were born in Iran, and almost 90 percent had a mother, a father or both born in Iran. About 60 percent said they currently spoke Parsee and a slightly larger proportion claimed Parsee mother tongue.

The information in Table 8 suggests that ancestry responses were strongly influenced by national origins and by time of arrival in the United States. For those groups whose peak period of immigration was more than a century ago, the proportion reporting a single ancestry was low, the proportion born in the country of origin was small, and the proportion of cases in

which the ancestry report agreed with none of the birthplace or language items was high. English language groups are an exception since persons reporting English, Irish or Scottish, for example, as an ancestry almost always reported a mother tongue and current language consistent with their ancestry. For those groups which recently arrived in the United States, the proportion reporting a single ancestry was high, and the proportion born in an appropriate country of origin was large, while the proportion of people whose reported ancestry response did not agree with birthplace or language was small.

This relationship is illustrated in columns 1 and 2 which show data for the birthplace of the respondent and the respondent's father. The percent of reports of agreement of father's birthplace with ancestry are highly correlated with the proportion of the group born in the country of origin. Most of those groups which arrived recently have a high proportion reporting father's birthplace consistent with their own country of birth. At the other end of the continuum are those groups which had peak periods of immigration to the United States or the colonies more than a century ago.

The comparison of data for Belgians and Iranians illustrates the different ways in which language and birthplace are related to reports of ancestry. Since it is extremely difficult to analyze data for many different ancestry groups, the information gathered in the November 1979 survey was used to create several groups of ancestries. The groups were to be as internally homogeneous as possible but maximally different from each other. A clustering routine in the SAS computer program package was used for this purpose. The variables used for the clustering were, for each ancestry, the proportion of respondents whose ancestry was consistent with own place of birth, with the birthplaces of each parent, and with current language or with mother tongue. Additionally, for each ancestry the proportion of respondents whose ancestry did not agree with any of these indicators of ethnic origin and the proportion who gave a single, rather than a multiple response to the ancestry question were determined. As a result, the values of the seven variables shown in Table 8 were used in the clustering procedure.

In this clustering program, ancestries were grouped so that the distance between ancestries in a group was minimized but the distance between the groups of ancestries was maximized. The statistical analysis reported that the ideal fit, that is, the fit which minimized the ratio of "within" distance to "between" distance was obtained when ancestries were grouped into six categories. These are reported in Table 9 along with the average values for the variables in each grouping. The first group includes descendants of the earliest non-English speaking settlers coming to the United States. Their countries of origin were in Western and Northern Europe with the exception of the African ancestry group. Few members of these ancestries were born abroad or spoke those languages which are used in the countries of their forebears. Since many of these individuals came from families which had been in the United States for generations, the proportion

whose ancestry report agrees with none of the five ethnic indicators is high. A large proportion of these people in the first group report a multiple, rather than a single ancestry. This implies that their forebears came from many different nations.

Table 9.--Classification of Ancestries by Birthplace, Origin, and Proportion Single: November 1978<sup>a</sup>

Ancestry groups	Number of ancestries	Proportion ancestry agreement with Birthplace of...					Proportion no agreement on 5 items	Proportion single ancestry
		Respon-ident	Resp's Father	Resp's Mother	Current language	Mother tongue		
Group 1	9	1.8	7.7	6.7	2.5	10.2	85.3	32.6
Group 2	6	9.0	17.8	15.8	99.0	97.4	0.2	30.8
Group 3	14	7.5	27.6	24.9	10.7	31.7	56.5	44.3
Group 4	6	35.5	53.0	52.9	43.0	57.5	27.5	71.4
Group 5	9	64.7	79.5	79.2	74.9	77.9	3.5	86.3
Group 6	4	98.9	97.9	98.2	98.3	97.7	0.1	67.4

Group 1: African, Belgian, Swiss, Danish, Swedish, Norwegian, French, German, Dutch.  
 Group 2: Canadian, West Indian, English, Irish, Welsh, Scottish.  
 Group 3: Austrian, Rumanian, Czechoslovakian, Russian, French-Canadian, Italian, Finnish, Lithuanian, Polish, Slavic, Hungarian, Lebanese, Portuguese, Ukrainian.  
 Group 4: Chinese, Filipino, Mexican, Greek, Yugoslavian, Japanese.  
 Group 5: Asian Indian, Iranian, Korean, Vietnamese, Colombian, Cuban, Jamaican, Dominican, Puerto Rican.  
 Group 6: Afro-American, American, American Indian, and Anglo-Saxon.

<sup>a</sup> Data clustered from Table 8. U.S. included with American; Scandinavian is excluded.

The second group includes ancestries coming largely from the British Isles or places settled by immigrants from Britain including Canada and the West Indies. Relatively few of these people were foreign born but almost all of them spoke English so they had a mother tongue and current language which was consistent with the ancestry they reported.

The third group includes those ancestries whose forebears did not speak English and who came to the United States in great numbers between the Civil and First World Wars. This includes most of the Eastern and Southern European ancestries. Only a few of the people who claimed the ancestries in group three were born in their country of origin but about one-quarter of them had a parent born in such a country and one-third of them reported a mother tongue consistent with their reported ancestry.

Groups four and five include newer immigrants. Many of the ancestries comprising group four were represented in the migration streams which came to the United States throughout the entire twentieth century while group five includes the most recent immigrants. These are the people who arrived in the United States in large numbers following the 1965 change in immigrant laws. About two-thirds of the individuals claiming these ancestries were born abroad and about 80 percent had parents born abroad and reported mother tongues consistent with their ancestry. A very high proportion of these recently-arriving people claimed only a single ancestry.

The sixth and final group resulting from the cluster analysis might be termed the American group. This includes those who persisted in saying they were Americans as well as those who reported Afro-American or American Indian as their ancestry. They are distinct from other

groups since almost all of these people were born in the United States, had parents born here and, because English was reported as a current language, almost all of them spoke a language consistent with ancestry.

#### SUMMARY AND CONCLUSIONS

We will summarize our discussion of historical comparability by referring to the concepts or questions which have been used in the decennial enumerations.

**Race.** Each of the nation's 20 decennial censuses tabulated race but the racial categories have changed frequently, sometimes making comparability difficult, and also giving little data useful for describing ancestry.

**Language.** Language questions have appeared on censuses since 1890 but the questions used in 1980 differ from those of previous enumerations. Language tabulations from the census of 1980 will hardly be comparable to those from earlier enumerations. In 1980, individuals in the sample were asked whether they spoke a language at home other than English. If so, they were asked to name the language and indicate their ability to speak English. According to preliminary tabulations, approximately 11 percent of the population aged 7 and over speak a language other than English in their homes. About 2 percent of the total population reported that they could not speak English well.

**Place of Birth and Place of Birth of Parents.** The census of 1980 included a question about place of birth just as the 12 preceding enumerations did. Thus, comparable data about the foreign-born population are available for a long span. Preliminary tabulations from the census of 1980 report that just under 14 million persons or 6 percent of the population were born abroad. For the first time in many decades, the foreign-born population grew more rapidly than the native. The census of 1970 counted only 10 million persons or 5 percent of the population as foreign born. Although place of birth of parents has been asked since 1870, this question was not asked in 1980. Historically comparable data about second generation migrants will not be available from this census.

**Spanish/Hispanic Origin or Descent.** The 1970 census was the first to include a question designed specifically to identify ancestry groups by collecting data on the population of Spanish origin. All persons were asked whether their origin was one of five Spanish designations. The question was repeated in 1980 although the wording and the Spanish groups were changed. In 1970, 9.3 million persons or 4.6 percent of the population indicated they were of Spanish origin or descent. By 1980, the Spanish origin question led to a preliminary estimate of 14.6 million persons or 6.4 percent of the total population.

**Ancestry.** An innovation in the 1980 census was the question which asked respondents to write in their ancestry or ancestries. The question provides unique information about the origins or ethnicity of many people who could not be identified through the traditional language or place of birth questions. However, these data will not be comparable to information gathered in any previous census.



No ancestry tabulations from the census of 1980 - not even preliminary ones - have yet been released so information from the November 1979 CPS supplement were analyzed. A public use file of these data issued by the Census Bureau should be particularly valuable to persons interested in ethnicity since the November 1979 CPS asked both the questions used in 1970 and those used in 1980.

Several conclusions can be drawn from this analysis. First, the ancestry question appears to elicit useful information from a very large fraction of the population. Approximately 89 percent of the respondents contacted in this survey listed one or more ancestries (including American). Very few persons gave religious groups or responses which could not be coded. If data had been gathered by asking about foreign birth or foreign languages, most respondents would have given an answer which would not have permitted a classification by origin or ethnicity.

Second, most responses to the ancestry question apparently reflect the respondent's country of origin. Further analysis is required to ascertain the manner in which respondents used these concepts in the selection of an ancestry. While counts of Asian groups from the race question on the 1980 census are quite similar to the counts obtained from the 1979 ancestry question, the count of American Indians is not. Apparently many individuals designate American Indian as one of their ancestries but not their race. Despite prodding from interviewers, American was claimed by about 7 percent of the persons who gave a classifiable response in November 1979.

Third, although the ancestry responses provide data about the size of groups, they will not end debate about this matter. In general, the estimates from the November 1979 CPS are larger than previous estimates for many national origins. This is partly due to multiple responses. Coding the additional ancestries given by a respondent or using a different approach to assess the size of these groups might lead to even larger estimates.

Fourth, reports of ancestry from persons whose forebears arrived in the American colonies or the United States long ago differ from the reports of people who arrived recently. For groups such as Irish, English, or German, very few of those claiming the ancestry were born in their country of origin or have parents born there. Many who report such ancestries do so in combination with reports of other ancestries. In contrast there are groups whose members arrived primarily since 1960 such as Cubans, Iranians and Vietnamese. Many persons claiming these ancestries were born in those countries, speak the language of that country and have parents born there. Very few of them claim multiple ancestries.

#### ACKNOWLEDGMENTS

The authors thank the following Census Bureau staff: Norman Kaplan and Milton Schuman for the computer programming, Donna Keegan for secretarial assistance, and the Racial and Ethnic Statistics Staff for helpful comments.

#### FOOTNOTES

- <sup>1</sup> Ancestry, ethnicity, ethnic identification, and origin are used interchangeably in this paper.
- <sup>2</sup> Data on ancestry are not yet available from the 1980 census. The estimates from the November 1979 CPS may not agree with data from the 1980 census. The CPS estimates are based on population controls from the 1970 census which have been updated for changes in the population since that time. Also, the CPS and the 1980 census used different procedures for collecting and processing the data. Self-identified Spanish origin data have been collected since the November 1969 CPS, but this is the first time ancestry has been asked using an open-ended format. A much broader range of ancestries were tabulated in the November 1979 CPS than in prior CPSs.