

U.S. Historical Statistics

Hispanics in the United States, 1850–1990

Estimates of Population Size and National Origin

BRIAN GRATTON

*Department of History
Arizona State University*

MYRON P. GUTMANN

*Department of History and Population Research Center
The University of Texas at Austin*

In this article, we present the first estimates based on representative data of the size of the Hispanic population of the United States from 1850 to 1960. We also provide estimates of the size of each of the major Hispanic subgroups. Our measures can be linked to estimates made since 1970 through the Hispanic-origin variables used in the U.S. census. The data sets for Hispanics constitute a new nationally based source for scholarship, one grounded in the U. S. censuses. These files are now publicly available so that other scholars may use the data in their research.¹

The data sets rely on the Integrated Public Use Microdata Series (IPUMS), which provides researchers with probability samples of the U.S. population from eleven of the U.S. decennial censuses between 1850 and 1990.² After outlining the basic logic for selecting subsamples from IPUMS, we provide population estimates for Hispanics, using two approaches. The first relies on a set of identifying variables that are largely consistent across more than a century. The second uses additional variables available in certain censuses to improve the accuracy of the estimates of the Hispanic population.

We then discuss the techniques employed to distinguish among Hispanics of different national origin. For the period 1850–1970, we estimate the size and proportionate representation of five subgroups defined by origin: Cuban, Mexican, Puerto Rican, Spanish, and other (the last made up of known, but minor, national categories).³ We compare the overall and subgroup results with previous estimates and with those based on the 1970 Hispanic-origin variable. Finally, we critically examine the series founded on the consistent set of variables, exploring the process of qualification and measuring possible errors of omission and commission.

Hispanic "Ethnicity"

Hispanics do not constitute an ethnic group, although they share important common characteristics, the most essential being a historical connection to Spain: All Hispanics have origins in regions in which language, religion, law, and custom have been influenced by Spanish rule. This common historical inheritance leads to our use of the term *Hispanic* instead of *Latino*. The first refers to Spain and the Spanish language, whereas the second links persons to Latin American societies. For our purposes, the latter reference would be inaccurate in several respects. Many Latin Americans do not have a Spanish heritage (Haitians and Brazilians being obvious examples); in addition, many Hispanics in the United States have direct Spanish ancestry. Such conditions recommended the use of the term *Hispanic*.

Although often considered as a single group, particularly by policymakers and politicians, the Hispanic population is made up of highly diverse peoples, whose differences are often more salient than their similarities. Given the striking variance among subgroups in racial, cultural, and historical background, the researcher must distinguish among subgroups of different origins, and doing so constitutes one of our central tasks.

Previous Estimates

Before 1970, estimating the size of the Hispanic population or that of its subgroups was problematic. Contemplating the large and politically important Mexican-origin population at the beginning of the twentieth century, the historian Roger Daniels (1990, 308) lamented that "it is all

but impossible to get accurate estimates of their total number." In *The Hispanic Population of the United States*, sociologists Frank Bean and Marta Tienda (1987, 38) warned researchers that they faced considerable "difficulties . . . using census data to demarcate ethnic boundaries" in the Hispanic population as a whole.⁴ Surveying a variety of census efforts, Carlos Cortés (1980, 697) noted that the "U. S. Census Bureau has never been quite sure how to categorize or count residents of Mexican ancestry."

Before the work reported here, scholars interested in nationally representative estimates generally had access only to census or immigration bureau publications that gave the number of foreign born from Hispanic countries and, in certain censuses, the number of persons with one or two parents born in such countries. Rough evaluations of the size of the resident population in regions taken from Mexico in the Mexican-American War (Nostrand 1992), and the use of a "Mexican" identifier on the race variable in 1930, offered alternative evidence, but of dubious quality. The first depended on extrapolation of data from a variety of sources; the second, on an inconsistent interpretation of census instructions (Grebler, Moore, and Guzman 1970; Cortés 1980). The 1930 racial designation officially added nothing to previous information on those of immigrant origin: enumerators were to record as Mexican "all persons born in Mexico, or having parents born in Mexico, who are not definitely white, Negro. . . ." The race clause actually restricted Mexican descent, because persons whose appearance met enumerators' standards for white, black, or other racial categories were not to be recorded as Mexican. Racist attitudes in some regions, as well as the form itself, encouraged census takers to mark the racial category of Mexican, regardless of the instructions (Longmore and Hitt 1944). It is evident that some enumerators classified third- and higher-generation individuals as Mexicans, because 1930 census publications reported data on native-born Mexicans of native parentage. Still, others did not record such persons to be of Mexican race (particularly in New Mexico). The meaning of the published results remains unclear, and the data produced are certainly not comparable to other measures (Longmore and Hitt 1944; Samora and Simon 1977).

Estimates for Hispanics not of Mexican origin or ancestry are still rarer. Joseph P. Fitzpatrick (1980) used first- and second-generation census data to trace the slow increase in the Puerto Rican population before 1950 and the explosive growth thereafter. Frances Leon Quintana (1980) noted both the lack of attention to Spanish emigration to the United States and the difficulty in using published census data to estimate their numbers.

In 1950, the census first experimented with a more general Hispanic indicator, one that identified individuals beyond the second generation. Census Bureau personnel recorded white persons of Spanish surname in five southwestern states: Arizona, California, Colorado, New Mexico, and Texas.⁵ Undoubtedly designed to focus on Mexican-origin

Hispanics who lived in the Southwest, it did not provide any information outside the region, nor did it distinguish among subgroups. Although the Census Bureau knew that the list of Spanish surnames omitted certain persons who were Hispanics (U.S. Bureau of the Census 1953), the opportunity to reach beyond the second generation encouraged the use of surname indicators. More extensive lists were employed in the 1960, 1970, and 1980 censuses, still limited to the five southwestern states (Ito, Gratton, and Wycoff 1997).

Additional variables—especially the intermittent recording of language characteristics beginning in the 1910 census—offered means for identifying Hispanics. The 1970 census marked a more decisive change. In that year, two new approaches were used to classify the population. The first method relied on a combination of variables that located "Spanish Americans," including Spanish as a mother tongue, Puerto Rican family background in three northeastern states, and Spanish surname in the five southwestern states. The second method asked respondents to identify themselves in four Spanish-origin categories (Mexican, Cuban, Puerto Rican, other). The latter procedure, using self-reported Hispanic origin, became increasingly detailed in the 1980 and 1990 censuses and has enabled researchers to make estimates of the Hispanic population and its subgroups with much greater confidence (Bean and Tienda 1987).

The development of microdata samples of the U.S. censuses before 1960 has also opened new doors for scholars interested in Hispanics (Ruggles and Sobek 1997a). Sampling teams have transcribed or are transcribing census enumeration records for all decennial years from 1850 until 1950, except for 1890 (which has been destroyed) and 1930 (which will not be free of confidentiality restrictions until 2003). The complete samples for 1860 and 1870 are not yet available. Coding teams working from the original schedules can identify Spanish surnames. Using the 1980 census list, data-entry operators were instructed to indicate whether the respondent had a Spanish surname in the samples taken from the 1940 and 1950 censuses, for example. This additional information thus allowed researchers to make national estimates of the Hispanic population and its subgroups beyond those based on birthplace variables (Gratton, Rosales, and DeBano 1988). The developers of the IPUMS at the University of Minnesota have followed a similar approach, so that Spanish surname variables are available for most years. The new surname information and the capacity to manipulate original variables in novel ways underpin the strategies employed here for selecting Hispanic subsamples.

Selecting the Hispanic Subsamples

Selection Variables

As a first step in developing a logic for selecting subsamples, we reviewed the IPUMS for variables useful in identi-

fying Hispanics. Table 1 displays the variables that might be used and reveals the constant availability between 1850 and 1970 (excepting 1850 and 1900) of three variables: birthplace, parental birthplace, and Spanish surname. The review also indicated that certain censuses offered additional variables (e.g., language characteristics), which led to a two-part strategy: (1) to select subsamples that depended on a *consistent* set of criteria (birthplace, parental birthplace, and Spanish surname) and (2) to create *maximized* subsamples when a census provided additional variables.

The consistent subsamples are especially useful for evaluating trends over time and for comparisons between censuses. Nonetheless, they underestimate the number of Hispanics at any census point. The second approach, which we used to create the maximized subsamples, provides a closer estimate of the Hispanic population and its national subgroups and, in certain censuses, compares favorably to self-reported Hispanic origin as an identifying tool. Because the maximized samples depend on variables that are inconsistently available, and which change in definition and universe, they are less useful for comparative research across time.

Even variables consistently available may undergo changes in definition and universe across the census samples that make up the IPUMS. Among the three variables that we used for the consistent subsamples, birthplace coding remained essentially the same, although the universes for parental birthplace shifted.⁶ The surname list changed in both definition and universe.⁷ In the examination of the estimates, we discuss the effects of such changes on the series of consistent subsamples.

Selection Rules for Consistent Subsamples

Our first selection rule was to cast the net as widely as possible, that is, to err on the side of accepting as Hispanic

someone who was not, rather than to risk excluding a legitimate case. We also followed a rule of hypodescent, in the sense that a distant Hispanic relative qualified a case. Researchers who use the data sets we have created may choose to remove cases they judge inappropriate. In the evaluation of estimates subsequently given, we examine categories that might be considered for exclusion; in the selection code, we provide detailed internal identifiers that allow elimination of cases.

Selection focused first on country. We considered birthplace, parents' birthplace, or grandparents' birthplace in a Hispanic country to be a clear, reliable, and valid indicator of being Hispanic. Subsequent study of persons selected in this process strongly confirmed this assumption because they exhibited other Hispanic characteristics, such as use of the Spanish language. Birthplace qualification used twenty Hispanic nations that the historical literature indicates had a Spanish historical background, especially countries in which the Spanish language is predominant (Lavrin 1996; Ruggles and Sobek 1997a):

Argentina	Honduras
Bolivia	Mexico
Chile	Nicaragua
Colombia	Panama
Costa Rica	Paraguay
Cuba	Peru
Dominican Republic	Puerto Rico
Ecuador	Spain
El Salvador	Uruguay
Guatemala	Venezuela

We classified as Hispanic any person born in one of those countries, any person whose mother or father was born there, and any person for whom we could locate a grandparent born there.⁸ The spouse of a person classified as Hispanic also received Hispanic designation. If a person desig-

TABLE 1
Census Variables

Variable	1850	1880	1900	1910	1920	1940	1950	1960	1970 ^a		1980	1990
									(15%)	(5%)		
Ancestry												
Hispanic origin											XX	XX
Birthplace	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX
Parental birthplace		XX	XX	XX	XX	SR	SR	XX	XX		XX	XX
Mother tongue				FB	FB	SR		FB	XX			
Parents' tongue				FB	FB							
Spanish surname	XX	XX		OS	XX	XX	XX	5 st.	5 st.	5 st.	5 st.	
Language				XX							XX	XX
Race				OS	XX						XX	XX

Note: XX = available; OS = 1910 Hispanic oversample; 5 state = Arizona, California, Colorado, New Mexico, and Texas; FB = foreign born only; SR = sample record only (1940 and 1950).

^aIn 1970, a 15% sample of housing units received one form for additional questions; another 5% sample received a different form. The forms presented different questions.

nated as a Hispanic headed a household (including those identified through a spouse), all persons related to him or her were drawn in. Thus, children, adopted children, grandparents, and cousins were classified as Hispanics, whereas boarders and employees were not. The household approach and spousal designation do not identify a large number of cases not selected by more direct criteria, but we consider the methods to be useful, especially in redressing the gender bias that occurred when women with a Spanish surname lost that indicator by marriage (Perkins 1993).

In 1850, and in person-record portions of the 1940 and 1950 Public Use Microdata Samples (PUMS), parental birthplace was not recorded.⁹ In samples drawn from those PUMS, we derived parental birthplace when the child resided with the parent, accessing the information from the parent's record. As a result, in those subsamples the range of the parental-birthplace variable is both smaller and selected for children living at home and is not strictly comparable to those in other subsamples.

The second step in the process used Spanish surname. For persons not yet identified as Hispanic on the basis of birthplace, parents' and grandparents' birthplace, or by family birthplace connection, we applied two tests: if they had a Spanish surname, and if they were native born and had a native-born father, we accepted them as Hispanics. If they had previously qualified under country as spouses or relatives, we reclassified them as qualifying under surname as well. If they were household heads, relatives in the household without a Spanish surname were admitted; nieces, parents, cousins, and others in the household thereby qualified. Again, few cases were added by the household-surname rule because most qualified independently by country or surname; but this device captured additional members of Hispanic households, many of whom we thought were Hispanic, including daughters who had married and lost their Spanish surname. Persons identified by country, by surname, or by family connection to a household head or spouse with such qualifications became part of our Hispanic subsample.

In the surname identification, we excluded persons whose fathers were born in a non-Hispanic foreign country. Caution over native background stems from the fact that surname lists produce errors of commission. They classify as Spanish certain surnames that are also found in other countries (e.g., Italy) where Romance languages are spoken (Gratton, Rosales, and DeBano 1988). Because we had already selected those with Hispanic-country backgrounds before the selection by surname, we wanted to avoid *falsos amigos* (e.g., Italians whose surnames happened to be on the Spanish surname list). By permitting neither the foreign born nor those whose fathers were foreign born to enter, we reduced the number of persons with surnames incorrectly classified as Spanish. Thus we could eliminate an Italian with a surname equivalent to a Spanish surname, or a young woman who received a cognate

surname from her Portuguese-born father. We could not altogether avoid errors: it was possible for children with native-born fathers to qualify, although the immigrant grandfather was Italian. Such errors can be eliminated by locating the few cases in which a resident father is not in the Hispanic sample.

Finally, in 1940 and 1950, we drew separate samples from persons for whom the census gathered further information in those years (sample-record persons).¹⁰ Although those records have complete parental birthplace information as well as useful maximizing variables, they do not provide flat probability samples of all U.S. households and persons (Ito, Gratton, and Wycoff 1997). For those subsamples, we identified Hispanics by the birthplace variables and by surname, restricting selection to persons with sample records. Those cases constitute the basic subsamples; to the classified Hispanics on the sample record we have added all relatives living in the same household; even though such persons are likely to be Hispanic, they are not part of a probability sample.

Selection Rules for Maximized Subsamples

Selection of the maximized subsamples depended on the particular variables available in each census. The following variables were added to the consistent subsample variables in each PUMS through the 1970 (15 percent) sample:¹¹

1850: no other variables apply
 1880: no other variables apply
 1900: no other variables apply
 1910: race as Spanish, mother tongue, parents' mother tongue, language
 1920: race as Spanish, mother tongue, parents' mother tongue
 1940: mother tongue
 1950: no other variables apply
 1960: mother tongue
 1970: (15 percent): mother tongue, Spanish American status¹²

Although the basic logic of selection mirrored that for consistent subsamples, the nature of the additional variables required adaptations to the selection process. We divided additional variables into two categories, separating characteristics that applied only to individuals from those denoting a transmission of Hispanic heritage to others. Hence, language-based variables were spread to children and grandchildren and extended to the spouse and collateral relatives of a household head. In contrast, the intermittent classification as Spanish (or Mexican) on the race variable was applied only to individuals with that characteristic and not distributed to other members of the household.¹³

The variables used to maximize Hispanic identification varied from census to census; even those with the same label underwent important changes. As an example, mother tongue applied only to foreign-born persons in 1910, 1920,

and 1960 and meant the language spoken in the home before immigration. In 1940 and 1970, it applied to all persons, foreign or native born, and it recorded the language spoken in the home when the individual was a child. Given such shifts, the maximized subsamples do not link together as a consistent, comparable series. Nonetheless, they provide estimates that more closely approximate the actual number of Hispanics and should be used when researchers focus on certain census years and fuller estimates are a research priority.

Regions Once under Spanish and Mexican Jurisdiction

The censuses of 1850, 1880, 1900, and 1910 presented a unique historical issue: treatment of regions once under Spanish (and Mexican) jurisdiction.¹⁴ Texas broke from Mexican rule in 1836. On 4 July 1848, New Mexico, California, and what would become Arizona came under U.S. jurisdiction under the Treaty of Guadalupe Hidalgo. Thus, persons born in Texas before 1836, and in the other regions before 1848, might have reported Mexican birth (or, if before 1819, Spanish birth). Still, when those regions were under Spanish and Mexican rule, settlement by migrants from the United States could have produced a divided population, with different ethnic origins and allegiances. Evidence for ambiguous origins can be seen in the high proportion of persons stating birth in "New Mexico Territory," despite having been born before such an entity existed. In 1880, there were 298 persons 33 years of age and older who had listed New Mexico Territory as their birthplace.

We chose, therefore, to examine enumeration data directly for all persons in these PUMS who claimed birth in the region before U.S. jurisdiction but who had not been identified as Hispanic by surname.¹⁵ The data proved that such persons were overwhelmingly Hispanic in New Mexico and generally so in California and Arizona. They reported given names that were clearly Spanish, lived with persons of Hispanic background, and had parents born under Mexican and Spanish rule. Indeed, their surnames were almost invariably Spanish but had been misspelled. Many could not speak English (a question to determine language spoken was asked in 1900 and 1910).¹⁶

Texans not already selected, on the other hand, were clearly not Hispanic. As a number of scholars have reported, by the early 1830s more Anglo-Americans and African Americans lived in Texas than persons of Mexican origin.¹⁷ Direct examination showed this to be so; few of those identified by birth year had Spanish given names (rather, the opposite was true), few reported parents born in Texas, and nearly all spoke English.

Those results led us to apply the country rule to New Mexico, California, and Arizona: any person born before U.S. jurisdiction, or any person with a parent or grandparent born before U.S. jurisdiction, became part of our His-

panic subsample. We did not apply this rule to Texas, adding only those few cases that direct examination indicated had been missed.

Results: Hispanic Population

Table 2 displays the sizes of the Hispanic subsamples selected by our procedures, the size of the PUMS sample from which the subsample was selected, and a weighted estimate of the Hispanic population living in the United States in each census year. The final column indicates the percentage of the U.S. population represented by the Hispanic subsample. We provide estimates under consistent and maximized subsample criteria between 1850 and 1970 (15 percent sample). For 1970 (5 percent), 1980 (1 percent), and 1990 (1 percent), we display results from the Hispanic-origin variables. Thus, in 1920, our consistent Hispanic sample has 12,746 cases, representing 1,286,154 Hispanics—about 1.22 percent of the U.S. population in that year. The maximized sample size is 13,443, or 1,356,499 persons—1.28 percent of the population.

The consistent series reported in table 2 provides the first reliable national estimates of the size of the Hispanic population before 1970. It confirms the rising share of the U.S. population commanded by Hispanic groups, beginning from a relatively small base in the nineteenth century. The estimates suggest a much larger presence by 1910 than was hitherto believed (Gutmann, Frisbie, and Blanchard 1999; Cardoso 1980). Before the recognized effects of the Mexican Revolution, Hispanics represented about 1 percent of the entire U.S. population and constituted much higher percentages in the Southwest and Florida (Gutmann et al. 1999). The heavy immigration of the subsequent two decades had continuing impact at mid-century, when immigrants and their descendants pushed the Hispanic population above 2 percent of the population. After World War II, added streams from Cuba and Puerto Rico, along with rapid natural increase in the larger native-born base, continued to raise the proportion of Hispanics in the total U.S. population, which reached about 5 percent in 1970.

Identifying Hispanic Subgroups

Although the size of the Hispanic population as a whole is of considerable interest to scholars and to the public, most investigators emphasize the critical differences among Hispanics of different national origin. This emphasis has encouraged the increasingly fine distinctions made in the Hispanic-origin variables since 1970. As we have seen, before the 1970 census the lack of such identification had confounded estimates of the size and specific characteristics of different Hispanic groups. Generally, scholars interested in one subgroup had to rely on published data regarding the foreign born or those of foreign origin from that particular country.

Extending Historical Geography

We considered national origin an essential issue that could be elegantly addressed in the IPUMS by extending a commonly used geographical approach, which classifies origin by state of residence. Bean and Tienda (1987), for example, defined Hispanics without known background by classifying those living in the five southwestern states as of Mexican origin; those in New York, Pennsylvania, and New Jersey to be of Puerto Rican ancestry; and those in Florida to be of Cuban background. In a state with a strong history of domination by one national group, a native-born Hispanic of undetermined origin is likely to have the same background as the dominant group. We elaborated on this conventional design by building a historical geography based on the state residence of foreign and foreign-parentage Hispanics in our samples. By examining state settlement patterns across the period 1850–1970, we expanded the geographical framework of identification.

We began by creating five major subgroups and distributing Hispanics into categories on the basis of their—or their parents' or grandparents'—foreign birthplace. The subgroups defined by origin are Mexican, Spanish, Cuban, Puerto Rican, and other (the last indicating that national origin was known but was not among the four previous groups). Thus, all persons who were born in Spain, or who had a parent or grandparent born in Spain, were placed in the category Spanish, whereas those born in Ecuador were listed as other.¹⁸ Those not defined by birthplace variables

remained “undetermined.” The overall rate of classification at this stage depended on the universe of the parental-birthplace variable and on how large first-, second-, and third-generation groups were in the Hispanic population in that year. For example, we classified 78 percent of all Hispanics by birthplace variables in 1920, but only 72 percent in 1960.

In the second stage, we examined the states of residence of those who had been classified, assessing for each census year whether one group dominated the Hispanic population of a state. Several rules governed our decisions. We assumed that all unidentified Hispanics who lived in the five southwestern states were of Mexican origin.¹⁹ For the remaining states, if a majority of a state's classified Hispanic population belonged to one category, such as Puerto Rican, we classified the state as Puerto Rican in that year. If there were fewer than nineteen determined cases, or no national origin predominated, we considered the state undetermined. Thus, Massachusetts, with a large number of Hispanics but relatively equal percentages of national groups, was usually undetermined, as was Georgia, which reported a very small number of Hispanics.

Finally, national origin had to be dominant in at least two contiguous censuses (e.g., 1920 and 1940). Application of national identification by state of residence was then lagged ten, twenty, and thirty years, unless interrupted by a new national domination. The lagged effects assume that undetermined cases were the descendants of those earlier residents. A preponderance of Mexican residents in a state in 1900 and 1910 predicted the national origin of unclassified

TABLE 2
Estimates of Size of U.S. Hispanic Population, 1850–1990

Census year	Size of census sample	Total U.S. population	Consistent subsample estimates			Maximized subsample estimates		
			No. of Hispanics in sample (unweighted)	Est. no. of Hispanics (weighted)	% of U.S. population	No. of Hispanics in sample (unweighted)	Est. no. of Hispanics (weighted)	% of U.S. population
1850 ^a	197,678	19,600,000	1,158	117,107	0.60			
1880	502,913	50,155,783	3,946	393,555	0.78			
1900	100,425	75,994,575	663	503,189	0.66			
1910	461,756	91,972,266	71,789	797,994	0.87	74,421	897,685	0.98
1920	1,050,634	105,710,620	12,746	1,286,154	1.22	13,443	1,356,499	1.28
1940	1,351,732	131,669,275	23,402	2,142,716	1.63			
Sample record	186,010	131,669,275	5,130	2,021,820	1.54	5,700	2,245,143	1.71
1950	1,922,198	150,697,361	58,966	3,558,761	2.36			
Sample record	461,130	150,697,361	9,887	3,231,409	2.14			
1960	1,799,888	179,323,175	58,371	5,814,784	3.24	58,869	5,864,383	3.27
1970 (15%)	2,029,633	203,302,031	89,058	8,920,940	4.39	104,757	10,493,419	5.16
<i>Estimates based on Hispanic-origin questions in census</i>								
1970 (5%)	2,030,276	203,302,031	90,748	9,074,800	4.46			
1980 (1%)	2,267,320	226,545,805	147,202	14,707,759	6.49			
1990 (1%)	2,500,052	248,709,873	210,218	21,727,406	8.74			

Source: Gratton and Gutmann 1999.

Note: No maximized subsample estimates are reported for 1850, 1880, 1900, and 1950 because no variables are available beyond those used in the consistent subsamples. For 1940 and 1950, both person-record and sample-record subsamples have been produced.

^aFree population only.

cases in 1920 and 1940. This process created a geography of settlement by national origin between 1850 and 1970. This Hispanic historical geography can be seen in the appendix, which lists, state by state, national subgroup predominance, and records the censuses in which geographic identification was applied. That method raised the level of subgroup identification in 1920 from 78 percent to 94 percent and, in 1960, from 73 percent to 99 percent.

The historical geography extended designation of subgroup origin to many Hispanics in the third and higher generations. For example, empirical results showed that Mexicans were the only important Hispanic settlement group in Kansas. By 1950, many third-generation Hispanics in that state could not be linked directly to Mexico, but the pattern of prior settlement made it clear that they were likely to be descendants of Mexicans. We tested those rules, especially in the more ambiguous states, by selecting undetermined cases and examining national classification of family and household members, surnames, birthplaces, and so forth. The tests indicated that the process correctly identified likely subgroups.

Results: Size of Major Hispanic Subgroups

Application of the extended historical geography led to the estimates shown in table 3, which uses consistent subsample criteria. The table shows the weighted number of Hispanics in each group and their percentage distribution in the Hispanic population as a whole. For example, we estimated that in 1920 there were at least 999,535 persons of Mexican origin, who represented 77.7 percent of the Hispanic population in the United States.

The series shows the expected dominance of persons of Mexican origin across the period 1850–1990, as well as the decline in the Mexican proportion after 1950. It also reveals the surprising secondary position of Spain as a nation of origin in the Hispanic population through 1940 (Gomez 1963; Mormino and Pozzetta 1987), the steep rise in Puerto Rican

representation after World War II, and the new part played by Cuban Americans after 1960.

Correspondence with Other Estimates

Foreign-born and second-generation Hispanics

Until 1970, published census figures generally enumerated only foreign-born and, at times, second-generation Hispanics, often ignoring certain national groups. For the foreign born, we have very high confidence in the estimates produced from the IPUMS. For example, in 1910, we estimated 229,635 Mexican-born persons, whereas the census reported 221,915 (U.S. Bureau of the Census 1975). The census reported 15,133 foreign-born Cubans in the United States in 1910; our collection yielded a weighted estimate of 15,986. In 1950, the 6,859 Mexican-born Hispanics in our sample implied a weighted population of 456,901 Mexicans in the U.S. population; the full-count published census figure was 454,417. In 1960, the census estimated 44,999 Spaniards, and we calculated 44,622 (U.S. Bureau of the Census 1963).

Our estimates for the second generation (and hence of “foreign stock,” made up of the first and second generation) may differ from published figures. In assigning national origin to second-generation persons, the census generally excluded those of mixed foreign parentage, that is, those who had a father born in one foreign country and a mother in another. At other times (in the 1960 and 1970 censuses, for example), the census gave the father’s birth priority because it asked only his place of birth if both parents were foreign born. Our estimates followed the latter approach: in marriages in which both parents were foreign born, the father’s birthplace determined national origin. If the father was native born, the mother’s birthplace determined such origin.

The 1910 census indicated that there were 162,959 second-generation persons of Mexican origin (U.S. Bureau of the Census 1975); using the census method, we estimated 170,808. A census document published shortly after the

TABLE 3
National Origin of Hispanics, 1850–1970

Census year	Undetermined		Mexican		Cuban		Spanish		Puerto Rican		Other		Total
	%	n	%	n	%	n	%	n	%	n	%	n	
1850	23.4	27,407	69.6	81,508	1.8	2,124	4.2	4,955	0.0	0	1.0	1,113	117,107
1880	11.9	46,769	73.9	290,642	3.6	12,267	8.3	32,504	0.0	0	2.9	11,373	393,555
1900	2.9	14,421	79.8	401,491	4.4	22,006	9.4	47,055	0.0	0	3.6	18,216	503,189
1910	2.7	21,414	80.2	640,104	4.4	34,903	8.6	69,020	0.4	2,937	3.7	29,616	797,994
1920	5.5	70,919	77.7	999,535	2.8	35,809	9.3	120,042	1.6	20,384	3.1	39,465	1,286,154
1940 ^a	3.6	72,189	77.5	1,567,596	2.5	49,938	7.4	150,332	4.7	95,124	4.3	86,636	2,021,820
1950 ^a	2.9	93,145	77.0	2,489,477	2.2	70,919	4.2	134,659	10.1	326,186	3.6	117,023	3,231,409
1960	1.0	60,865	70.3	4,087,546	2.8	163,241	3.5	202,822	17.7	1,027,338	4.7	272,972	5,814,784
1970 (15%)	0.8	67,039	63.2	5,641,956	7.2	637,931	2.8	248,439	18.2	1,620,777	7.9	704,798	8,920,940

Source: Gratton and Gutmann 1999, consistent series.

^aSample-record estimates.

1920 census found 725,332 persons of Mexican birth or Mexican parentage living in the United States (U.S. Bureau of the Census 1922); we estimated 748,776 such persons. First- and second-generation Spaniards totaled 76,600 in that year; in our subsample, the weighted estimate was 85,143. In 1940, we estimated there were 15,769 white Cubans in the second generation (second-generation estimates for 1940 and 1950 use sample-record subsamples); the census listed 16,980 (U.S. Bureau of the Census 1945). Our estimate for Mexicans was 668,745 versus 699,220 in the census. In the 1960 census, first- and second-generation Spaniards totaled 126,163; in our subsample, the total was 128,409 (U.S. Bureau of the Census 1963). It is difficult to estimate persons of Puerto Rican ancestry from published census data before 1950; for 1940, Fitzpatrick (1980) used a figure of 69,967 but relied on first-generation reports only; our first- and second-generation estimate was 88,659.

Other Census Indicators

In 1920, the census reported that 556,111 foreign-born whites had Spanish as a mother tongue (U.S. Bureau of the Census 1922); we estimated 552,936. In 1950, 1960, and 1970, the census used a Spanish-surname question in the five southwestern states. The responses indicated that the following numbers of persons had Spanish surnames equivalent to those on the census list: 1950—2,281,710, 1960—3,513,684, and 1970—4,667,975 (U.S. Bureau of the Census 1973a). Weighted estimates from the Hispanic subsamples show the following: 1950—2,072,068, 1960—3,457,863, and 1970—4,481,037. Our estimates were consistently lower for good reason: We excluded Spanish-surnamed persons with a non-Hispanic foreign background (e.g., those who were born in Portugal or Italy). The differences reflect error in the census selection; such error is relatively small because of the high probability that persons in the Southwest were native born or from a Mexican, rather than Portuguese or other, background.

The 1970 Spanish American variable recorded 9,294,509 persons (U.S. Bureau of the Census 1973b); the weighted estimate from our subsample for this variable was similar—9,308,410. This specification clearly introduces commission error, classifying some persons by surname who are not Hispanic (hence, we excluded cases it added to our maximized sample). These various comparative results make us confident that subgroups defined by birthplace variables and other census measures constitute representative subsamples of their respective populations. We cannot, of course, control for errors intrinsic to the census itself, such as underenumeration (Willette et al. 1982).

Hispanic Population and Subgroups

Because we used both birthplace and surname to reach third- and higher-generation persons, our estimates of the

overall Hispanic population were higher than those based on birthplace and parental birthplace. In 1850 and 1900, and in the person-record samples for 1940 and 1950, however, we lacked critical variables. The 1850 PUMS did not report parental birthplace; hence our estimate of 117,107 Hispanics and 81,508 persons of Mexican origin is certainly too low. Nonetheless, it corresponds with previous measures. Citing Carey McWilliams's seminal study, *North from Mexico*, David J. Weber (1973) estimated that 75,000 Spanish-speaking persons lived in Arizona, California, New Mexico, and Texas in 1854. If we confine our sample to those states and territories, we find 78,982 Hispanics in the Southwest. Richard L. Nostrand (1992, 19) relied on the 1850 census returns and "estimates for several gaps" to find "80,302 Mexican Americans in the Southwest." He asserted that "a more realistic figure would be at least 100,000."²⁰

Oscar J. Martínez (1975) provided a range for the number of Chicanos, or Mexicans, between 1850 and 1900. Taking the 1850 census as a base for his lower bound, Martínez counted all white persons in New Mexico, California, Texas, and Arizona as of Mexican origin, added the small number of Mexican-born persons outside the Southwest, and made several other adjustments. He argued that non-Hispanic white intrusion into Mexican territories before the war was slight (certainly an error for Texas). For subsequent censuses, he added immigrants and estimates of birthrates for the immigrant and native populations. To calculate the upper bound, he multiplied all but the New Mexicans in 1850 by 1.4, a formula based on complaints made by Californians of census undercounts and an added effect for undercounting of minorities. New Mexicans received a 1.2 correction. (Results do not always correspond to these multiplying factors.) The resulting ranges were as follows: 1850—87,000 to 118,000, 1880—229,000 to 333,000, and 1900—381,000 to 562,000. Our estimates for 1850, 1880, and 1900 were as follows: 81,508, 290,642, and 401,491, respectively. Limitations in the variables available in 1850 and 1900 suggest that the Mexican population was probably considerably greater than we (and others) have estimated in those years. The 1880 estimate is closer, but, as we argue generally for all consistent samples, it still underestimates the actual population of Mexican background.

In 1910, our consistent subsample estimate of 797,994 Hispanics can be compared with published census figures equaling 477,508 for foreign-born and foreign-parentage persons from Mexico, Cuba (and West Indies), Spain, and Central and South America (U.S. Bureau of the Census 1922). Such published census data excluded Puerto Ricans, persons of mixed parentage (who had, for example, a Mexican father and a German mother), and third- and higher-generation Hispanics. In the innovative Hispanic Oversample of the 1910 U.S. Census of Population using all available indicators (e.g., language), Myron Gutmann, W. Parker Frisbie, and K. Stephen Blanchard (1999) estimated that about 845,000 Hispanics lived in the United

States and its territories in 1910. The maximized subsample, which used the same PUMS and Hispanic oversample, reported a similar figure, 897,685. By either measure, the number of Hispanics was certainly much higher than the foreign and foreign-parentage totals given in the census.

The Mexican race classification in the 1930 census provided a total of 1,422,533 Mexicans, plus 65,968 who were separately classified as "white" but were of Mexican stock (U.S. Bureau of the Census 1933).²¹ Although no PUMS for 1930 yet exists, the 1920 consistent subsample estimate for persons of Mexican origin nears 1 million (999,535). In 1940, we found more than 1.5 million Mexicans, despite repatriation during the 1930s. Given the volume of immigration and natural increase during the 1920s, the 1930 race classification surely missed a large number of Mexican-origin persons, a conclusion reached as well by other scholars (Grebler, Moore, and Guzman 1970).

Cary Davis, Carl Haub, and JoAnne L. Willette (1983, 1988) used backward projection from the 1980 Hispanic-origin count, mortality estimates, and U.S. Immigration and Naturalization Service immigration records for estimation. They calculated 4.0 million Hispanics in 1950 (versus our consistent estimate of 3.6 million) and 6.9 million in 1960 (substantially higher than the maximum sample estimate of 5.9 million). Their calculation for 1970, at 10.5 million, approximates our maximum estimate. Willette et al. (1982) detailed the projection methods used in the Davis, Haub, and Willette study and also provided estimates for subgroups, which suggested 2.8 million Mexican Americans in 1950, 4.6 million in 1960, and 6.6 million in 1970. Cubans were estimated at 0.2 million in 1950, 0.3 million in 1960, and 0.6 million in 1970. Puerto Ricans (calculated by actual census reports rather than reverse projection) constituted 0.3 million in 1950, 0.9 million in 1960, and 1.4 million in 1970. In 1950 and 1960, Willette et al.'s estimates of Mexican-origin and Cuban-origin persons were higher than ours, whether consistent or maximum. In all three census years, their estimates of Puerto Ricans, based on first- and second-generation data only, were lower.

F. G. Mittelbach and G. Marshall (1970) used the Spanish-surname variable in the southwestern states, information on foreign birth and parentage (foreign "stock") for the United States, and the ratios of foreign stock to more distant generations to estimate the total number of Mexican Americans in the United States. Their estimate for 1960, which did not use microdata samples, was 3,842,100, as opposed to the maximum estimate of 4,110,238. Using microdata samples, Bean and Tienda (1987) employed birthplace, parental birthplace, surname, Spanish as a mother tongue, and state of residence to identify Hispanics and then divide them into subgroups. Although they provided no numerical estimates, they calculated the proportionate weight of the subgroups in 1960 as 71.3 percent Mexican, 17.5 percent Puerto Rican, 2.3 percent Cuban, and 8.9 percent other. Their estimates mirrored ours (as presented in table 3) when

the categories of undetermined, Spanish, and other are all combined into "other."

The 1970 Comparison

The Spanish-origin variable provided in the 1970 census offered the first straightforward test of the validity of the consistent and maximized subsample approaches, both for estimates of the size of the entire Hispanic population and for its subgroups. Based on this variable (available only in the 5 percent sample), the Census Bureau reported 9.1 million Hispanics in the U.S. population (U.S. Bureau of the Census 1973c). Our consistent subsample criteria found 8.9 million persons to be Hispanic in origin, about 1.5 percent less than the Census Bureau's result. However, the 1970 Spanish-origin variable has been criticized for undercounting Hispanics.²² The 1973 Current Population Survey (CPS) (see U.S. Bureau of the Census 1974) found 10.6 million Hispanics, implying an improbable 16 percent increase in three years, most of it among Mexicans—a 40 percent increase (Bean and Tienda 1987). The maximized subsample estimate of 10.5 million for 1970 corresponds with CPS results.

The origin variable allowed estimation of national-origin subgroups. It indicated that 50 percent of Hispanics were Mexican, 15.8 percent Puerto Rican, 6 percent Cuban, and 28.3 percent other, respectively. Criticism of the design of the Spanish-origin variable, and of the very low percentage indicated for Mexican origin, led Bean and Tienda (1987) to decompose Hispanics into national groups using the 1970 (15 percent) sample that we employed, rather than the 1970 (5 percent) sample with the Hispanic-origin question.²³ Using birthplace and parental-birthplace variables, Bean and Tienda first determined known national origin. They then assigned categories to native-born Hispanics with native-born parents by means of a geographical scheme: Those in the five southwestern states were Mexican; those in New York, New Jersey, and Pennsylvania were Puerto Rican; and those in Florida were Cuban. Table 4 displays how this more attenuated geographical scheme compared with our own for 1970 (15 percent). In the second column,

TABLE 4
Comparison of National-Origin Estimates,
1970 (15%) Sample

Origin	Bean and Tienda (%)	Gratton and Gutmann (%)
Mexican	59.1	63.2
Puerto Rican	16.2	18.2
Cuban	6.7	7.2
Central and South American	6.7	7.9
Other	11.4	3.5

Sources: Bean and Tienda 1984; Gratton and Gutmann 1999, consistent series subsample.

we adjusted our national-origin estimates to the categories Bean and Tienda used, and we produced a good fit. The important difference is the rise in the percentage we could classify within the main groups and the decline in the unspecified percentage "other." In the wider geographical system we employed, more cases were assigned to likely national categories.

Examining the Consistent Subsample Series

In this section, we review the characteristics of the estimates produced by the consistent subsample series. We assess how subsamples for individual years differed from one another, and we examine potential sources of error. We first discuss the process of qualification, detailing the relative roles played by selection criteria in each PUMS. In each census year, differing proportions of persons are recorded as Hispanic because of surname, place of birth, and so on. We then turn to two broad categories of potential error in the consistent subsample series. We begin with errors of omission, those that keep individuals from being ascertained as Hispanic despite valid characteristics. We conclude by assessing errors of commission, which incorrectly identify persons as Hispanic.

Qualification

A variety of indicators linked to either birthplace or surname qualified cases for entry into the consistent subsample series. In each year, the contribution of qualifying factors depended on a variety of forces at work at the time of the census and in the preceding years: the level of immigration, the precise definition of the variables used, and the universe to which the variables applied. Although the last two factors

undoubtedly influenced the samples, the overall pattern of qualification mirrors the general immigration history of Hispanics. The proportionate effect of the fundamental criteria for qualification is displayed in table 5.

The qualifying criteria are as follows:

- foreign born: birth in a Hispanic country
- parent foreign born: at least one parent born in a Hispanic country
- grandparent foreign born: at least one grandparent born in a Hispanic country
- household head's country: relative of a head of household who had qualified by one of the above, or whose spouse had so qualified
- surname: native born with a native-born father who reported a Spanish surname
- household head's surname: relative of a head of household who had qualified by surname

The series is affected by differences in the availability of variables. In 1850, and in the person-record samples for 1940 and 1950, parental birthplace can only be ascertained for children living in their parents' households; no birthplace for grandparents can be determined. In 1960, parental birthplace is not available for the foreign born. In 1960 and 1970, only the father's foreign birthplace was reported. Such conditions reduce qualification by birthplace variables. In 1900, no surname identification was possible; in 1910, surnames were only available for the Hispanic oversample; and, in 1960 and 1970, the universe of the surname variable was restricted to the five southwestern states. Hence, in those years, surname effects were reduced, and the geographic limitation disproportionately selected Hispanic subgroups living in the Southwest. In the sample-record sub-

TABLE 5
Selection Criteria Effects, Hispanic Samples, 1850–1970: Unweighted Data

Census year	Foreign born		Parent foreign born		Grandparent foreign born		Household head's country		Surname		Household head's surname		Total
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	
1850	764	66.0	44	3.8	N.A.	N.A.	50	4.3	290	25.0	10	0.9	1,158
1880	1,801	45.6	621	15.7	148	3.8	565	14.3	788	20.0	23	0.6	3,946
1900	267	40.3	195	29.4	71	10.7	130	19.6	N.A.	N.A.	N.A.	N.A.	663
1910	32,911	45.8	22,501	31.3	4,139	5.8	2,225	3.1	9,894	13.8	119	0.2	71,789
1920	5,958	46.7	3,246	25.5	702	5.5	485	3.8	2,272	17.8	83	0.7	12,746
1940 ^a	6,012	25.7	7,266	31.0	N.A.	N.A.	925	4.0	8,804	37.6	395	1.7	23,402
Sample records	1,804	35.2	1,943	37.9	N.A.	N.A.	N.A.	N.A.	1,383	27.0	N.A.	N.A.	5,130
1950 ^a	11,554	19.6	14,994	25.4	N.A.	N.A.	1,310	2.2	30,074	51.0	1,034	1.8	58,966
Sample records	2,523	25.5	3,616	36.6	N.A.	N.A.	N.A.	N.A.	3,748	37.9	N.A.	N.A.	9,887
1960	14,732	25.2	17,487	30.0	10,078	17.3	3,019	5.2	12,526	21.5	529	0.9	58,371
1970 (15%)	26,274	29.5	26,648	29.9	13,162	14.8	4,825	5.4	17,287	19.4	862	1.0	89,058

Source: Gratton and Gutmann 1999, consistent series.

N.A. = not available.

^aAll records.

samples for 1940 and 1950, grandparents' birthplaces are not available, and we did not distribute the head of household's country or surname to other household members.

Despite those limitations, qualification follows what might be predicted from the history of Hispanic immigration to the United States. Selection by foreign birthplace declined after 1920, as depression and war reduced immigrant flows. Despite renewed immigration, the effects of foreign birthplace increased slowly after World War II because of the sharp natural growth in subsequent Hispanic generations: Parental birthplace and, when available, grandparents' birthplace revealed those new, native sources of growth. Surname's effects followed a similar pattern, rising when immigration was low, and information about parents and grandparents was lacking. Two categories, household head's country and household head's surname, assigned statuses to persons who would not have otherwise qualified. Such categories are rarely very large. Only in 1880 and 1900 did the head of household's country exceed 6 percent of the cases. In 1900, this was largely a product of the lack of surname identification tools. Household head's surname had even smaller effects.

Errors of Omission

A review of the consistent subsamples indicates that the birthplace and surname variables did not locate all persons likely to identify themselves as Hispanic. The difference between the consistent and maximized subsample estimates ranged from less than 1 percent to 18 percent. Differences depended not only on what additional classifying variables could be found but also on the underlying composition of the Hispanic population (e.g., the proportion consisting of recent immigrants).

Whereas the birthplace variables occasionally missed true Hispanics (e.g., persons with Hispanic heritage whose parents were born in a non-Hispanic country), we consider this error to be trivial. By far, the greatest amount of omission came from the Spanish-surname variable, which failed to register the surnames of persons who were clearly Hispanic.

TABLE 6
Rates of Omission: Persons Born in Hispanic Countries with Surnames Not Listed as Spanish, 1920

Country of origin	Born in Hispanic country, but surname not listed as Spanish (%)
Mexico	50.4
Cuba	65.7
Argentina	85.8
Spain	65.2
Puerto Rico	57.0
All Hispanic countries	54.0

Source: Gratton and Gutmann 1999, consistent series subsample.

Surname lists are known for omission error (Gratton, Rosales, and DeBano 1988; Perkins 1993), and their impact on the selection process increased in periods of low immigration (or when birthplace variables were lacking), when country indicators were less robust. Error rose the further back one applied lists devised in the 1970s and was more common in regions of the country with lower concentrations of Hispanics; the magnitude of error depended as well on the training and skills of enumerators and data-entry operators.²⁴

When we examine surname classification among foreign-born Hispanics, we ascertain the importance of surname omission. In 1880, 55 percent of all persons born in Hispanic countries did not have a Spanish surname recorded; in 1910, 43 percent did not; in 1920, 54 percent did not. In the 1940 person-record subsample, the overall rate of error for foreign-born Hispanics was 45 percent; it then dropped to 28 percent in the same subsample in 1950, aided by an additional surname variable in that year.²⁵ Finally, omission rates differed substantially by national origin (see table 6; see also Perkins 1993). Mexican-born persons, upon whose surnames the lists were based, tended to be more accurately reported.

Studying the characteristics of those omitted from the surname list, but born in a Hispanic country, we reached several conclusions. In general, when they differed, the country indicator was right and the surname indicator wrong. Other variables, especially language, identified as Hispanic those who had been omitted on surname lists. In 1920, although only 50 percent of the Mexican born reported a Spanish surname, 97 percent reported Spanish as a mother tongue. Among all persons chosen directly by birthplace, 95 percent reported Spanish as a mother tongue; if only those without a Spanish surname were selected, the percentage with Spanish as a mother tongue was 92 percent; similar results obtained in 1910, when 83 percent of all foreign born, and 80 percent of the foreign born without Spanish surnames, reported Spanish as a mother tongue. The results demonstrated that lack of surname is not a clear indicator of lack of Hispanic traits.

The 1920 census reported the mother tongue of all foreign-born parents. Of all Hispanics selected because one or both parents were foreign born, 93 percent had a parent with Spanish as a mother tongue. If we restricted this group to those without a Spanish surname, 90 percent had a parent, and 76 percent a father whose mother tongue was Spanish. In the 1940 sample-record subsample, the surname list failed to identify 38 percent of Mexicans, but 83 percent of those missed had lived as children in Spanish-speaking households. The list missed 55 percent of Cuban immigrants, but 42 percent of those Cubans reported Spanish as the language spoken in their childhood homes. Only 24 percent of the foreigners ascertained by country in 1940 did not report Spanish as a mother tongue.

A small number of omissions are justified, especially when persons of non-Hispanic background happened to have been born in a Hispanic country. Inspection of the

actual surnames (available in earlier PUMS), however, indicates that the dominating factor was quite different: Hispanic surnames were not being accepted because of the limitations of the list used, misspellings, enumerators' faulty recording of obviously Hispanic surnames, or the ambiguity of the surname itself. For example, the following surnames from Colorado in 1880 were not returned as Spanish: Valazuella, Escolosa, Slazar, Mendis, Vasquer. In 1920, Rodrigues, Rodregez, Martineas, Chaves, Lopes, and Rio were not accepted. Not only is it highly likely that they were misspellings of actual Spanish surnames (e.g., Valenzuela and Salazar), but other characteristics—such as the given name, language, and undeniably Hispanic background of remaining household members—also attest to the Hispanicity of omitted cases. In 1910, in South Texas, the surname for the Voldez family was not returned as Spanish; the given names in the household were Jose, Jesus, Refugio, and Pedro, and the street upon which they lived was Alto Vonto. Such errors are obviously more important for Hispanics who have resided in the United States for a longer time and for whom the country criteria do not apply.

Whereas country variables can repair omission errors for those with a foreign background, the surname list is critical to reaching the native born, especially those who have lived in the United States for several generations. Omission error is most important for this group; an unknown number of native-born Hispanics were not recorded by surname. The impact of this error increased in periods of low immigration when the country indicator was less dominant and in early censuses when current surname lists became less useful. The extension of country and surname through the status of the household head helps to reduce such omission error.

The overall level of omission may be roughly gauged by assessing how well consistent subsample variables could select and identify persons that a maximized sample variable will find. In 1910, the consistent variables identify 97 percent of persons recorded as being of Spanish "race." In 1920, consistent subsample indicators located 95 percent of all foreigners who spoke Spanish as a mother tongue; in the same year, they selected 74 percent of all persons listed to be of Spanish race. In 1940, consistent selection procedures captured 85 percent of native-born persons for whom Spanish was the household language when they were children. In 1960, they located 97 percent of all foreigners who spoke Spanish as a mother tongue. In the 1970 census, native-born persons reported the language spoken by the head (or spouse) of the household in which they had lived as children. Consistent subsample procedures located 86 percent of the cases that reported Spanish as a mother tongue. If language is taken as the best single criterion variable other than self-identification, the consistent subsample series captures most but not all Hispanics.²⁶

Comparison with the Hispanic-origin self-identification variable available in 1970 suggests a low level of error. In the 15 percent subsample of that year, the consistent sub-

sample variables located 8,920,940 persons, whereas the Hispanic-origin variable available in the 5 percent sample reported 9,074,800—a difference of 1.7 percent. (The census itself reported 9,072,602 [U.S. Bureau of the Census 1973c].) However, as we have previously shown, the first attempt at Hispanic self-identification clearly underestimated their number. The addition of other criteria, especially language, taps such groups; the maximized subsample estimate at 10.5 million very closely approximates the Hispanic population as found in 1973 CPS surveys.

In summary, consistent subsample selection procedures have the advantage of consistency across the censuses, but they miss native-born Hispanics without common Spanish surnames. Many of the omitted persons are picked up in maximized samples.

Errors of Commission

The subsamples may also incorrectly classify persons as Hispanics: children may enter by surname, despite having a grandfather born in a non-Hispanic country. For the most part, we find that errors of commission are relatively few, and rarely clear cut, so that dubious cases could well be considered Hispanic. The decision about whether or not to include such cases will depend on each investigator's perspective and research objectives. To facilitate such decisions, we offer an evaluation of five categories of potential error:

1. Persons born in a Hispanic country, neither of whose parents was born in the listed nations; these persons range from 1 percent to 6 percent of the cases in the consistent series.²⁷ For example, in 1920, less than 2 percent of all cases had this characteristic. Of those, 27 percent reported Spanish as a mother tongue. In the 1940 sample-record subsample, about 5 percent of all cases had this characteristic: 29 percent reported Spanish as a mother tongue (of those with language information), and 37 percent of their fathers had missing or blank birthplace information. In 1970, 2.4 percent of the Hispanic foreign born had neither parent born in such countries; 61 percent reported Spanish as a mother tongue. These rates included persons born in Puerto Rico whose parents were not reported as born on the island. Such instances artificially inflate error, adding cases in which parents were born on the mainland, returned to Puerto Rico, and had children. There were also a large number born in Puerto Rico whose parents' birthplaces were listed as "United States, n.s." (not specified), "Abroad, n.s.," or "Caribbean." In 1970, 44 percent of the fathers' birthplaces were "United States, n.s." Other parental birthplaces—in the Southwest, Florida, and New York—suggested that parents were returning to Hispanic countries. In general, researchers should exercise caution in eliminating cases that lack Hispanic parental birthplace and should review other characteristics before exclusion.

2. Persons born in non-Hispanic foreign countries, none

of whose parents or grandparents was born in a Hispanic country. (In several PUMS—for example, 1960—the number of such cases cannot be ascertained.) Brought in by family connections, this group averaged less than 1 percent in the subsamples. In 1920, there were 99 such cases; in 1940, in the sample-record subsample, there were none, and in the person-record subsample, 218. In 1970, when the 775 individuals represented less than 1 percent of the cases, 26 percent reported Spanish as a mother tongue, and 26 percent had fathers with unspecified birthplaces.

3. Relatives, other than children, who qualify by the Hispanic country of the household head or spouse. The possibility exists that a German-born man might be considered Hispanic if his father-in-law was born in Mexico. Extension of Hispanic status to his co-resident mother, or to a cousin living in the household, may be equally inappropriate. However, family designation not only allows selection of children who would otherwise be lost to the sample, but in endogamous marriages it identifies spouses and other relatives of more distant Hispanic origin. Automatic exclusion of all spouses and collateral relatives loses these cases. Rather than fathers-in-law, grandparents, and other more distant relatives, the great majority of relatives other than spouses are grandchildren, nieces, nephews, adopted children—that is, subsequent generations. There is good reason to consider such family members as recipients of a Hispanic heritage. In 1910, of 2,225 cases, 833 were children. The 1,392 remaining constituted 2 percent of the consistent sample. Spouses represented 67 percent of these relatives; grandchildren 10 percent; adopted children, nieces, and nephews another 10 percent. Together, parents, parents-in-law, and siblings of the household head summed to a similar percentage. In response to a question seeking information on whether respondents could speak English at all, among these relatives, 29 percent reported that they could speak *only* Spanish.

In 1920, 4 percent were selected in this manner, and three-quarters of them were spouses. Many relatives reported birthplaces and parental birthplaces in the southwestern states, but about one-fifth had foreign backgrounds and might have been excluded. In the 1940 person-record sample, 4 percent of the cases had those characteristics, and two-thirds were spouses. Grandchildren constituted 8 percent, and a variety of “in-law” categories another 11 percent. Hispanic states led birthplace distribution. In 1960, spouses provided more than 80 percent of the cases—southwestern states or “United States, n.s.” dominated all birthplace distributions. These results suggest considerable endogamy, in which household members were Hispanic even when primary indicators were missing. Foreign birthplace or in-law status may more clearly imply a non-Hispanic background, subject to exclusion.

4. Persons who do not have a Spanish surname but who qualify as relatives of a head of household who does. These cases do not reach 2 percent in any subsample. In the main, they expose omission errors in the surname list, or the loss

of surname by marriage, because large numbers are children, stepchildren, and grandchildren of the household head. In 1880 and 1910, about half had given names and surnames that were clearly Spanish, although not reported on the surname list. Indeed, in 1910, the overwhelming majority of such cases followed directly from surname error: in that year, 40 percent could speak *only* Spanish. Birthplace and parental birthplace information in the latter two censuses revealed origins in New Mexico and Texas, although a few cases had foreign backgrounds. In 1970, 34 percent of the cases reported Spanish as a mother tongue, another 34 percent English, and 21 percent no language reported.

5. Persons of races other than white. Because Hispanics can be of any race, an automatic disqualification rule would be erroneous. Nonwhite cases represent about 14 percent of all those from 1880 to 1910 (the 1850 census did not count slaves) and less than 7 percent after that date. Blacks, mulattos, Indians, and “other,” appear in nontrivial numbers in the earlier censuses. Spanish-surname qualifiers occasionally selected African Americans and Asians whose Hispanic ancestry was dubious, and they regularly selected Filipinos and American Indians (Perkins 1993). Although Filipinos are a small number in the subsamples, they sometimes represented very large proportions of that nationality in the PUMS. The Filipinos often reported Spanish as a mother tongue—Indians less frequently. In 1880, 5 percent of the cases were mulattos—more than half lived in Colorado and were certainly of Mexican origin (in Colorado, enumerators used an M to denote that “racial” status).

Most nonwhite cases, however, qualify by country, and a large proportion reported Spanish as a mother tongue and surname. In earlier PUMS, Indians often qualified by country (primarily Mexico); in 1920, 33 percent of the Indian cases had been born in Mexico. For Indians, the mother-tongue and parental mother-tongue variables provide evidence that researchers may use for exclusion rules—about 15 percent reported Indian languages in 1920, but a larger percentage reported Spanish. In 1850, about half of the twenty-two mulattos listed had Spanish given names or qualified by country. In 1910, 76 percent of blacks qualified by country (primarily Cuba), 39 percent being foreign born. Most had Spanish surnames and given names, and most reported Spanish on mother-tongue variables. Such Hispanic characteristics are still more pronounced for mulattos. In 1920, among blacks and mulattos, about 50 percent qualified by birthplace or their parents’ or grandparents’ birthplace: more than one-third of their fathers had Spanish as a mother tongue. Similar characteristics appeared in 1970. Some blacks from interior southern states qualified solely by surname and did not exhibit other Hispanic characteristics.

Hispanics identified as “other” race constitute a large proportion of all such cases in the PUMS and were a particularly important group in 1900 and 1910. Likely to be of Mexican origin, they reported Spanish-language characteristics. In 1910, 96 percent qualified by country: 55 percent

had been born in a Hispanic country, over half in Mexico. About 60 percent spoke only Spanish. Most cases labeled as "other" appeared to have Hispanic backgrounds. Whereas researchers may wish to exclude certain cases, the results strongly recommend that there be no blanket exclusion by race.

The review of the five categories of possible errors of commission suggests both a relatively small magnitude as well as the need for researchers to consider carefully whether exclusion is justified. If we use 1920 as an example—excluding the first four categories and also removing all Indians and Asians—we reduce the subsample size by 7 percent. However, 11 percent of these cases reported Spanish as a mother tongue, and others had parents who spoke Spanish. Many were born in the southwestern states. Even among those instances, country qualification, relationship variables, and first and last names indicated Hispanic characteristics. Among those not reporting Spanish as a mother tongue, one notes the following given names—Canato, Carlos, and Cresencias—and the following surnames—Villagas, Sobrero, and Placencio. Researchers should use the identifying variables available in each census before deciding to eliminate cases.

Conclusion

The estimates of the size of the Hispanic population of the United States, and of its constituent subgroups, fill important lacunae in our immigrant and ethnic history. The value of the data lies, however, in the subsamples on which the estimates have been based. They provide a new and representative source for research on the diverse Hispanic population. Consistent subsamples underestimate that population, but they provide comparable data for research across the period 1850 to 1970. Maximized subsamples are not always available, and selection variables are not consistent. But such samples appear to compare well with those selected by Hispanic-origin variables and come closer to representing the full Hispanic population.

Each series offers scholars a new avenue for research based on broadly based census data, rather than purposive or convenience samples. National in scope, and larger in size than data sets heretofore available, the series permit analyses of Hispanic subgroups and more reliable comparisons between them and other populations. As such, they promise to underwrite more accurate and valid research on the place of Hispanics in the history of the United States.

APPENDIX
Predominant Country of Origin, by State, 1850–1970 (15%)

State of residence	Predominant country of origin for persons of known origin		Country of origin assigned to persons of unknown origin ^a	
	Country of origin	Year	Country of origin	Year
Alabama	Spain	1880		
Alaska	Mexico	1960		
Arizona	Mexico	1880–1970	Mexico	1880–1970
Arkansas	Mexico	1920, 1950, 1970	Mexico	1960, 1970
	Spain	1940		
California	Mexico	1850–1970	Mexico	1850–1970
Colorado	Mexico	1880–1970	Mexico	1880–1970
Connecticut	Spain	1920	Puerto Rico	1970
	Puerto Rico	1960, 1970		
Delaware	Puerto Rico	1970		
District of Columbia	Other	1940, 1960, 1970	Other	1950–1970
Florida	Cuba	1880–1970	Cuba	1900–1970
Georgia				
Hawaii	Spain	1910		
	Puerto Rico	1960, 1970	Puerto Rico	1970
Idaho	Spain	1920, 1940	Spain	1940, 1950
	Mexico	1950–1970	Mexico	1960, 1970
Illinois	Spain	1880		
	Mexico	1940–1970	Mexico	1950–1970
Indiana	Mexico	1940–1970	Mexico	1950–1970
Iowa	Mexico	1940–1970	Mexico	1950–1970
Kansas	Mexico	1910–1970	Mexico	1920–1970
Kentucky	Mexico	1880		
	Spain	1940		
	Puerto Rico	1960		
Louisiana	Spain	1850, 1880, 1920	Spain	1880–1940
Maine				

(appendix continues)

APPENDIX—Continued

State of residence	Predominant country of origin for persons of known origin		Country of origin assigned to persons of unknown origin ^a	
	Country of origin	Year	Country of origin	Year
Maryland	Cuba	1850		
	Spain	1880		
Massachusetts	Spain	1920		
	Puerto Rico	1970		
Michigan	Mexico	1940–1970	Mexico	1950–1970
Minnesota	Mexico	1940–1970	Mexico	1950–1970
Mississippi	Spain	1880, 1910	Spain	1900–1920
	Mexico	1950		
	Other	1960		
Missouri	Mexico	1920–1970	Mexico	1940–1970
Montana	Mexico	1940–1970	Mexico	1950–1970
Nebraska	Mexico	1940–1970	Mexico	1950–1970
Nevada	Mexico	1920		
	Spain	1940		
	Mexico	1950–1970	Mexico	1960, 1970
New Hampshire				
New Jersey	Other	1880		
	Spain	1940		
	Puerto Rico	1960, 1970	Puerto Rico	1970
New Mexico	Mexico	1850–1970	Mexico	1850–1970
New York	Puerto Rico	1940–1970	Puerto Rico	1950–1970
North Carolina				
North Dakota				
Ohio	Mexico	1950		
Oklahoma	Mexico	1910–1970	Mexico	1920–1970
Oregon	Mexico	1880		
	Spain	1920		
	Other	1950		
Pennsylvania	Puerto Rico	1970		
Rhode Island	Spain	1940		
South Carolina	Spain	1950		
South Dakota	Puerto Rico	1970		
Tennessee	Other	1950		
	Mexico	1960		
Texas	Mexico	1850–1970	Mexico	1850–1970
Utah	Mexico	1920–1970	Mexico	1940–1970
Vermont	Spain	1960		
	Other	1970		
Virginia				
Washington	Mexico	1950–1970	Mexico	1960, 1970
West Virginia	Spain	1940–1960	Spain	1950–1970
Wisconsin	Mexico	1940, 1950, 1970	Mexico	1950–1970
Wyoming	Mexico	1920–1970	Mexico	1940–1970

Sources: Gratton and Gutmann (1999); see also text.

^aBlanks indicate that no country was assigned.

NOTES

We are grateful for the assistance provided by the National Institute of Child Health and Human Development through grants 1 F33 HD08147-0 (Brian Gratton), and 5 R01 HD32325 (Myron Gutmann). The work was begun while Gratton was an NIH postdoctoral fellow at the Population Research Center, the University of Texas at Austin.

1. We have prepared ASCII data files that can be merged by individual and family serial number with the 1998 versions of the master IPUMS files located at <http://www.ipums.umn.edu>

The Hispanic data files and instructions for their use can be found at <http://www.prc.utexas.edu/hispanic/>

Inquiries regarding the subsamples and their use should be directed to Brian Gratton, Dept. of History, Arizona State University, Tempe, AZ 85287-2501, 480-727-6118, e-mail: brian@asu.edu

2. If not otherwise cited, information regarding variables may be found

in Ruggles and Sobek (1997a). The IPUMS data come from Ruggles and Sobek (1997b).

3. We follow the conventional scholarly practice of identifying Puerto Ricans as a separate group, although they have been U.S. citizens since 1917 by virtue of the Jones Act (Davis, Haub, and Willette 1988). Puerto Rico itself is not included in the IPUMS data; only Puerto Ricans on the mainland are part of the Hispanic samples.

4. Bean and Tienda (1987, chap. 2) reviewed census variables through 1980 that may be used to identify Hispanics.

5. The surname identification for the Southwest is found in the IPUMS variable "Spanish surname, 1950 basis" or span95. In the PUMS for 1940 and 1950, the U.S. Bureau of the Census and the Center for Demography and Ecology at the University of Wisconsin applied a different list to all persons (see Ito, Gratton, and Wycoff 1997).

6. Country-of-birth lists changed as war, diplomacy, and refined Census Bureau techniques added and eliminated nations. National boundaries also

shifted. This historical process had minor effects on countries we identified as Hispanic. Changes in universe for parental birthplace presented more serious difficulties. In 1850, parental birthplace was not recorded, and, in 1940 and 1950, it was reported only for a subsample of persons. In 1960, it was not reported for the foreign born, and in that year and 1970, when both parents were foreign born, only the father's birthplace was reported. Those changes affected both selection procedures and results, as we discuss in our article.

7. In the 1900 PUMS, Spanish surname was not recorded. In 1910, it was added in the Hispanic oversample for persons in fifty-seven counties in Arizona, California, New Mexico, Texas, Kansas, and Florida (Gutmann et al. 1997). In 1960 and 1970, it was reported only for persons residing in Arizona, California, Colorado, New Mexico, and Texas. In 1950, both the original list, applied to the five southwestern states, and a national list were available.

The set of Spanish surnames used was not perfectly consistent. Jeffrey Passel and David Word's (1980) Census Bureau list is the foundation for IPUMS applications from 1850 through 1950. In 1850, 1880, and 1910, we augmented their list with surnames identified by other researchers as clearly Spanish (Forbes and Frisbie 1991); in the 1910 Hispanic oversample, coders were instructed to accept clearly Spanish surnames. In 1940 and 1950, persons surnamed Martin, Roman, Santos, or Silva also received the designation if a parent was born in a set of Hispanic countries (Ruggles and Sobek 1997a). In the southwestern lists used in 1950, 1960, and 1970, the census applied predecessors of the Word and Passel list.

8. The code for this procedure identified the birthplace of maternal and paternal grandparents through the parental birthplace variables on a child's mother's and father's records. Such a procedure can only identify grandparents for children living in households with their parents. Hence, it is not an unbiased identifier of all third-generation persons. For sample-record children in 1940 and 1950, the parental record did not provide data for grandparents.

9. Most census data contain a household record for information pertaining to all household members and a person record for each individual. In 1940 and 1950, enumerators collected additional information for individuals who appeared on a designated sample line of the enumeration form, creating a sample record in those years. See Ruggles and Sobek (1997a).

10. This approach was dictated by the peculiar sampling schemes used in the PUMS for 1940 and 1950. See Ito, Gratton, and Wycoff (1997).

11. The Census Bureau administered two long-form schedules in 1970, one to 5 percent of the population (called Form 1 by the Census Bureau), and another to 15 percent of the population (called Form 2). The questions asked were different. Whereas the 5 percent form asked respondents if they were of Hispanic origin, the 15 percent form had questions about parental birthplace (Ruggles and Sobek 1997a). Because the 15 percent form provided information comparable to those in earlier censuses, we used it in our series and reported results from it. The 1970 5 percent schedule begins the series relying on self-identification.

12. As noted previously, the spanamer indicator combines a number of variables to identify Spanish Americans. Although we included this variable in the maximized selection procedure, the small number of cases it adds is produced by commission error, and we removed them from our 1970 estimate.

13. Self-enumerated since 1960, the race question had various degrees of input from respondents before that date.

14. The small number of cases after 1910 and access to language variables in 1920 led us to set 1910 as the last census for special treatment.

15. Previous research on the Hispanic population in this region (Nostrand 1992; Martínez 1975), which distinguished between Hispanic and non-Hispanic residents in former Mexican territories, also suggested direct examination of the data.

16. In California and Arizona, such cases were often Indians; given that Hispanics can be of any race, we elected to retain these cases in the Hispanic samples. We discuss race and Hispanic origin in the text, with specific reference to American Indians.

17. D. W. Meinig (1986, 39) stated that, by 1830, "at least three-quarters of the 25,000–30,000 people in Texas (excluding Indians) were Anglo-Americans (including Black slaves)."

18. In the event of a mixed marriage (a father born in Mexico and a mother in Cuba, for example), we gave precedence to the father's place of birth. This procedure accorded with census practice in the 1960 and 1970 censuses, in which, if both parents were foreign born, only the father's place

of birth was reported. For the other census years, investigators may make maternal birthplace dominate if they so choose.

19. This decision was confirmed by empirical results, although New Mexico and Colorado often had very high proportions of Hispanics of distant, hence "undetermined," origins. We recognize the tradition of a separate *norteño* culture in portions of these states, and researchers may want to distinguish "Hispanos" from those with more recent connections to Mexico. However, continued Mexican immigration into these states indicates that Mexican designation remains the best initial choice (Cortés 1980).

20. The region around El Paso, Texas, was not enumerated in 1850 because of jurisdictional squabbles between Texas and the New Mexico Territory.

21. Abraham Hoffman (1974) followed this estimate. See also Grebler, Moore, and Guzman (1970) for errors in interpretation.

22. The consensus is that the 1970 variable underestimated the Hispanic (especially the Mexican) population (see Bean and Tienda 1987). Modification in the wording of the Hispanic-origin variable in the 1980 census, publicity about its use, and its inclusion in the full census schedule led to sharp increases in estimates of the number of Mexicans, in particular, and Hispanics, in general. The increase indicated that if the 1970 data were accurate, the overall rate of growth in the Hispanic population had been 60 percent across the decade, five times the rate of the general population. Were the 1970 estimate of Mexican-origin persons valid, then the Mexican population had indeed risen by 93 percent.

23. No direct comparison is possible: parental birthplace and mother tongue are only available in the 15 percent sample, whereas the Hispanic-identification variable is only included in the 5 percent sample.

24. Perkins (1993) tested the Word-Passel (1980) list in a data set that also had self-reported Hispanic origin. In addition to geographic and national origin bias, he noted the familiar absence of women who lose their surname by marriage. Among persons reporting Hispanic origin and surname in 1990, the Word-Passel list had an omission rate of 21 percent and a commission rate of 10 percent. Perkins concluded that surname lists capture Hispanic populations well for sex, age, and household relationship but show greater disparities for geographic location and race.

25. In 1960 and 1970, rates of omission fell substantially (11 percent and 20 percent, respectively), but only the surnames of persons in the five southwestern states were reported.

26. Many Hispanics did not grow up in homes where Spanish was the dominant language, and they do not speak it. In 1940, selection by language alone would have missed 22 percent of persons identified by other indicators; the same proportion would have been missed in 1970.

27. The 1850, 1960, and person-record samples in 1940 and 1950 lack complete parental birthplace information, so these characteristics cannot be fully assessed.

REFERENCES

- Bean, F. D., and M. Tienda, 1987. *The Hispanic population of the United States*. New York: National Committee for Research on the 1980 Census, Russell Sage.
- Cardoso, L. A. 1980. *Mexican emigration to the United States, 1897–1931*. Tucson: University of Arizona Press.
- Cortés, C. E. 1980. Mexicans. In *Harvard encyclopedia of American ethnic groups*, edited by S. Thernstrom, A. Orlov, and O. Handlin, 697–719. Cambridge: Harvard University Press.
- Daniels, R. 1990. *Coming to America: A history of immigration and ethnicity in American life*. New York: HarperCollins.
- Davis, C., C. Haub, and J. Willette. 1983. U.S. Hispanics: Changing the face of America. *Population Bulletin* 38, No. 3. Washington, D.C.: Population Reference Bureau.
- . 1988. U.S. Hispanics: Changing the face of America. In *The Hispanic experience in the United States*, edited by E. Acosta-Belén and B. R. Sjöström, 3–56. New York: Praeger.
- Fitzpatrick, J. P. 1980. Puerto Ricans. In *Harvard encyclopedia of American ethnic groups*, edited by S. Thernstrom, A. Orlov, and O. Handlin, 849–67. Cambridge: Harvard University Press.
- Forbes, D., and W. P. Frisbie. 1991. Spanish surname and Anglo infant mortality: Differentials over a half-century. *Demography* 26: 639–60.
- Gomez, R. A. 1963. Spanish immigration to the United States. *The Americas* 19: 59–78.
- Gratton, B., F. A. Rosales, and H. DeBano. 1988. A sample of the Mexi-

- can American population in 1940. *Historical Methods* 21: 80-87.
- Gratton, B., and M. P. Gutmann. 1999. Hispanic subsamples from the 1998 integrated public use microdata samples. Data File: <http://www.prc.utexas.edu/hispanic/>
- Grebler, L., J. W. Moore, and R. C. Guzman. 1970. *The Mexican American people: The nation's second largest minority*. New York: Free Press.
- Gutmann, M. P., W. P. Frisbie, and K. S. Blanchard. 1999. A new look at the Hispanic population of the United States in 1910. *Historical Methods* 32: 5-19.
- Gutmann, M. P., S. Ruggles, W. C. Block, C. A. Fitch, T. K. Gardner, P. K. Hall, D. Kallgren, and M. Sobek. 1997. The Hispanic Oversample of the 1910 U.S. Census of Population: User's guide. Minneapolis: Department of History, University of Minnesota.
- Gutmann, M. P., R. McCaa, R. Gutierrez-Montes, and B. Gratton. 1999. The demographic impact of the Mexican Revolution in the United States. Texas Population Research Center Papers No. 99-00-01. Austin, Tex.: Population Research Center.
- Hoffman, A. 1974. *Unwanted Mexican Americans in the great depression: Repatriation pressures, 1929-1939*. Tucson: University of Arizona Press.
- Ito, R. M. II, B. Gratton, and J. Wycoff. 1997. Using the 1940 and 1950 public use microdata samples: A cautionary tale. *Historical Methods* 30: 137-47.
- Lavrin, A. 1996. Personal communication with B. Gratton.
- Longmore, T. W., and H. L. Hitt. 1944. A demographic analysis of first and second generation Mexican population of the United States: 1930. *Southwestern Social Science Quarterly* 24:138-49.
- Martínez, O. J. 1975. On the size of the Chicano population: New estimates, 1850-1900. *Aztlan* 6: 50-56.
- McWilliams, C. 1990. New ed. updated by M. S. Meier. *North from Mexico: The Spanish-speaking people of the United States*. New York: Praeger.
- Meinig, D. W. 1986. *The shaping of America: A geographical perspective on 500 years of history*. Vol. 2. New Haven, Conn.: Yale University Press.
- Mittelbach, F. G., and G. Marshall. 1970. An estimate of the number of Mexican Americans in 1960. In *The Mexican American people: The nation's second largest minority*, edited by L. Grebler, J. W. Moore, and R. C. Guzman, 605-7. New York: Free Press.
- Mormino, G. R., and G. E. Pozzetta. 1987. *The immigrant world of Ybor City: Italians and their Latin neighbors in Tampa, 1885-1985*. Urbana, Ill.: University of Illinois Press.
- Nostrand, R. L. 1992. *The Hispano homeland*. Norman: University of Oklahoma Press.
- Passel, J. S., and D. L. Word. 1980. Constructing the list of Spanish surnames for the 1980 census: An application of Bayes' theorem. Paper presented at the Annual Meeting of the Population Association of America, Denver, 10-12 April.
- Perkins, R. C. 1993. Evaluating the Passel-Word Spanish surname list: 1990 decennial census post enumeration survey results. Population Division Working Paper No. 4. U.S. Census Bureau. <http://www.census.gov/population/www/documentation/twps0004.html>
- Quintana, F. L. 1980. The Spaniards. In *Harvard encyclopedia of American ethnic groups*, edited by S. Thernstrom, A. Orlov, and O. Handlin, 948-50. Cambridge: Harvard University Press.
- Ruggles, S. and M. Sobek. 1997a. With C. A. Fitch, P. K. Hall, and C. Ronnander. *Integrated public use microdata series. Version 2.0, vol. 1: User's guide*, edited by C. A. Fitch. Minneapolis: Historical Census Project, University of Minnesota.
- . 1997b. *Integrated public use microdata series. Version 2.0. vol. 3: Counting the past*, edited by P. K. Hall. Minneapolis: Historical Census Project, University of Minnesota.
- Samora, J., and P. V. Simon. 1977. *A history of the Mexican American people*. Notre Dame, Ind.: University of Notre Dame Press.
- U.S. Bureau of the Census. 1922. *Fourteenth census of the United States taken in the year 1920: Vol. 2. Population. General report and analytical tables*. Washington, D.C.: Government Printing Office.
- . 1933. *Fifteenth census of the United States: 1930. Population. Vol. 2. General report*. Washington, D.C.: Government Printing Office.
- . 1945. *Statistical abstract of the United States: 1944-45*. Washington, D.C.: Government Printing Office.
- . 1953. *U. S. Census of Population: 1950. Vol. 4. Special reports, No. 3C. Persons of Spanish surname*. Washington, D.C.: Government Printing Office.
- . 1963. *U.S. Census of Population: 1960. Final report PC(1)-1D. Detailed characteristics. U.S. summary*. Washington, D.C.: Government Printing Office.
- . 1973a. *1970 Census of Population. Subject reports. Persons of Spanish surname*. Washington, D.C.: Government Printing Office.
- . 1973b. *1970 Census of Population. Vol. I. Characteristics of the population. Part 1. U. S. summary. Sect. 2*. Washington, D.C.: Government Printing Office.
- . 1973c. *1970 Census of Population. Subject reports, PC(2)-1c. Persons of Spanish origin*. Washington, D.C.: Government Printing Office.
- . 1974. *Current Population Reports, series P-20, no. 264, table A. "Persons of Spanish origin in the United States: March 1973"*. Washington, D.C.: Government Printing Office.
- . 1975. *Historical statistics of the United States: Colonial times to 1970. Part 1*. Washington, D.C.: Government Printing Office.
- Weber, D. J., ed. 1973. *Foreigners in their native land: Historical roots of the Mexican Americans*. Albuquerque: University of New Mexico Press.
- Willette, J., R. Haupt, C. Haub, L. Bouvier, and C. Davis. 1982. *Demographic and socioeconomic characteristics of the Hispanic population in the United States, 1850-1980*. Report submitted to the Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services. Washington, D.C.: Development Associates and the Population Reference Bureau.