TECHNICAL WORKING PAPER NO. 13
March 1996

Building a Spanish Surname List for the 1990's— A New Approach to an Old Problem

by
David L. Word and R. Colby Perkins Jr.

Population Division
U. S. Bureau of the Census
Washington D.C.

The data and results appearing in this working paper were originally introduced at the Annual Meeting of the Population Association of America (PAA) Miami, Florida May 1994.

The views expressed in this paper are solely attributable to the two authors and do not necessarily reflect the position of the United States Bureau of the Census.

ABSTRACT

The United States Census Bureau produced and released Spanish surname products for 1950, 1960, 1970 and 1980. This 1990 version is another way station in an ongoing research journey. This paper, "Building a Spanish Surname List for the 1990's—A New Approach to an Old Problem," differs from its predecessors in two significant respects.

(1) Until 1990, name has never been part of a permanent Census electronic record. Following the 1990 Census, the Census Bureau appended name to 7 million Census records for the purposes of determining undercount. The "List" is constructed by tabulating the responses (surname by surname) to the Spanish origin question for persons in that sample. Well over 90 percent of male householders with the surnames: GARCIA, MARTINEZ, RODRIGUEZ, and LOPEZ responded affirmatively to the Spanish origin question while less than 1.0 percent of male householders named SMITH, JOHNSON, and BROWN provided a positive response to the Spanish origin question.

(2) In the past, a name was either on the list (e.g., Garcia) and was taken to be Spanish or it did not appear on the list. The assumption was that any name not on the list was not Spanish. Since neither BROWN nor SILVA appeared on the 1980 Spanish Surname list, one would naturally assume that neither name was Spanish. In the electronic version of the 1990 "List" we append auxiliary data for 25,000 surnames including both SILVA and BROWN that allow users to form their own lists. Almost 60 percent of the SILVA's in our 1990 Census sample responded that they were Hispanic while less than 1 percent of BROWN's claimed to be Hispanic. Moreover, another auxiliary item suggests that the letters S I L V A form a potentially Spanish word. That same statement cannot be made for B R O W N. From this data, some users might include SILVA on their own personal Spanish surname list, while others would justifiably arrive at an opposite conclusion.

We must emphasize that this product does not violate the confidentiality of Census responses. On average, each captured surname represents about 40 householders. Moreover, we provide no subnational geographic data nor is there any indication of first name or age of respondent. Given these conditions, we are confident that this file does not provide information that could identify any individual enumerated in the 1990 Census.

ACKNOWLEDGEMENTS

This paper could not have been written without the help of our colleagues at the Census Bureau. Six of our co-workers provided so much assistance that they are singled out for special thanks.

1. Randy Klear single-handedly built the data base used in the surname extraction operation. He wrote the programs to normalize names (JOHN SMITH JR is normalized to JOHN SMITH) as well as creating the algorithms for inverting names (JOHNSON CYNTHIA is inverted to CYNTHIA JOHNSON) when appropriate.

2. Sam Davis designed the programs to delineate infrequently occurring surnames into various Hispanic categories.

3. Marie Pees created the electronic diskettes that are an important supplement to the paper. For persons needing specific information on individual surnames, the statistical material located on the diskette is crucial.

4. Signe Wetrogan gave the authors a great deal of her time, enthusiasm and expertise in their early efforts at organizing and writing this paper. Many of her suggestions on points of emphasis have been included in this document.

5. Gregg Robinson painstakingly read and re-read several versions of this paper. His sense for where to expand and where to modify the authors' original phrasing were almost always right on the money.

6. Finally, we want to commend Rheta Pemberton on her word processing skills and her patience in producing "just one more final draft". The typographical errors which have crept into this paper are the sole responsibility of the authors.

TABLE OF CONTENTS

TEXT TABLES

APPENDIX TABLES

Building a Spanish Surname List for the 1990's—A New Approach to An Old Problem

by

David L. Word and R. Colby Perkins Jr.

This paper describes a direct and reproducible method for creating an inventory of surnames characteristic of the Hispanic origin population in the United States. The individual surnames included in this inventory are created by combining distinct surnames into groups and then analyzing group responses to the 1990 Hispanic origin question. Persons wishing to purchase an electronic file need to be specific as to whether they want the long list (Section 10.1.2) or the short list (Section 10.1.3).

Both electronic versions are available through the Population Division's Statistical Information Office (301-457-2422). If you would like or need additional insight into the contents of this paper, David Word (301-457-2103) dword@census.gov and Colby Perkins (301-457-2428) rperkins@census.gov will welcome your comments.

1.0 INTRODUCTION

In 1980 the Census Bureau published a list of 12,497 different "Spanish" surnames. The central premise for including a surname on that list was the "similarity" of that name's geographic distribution to the geographic distribution of the Hispanic origin population within the United States. The 12,497 surnames appearing on the 1980 Spanish surname list were culled from a data base of 85 million taxpayers filing individual federal tax returns for 1977.

Each of the 1.4 million distinct names appearing on the 1977 IRS file was subjected to a complex mathematical function incorporating Bayes' theorem to determine the "odds" that any particular surname was Spanish (Word, et al 1978). When the arithmetic value of the function exceeded a predetermined standard, that surname became a potential candidate for inclusion on the 1980 Spanish surname list. If the numerical value of the multinomial function failed to reach that criterion, the surname being tested was immediately discarded. This procedure works remarkably well for commonly occurring surnames, but a great amount of "hands on" effort was required to dispose of infrequently occurring surnames that surfaced as "Spanish" on the initial selection pass.

In this paper, Perkins and Word discard that **indirect** Bayesian approach in favor of a **direct** method to reach the same ends. Here, instead of attempting to "classify" surnames through geographic distribution, we actually link ethnicity and name. The ideal data source for classifying surnames by proportion Hispanic origin would be the 1990 Census in its entirety. Because of disclosure concerns, name has never been part of the computerized permanent record even though the Decennial Census routinely requests name for followup purposes.

Nevertheless, a very large sample data set is available that does link name (first and last) to individual 1990 Census records. This individual record file, hereafter called the SOR—(Spanish Origin)—file contains 7,154,390 person records[1] and was originally created for the purpose of estimating undercount in the 1990 Census. Since slightly over 1.5 million of those records lack name and/or Hispanic origin information, we limited ourselves to the 5,609,592 records that include both a valid surname and a response to the Hispanic origin question.

[1]Following the 1990 Census, the Census Bureau instituted a large scale post-enumerative survey (PES) to measure undercount in the 1990 census (Hogan, 1993; 1992). The formal PES sample was limited to 377,000 persons residing in 171,000 households in 5300 preselected blocks. The much larger SOR sample includes those PES blocks AND surrounding ring blocks. The SOR sample file used in this analysis is nearly 20 times as large as the formal PES sample.

Most people within a household have the same surname and the same ethnicity, implying that 5,609,592 person records do not produce 5,609,592 independent observations. To mitigate the effect of clustering, we limit our universe to the 1,868,781 Householder[2] records that include valid responses to both surname and Hispanic origin. This "householder" data set contains 268,783 distinct surnames—167,765 occurring exactly one time. In fairness, a large portion of surnames occurring one time appear to be errors in keying or errors in interpreting handwriting. GOUZALEZ, GOMEZS, and RODRIGUF are the surnames of three householders appearing in the SOR file who designated themselves as Hispanic.

For reasons sited in footnote 2, all future discussions of frequency/appearances/observations for individual surnames in the SOR file, will be taken as householders not persons.

2.0 BACKGROUND

If it were possible to develop a Spanish surname list that identifies all Hispanics, and does not include any non-Hispanics, we could represent that condition by Table 1.

TABLE 1—TABULAR ENTRIES IN AN IDEAL SITUATION

	Hispanic Origin	Non-Hispanic Origin	All Origins
Spanish Surname	**X**	ZERO	**X**
Non-Spanish Surname	ZERO	**Y**	**Y**
All Names	**X**	**Y**	**Z**

In Table 1, each of the **X** persons denoting themselves as Hispanic possesses a Spanish surname, and no person of Hispanic origin has a non- Spanish surname. Moreover, not one single person among the **Y** non-Hispanics possess a Spanish surname. This pattern does not hold in the real world. Hispanic persons may possess surnames that are not "Spanish", and non-Hispanics,—especially married women—can have Spanish Surnames. Table 2 illustrates this "real world" situation.

TABLE 2—TABULAR ENTRIES IN A NORMAL SITUATION

	Hispanic Origin	Non-Hispanic Origin	All Origins
Spanish Surname	**X**	**p**	**S**
Non-Spanish Surname	**q**	**Y**	**T**
All Names	**H**	**U**	**Z**

If the surname list under consideration behaves normally, the entries "**p**" and "**q**" are small relative to the values of **X** and **Y**. Displaying the data in this form clarifies the two relationships which are crucial in **evaluating** any Spanish surname list.

[2]The term "householder" used in the context of this paper is limited to male or never married female householders plus any other male or never married female in the household not related to the householder. We expressly exclude ever married women from the calculations because our interest in the relationship of surname to ethnicity lies in the potential of a given surname to identify persons of Hispanic origin. As would be suspected, the existing 1980 Spanish surname list is less effective in identifying the ethnicity of ever married females than any other demographic group (Perkins, 1993).

1. The entry "**p**" represents the number of persons possessing any "Spanish surname" appearing on an existing Spanish surname list who do not identify themselves as Hispanic. We define **Error of Commission** to be the ratio of **p** to **S**. That is, of the **S** persons who have Spanish surnames, "**p**" **are not** Hispanic. As a rule of thumb, fewer than 10 percent of the persons with generally accepted "Spanish" surnames fail to identify themselves as Hispanic. Ambiguous surnames, such as SANTOS and SILVA, should be **excluded** from any Spanish Surname list if a user's goal is to minimize Error of Commission.

2. The entry "**q**" represents persons who identify themselves as Hispanic, but whose surname is not found on a given Spanish surname list. **Error of Omission** is analogous to Error of Commission and is the ratio of **q** to **H**. However, Error of Omission is not strictly a rate. It is the proportion of the Hispanic origin population whose last name does not appear on a particular Spanish surname list. Although fewer than 1 percent of persons with non-Spanish surnames identify themselves as Hispanic, non-Hispanics outnumber Hispanics by 10 to 1 in the United States. For that reason, it is virtually impossible for Error of Omission to dip much below 10 percent, regardless of "fringe" surnames that are added to an existing surname list. If one desires to lower the Error of Omission at the expense of Error of Commission, indefinite surnames such as SANTOS and SILVA need to be **included** on a Spanish surname list.

3.0 PURPOSE OF CONSTRUCTING A SPANISH SURNAME LIST

The existing 1980 Spanish surname list was originally created to code persons of Spanish surname in the five Southwestern States at the time of the 1980 Census (Passel and Word, 1980). But that surname list has had a far wider range of uses and users since its release. Five practical applications involving the use of Spanish surnames follow:

3.1 **Mortality Studies.** Until very recently (late 1960's) there was no attempt to identify the Latin American community with a single unifying term. As a result, Mexicans, Germans, Iraqis and Peruvians were terms for persons of four distinct ethnic groups. By the late 1970's, the term Spanish origin came into vogue and Mexicans, Peruvians, Puerto Ricans, etc. were combined under a single generic designation—Spanish origin population. (The term Spanish origin has gradually been replaced or used interchangeably with the term Hispanic origin.) At the same time (1980) the Social Security Administration (SSA) revised their application form to request ethnic ("Hispanic") information for Social Security applicants. But neither Social Security nor its sister agency, Health Care Financing Administration (HCFA/Medicare), felt that it was necessary to obtain direct information on Hispanic origin for persons who had applied for and received Social Security numbers prior to 1980.

In order to obtain information on mortality of the elderly Hispanic population, HCFA is contemplating a large scale mortality study of the Hispanic origin population enrolled in Medicare. For a large proportion of that population, "Hispanic origin" will be defined and assigned on the basis of surnames contained on either the existing 1980 or the new 1990 Spanish surname list.

3.2 **Population Estimates.** The Census Bureau's initial effort at producing local area population estimates for the Hispanic population (Word, 1989) relied on the premise that the domestic migration rate of the Hispanic origin population could be approximated from the migration of the Spanish surnamed population as defined in 1980.

3.3 **Customer Base.** A utility company knows its customer base (by surname) at time t_0 and time t_1. The ratio of Spanish surnamed customers at the end point relative to the starting point provides an excellent basis for estimating change in the Hispanic origin population from the beginning to the end of the time period.

3.4 **Marketing.** In the first three applications, it was more important to limit errors of commission than errors of omission. But for marketing purposes it is generally useful to approach persons who are tangential to the group being studied. Suppose that a publisher wishes to launch a magazine written in Spanish about items of interest to persons of Hispanic origin. In order to get the largest subscriber base, it would be worthwhile to contact persons with borderline Spanish surnames on the chance that they are Hispanic.

3.5 **Census Use.** The Census Bureau is continually faced with the problem of "estimating" data when the respondent does not supply data on a census form. This estimation process is called "editing" or "imputation". Given that name will be captured on the year 2000 census record, a possible option to be considered is to use name to improve editing the Hispanic origin question when a direct response is not available.

4.0 ONE DOZEN COMMON SPANISH SURNAMES

The paper contains many abridged tables illustrating the authors' logic in generating Spanish surnames. For frequently occurring surnames, the qualification standards are self evident—we need only to know the ratio of successes (persons with a particular name identifying as Hispanic) to failures (persons with that same surname identifying as non-Hispanic). For rarely occurring names, the procedures for deciding whether a surname is or is not Spanish require more innovation.

As a starting point, we tabulated for each surname (SMITH as well as GARCIA) the proportion of persons who indicate that they are Hispanic. Using this construct, the criteria for establishing numerical limits on what constitutes a Spanish surname can be left to the individual data user. In practice, 95 percent of male householders with frequently occurring surnames (e.g., GOMEZ, GONZALEZ, GARCIA, RUIZ, etc.,) said they were Hispanic while less than 1 percent of males with common Anglo-Saxon surnames report themselves to be Hispanic. There are a few surnames (e.g., SILVA and SANTOS) for which the proportion of Hispanics is close to one-half, but these difficult to classify surnames are quite rare.

Approximately 20 percent of the Spanish surnamed population in the United States is concentrated in an even dozen names. The relative positioning of those 12 Spanish surnames in 1977 and 1990 appear in Table 3.

TABLE 3—RANKING SPANISH SURNAMES BY HOUSEHOLDER

(Source: 1977 (IRS); 1990 (Census SOR file))

	1977			**1990**	
Rank	Name	Percent	Rank	Name	Percent
1.	Garcia	2.97	1.	Garcia	2.90
2.	Martinez	2.69	2.	Martinez	2.73
3.	Rodriguez	2.51	3.	Rodriguez	2.55
4.	Lopez	1.99	4.	Lopez	2.23
5.	Hernandez	1.89	5.	Hernandez	2.16
6.	Gonzalez	1.65	6.	Gonzalez	1.87
7.	Perez	1.57	7.	Perez	1.73
8.	Sanchez	1.41	8.	Sanchez	1.50
9.	Gonzales	1.18	9.	Rivera	1.24
10.	Ramirez	1.13	10.	Ramirez	1.20
11.	Torres	1.03	11.	Torres	1.15
12.	Rivera	0.98	12.	Gonzales	1.06
	TOTAL	**21.00**		**TOTAL**	**22.31**

The term "householder" in Table 3 is used for convenience and does not follow a precise census definition. For the 1977 entries, a more exact descriptor would be "primary taxpayers on 1977 IRS returns". The 1990 SOR source includes male householders but excludes all female householders currently or previously married.

Table 3 focuses upon the stability of surname positional rankings. Even though the Hispanic origin population in the United States increased by 70 percent over the 13 year period (1977 to 1990), the relative positioning of the 12 most frequently occurring Spanish surnames are invariant in both data sources. Were it not for the inversion of RIVERA and GONZALES, the individual positional rankings among the first 12 Spanish surnames would be identical.

We are now prepared to address the following question: "Just how effective are Spanish surnames in identifying the Hispanic origin population?" Table 4 attempts to answer that question by presenting surname data from the SOR research file for both "householders" (H.H.) and all persons (POP). Note how the inclusion of ever married females in the POP column depresses the effectiveness of both Spanish and non-Spanish surnames as classifiers of ethnic populations.

TABLE 4—PERCENT OF HOUSEHOLDERS AND PERSONS SELF-IDENTIFIED AS HISPANIC

(Source 1990 Census-SOR)

	Spanish Surnames				Non-Spanish Surnames		
Rank	Surname	H. H.	Pop.	Rank	Surname	H. H.	Pop.
1.	Garcia	94.5	91.0	1.	Smith	0.7	1.2
2.	Martinez	95.9	93.2	2.	Johnson	0.6	1.1
3.	Rodriguez	96.9	94.2	3.	Williams	0.8	1.1
4.	Lopez	94.6	91.8	4.	Brown	0.9	1.3
5.	Hernandez	97.0	94.2	5.	Jones	0.5	0.9
6.	Gonzalez	98.0	95.5	6.	Davis	0.7	1.1
7.	Perez	95.8	92.6	7.	Miller	0.6	1.3
8.	Sanchez	96.4	93.4	8.	Wilson	1.0	1.5
9.	Rivera	96.1	92.3	9.	Anderson	0.7	1.4
10.	Ramirez	96.7	94.3	10.	Moore	0.5	1.1
11.	Torres	95.3	92.9	11.	Taylor	0.7	1.1
12.	Gonzales	92.1	89.8	12.	Thomas	0.8	1.2
30.	Silva	57.3	60.0	13.	Martin	2.5	3.2
47.	Santos	60.3	61.5	209.	Oliver	3.1	3.0

Table 4 demonstrates just how effectively the top 12 Spanish and Anglo surnames classify the total population as to Hispanic or non-Hispanic origin. About 93 percent of the population and 96 percent of the householders with the 12 most common Spanish surnames identified themselves as Hispanic in the 1990 Census. On the other hand, only 1.2 percent of the population and 0.7 percent of the householders with the 12 most frequently occurring Anglo names answered the Hispanic origin question affirmatively.

Note that MARTIN and OLIVER are substantially more Hispanic than the other 12 Anglo surnames. The reason for this is that the pronunciation of MARTIN and OLIVER can be altered from English to Spanish by accenting the last syllable rather than the next to the last syllable. We do not doubt that persons pronouncing their surnames as MAR TEEN or O LEE VAIR are generally Hispanic. Given that a name's pronunciation cannot be guessed from its spelling, the surnames MARTIN and OLIVER should not be classified as Spanish in the United States. Only 3 percent of persons with names spelled M-A-R-T-I-N or O-L-I-V-E-R responded positively to the Hispanic origin question on the 1990 Census.

5.1 STATISTICAL PROPERTIES FOR FREQUENTLY OCCURRING SURNAMES

The primary goal of this research is to supply statistical data on surnames where a sizeable proportion of persons with these surnames self-identify as Hispanic. Approximately 95 percent of householders possessing the 12 most frequently occurring Spanish surnames (Table 4) identify as Hispanic, and that pattern holds for the majority of Spanish surnames on the existing 1980 list. To avoid the awkward construction "**x** percent of persons with surname **s** are Hispanic", we will employ the arbitrary, but easily understandable usage of "Heavily Hispanic", "Generally Hispanic", "Moderately Hispanic", "Occasionally Hispanic" and "Rarely Hispanic" for surname classification purposes. Table 5 defines these terms.

TABLE 5—CRITERIA FOR SPANISH SURNAME CLASSIFICATION

Spanish Surname Classification	Proportion of Householders Who are Hispanic
1. Heavily Hispanic	Over 75 Percent
2. Generally Hispanic	50 Percent < x ≤ 75 Percent
3. Moderately Hispanic	25 Percent < x ≤ 50 Percent
4. Occasionally Hispanic	5 Percent < x ≤ 25 Percent
5. Rarely Hispanic	Less than or equal to 5 percent
6. Indeterminant	Name not on file

Within the SOR file, there were 8,614 distinct "householder" surnames which appear 25 or more times. Based on an extrapolation of Social Security data (Social Security Administration, 1984), persons with those 8,614 surnames account for 70 percent of the American population. 715 of these 8,614 surnames matched entries appearing on the 1980 Spanish surname list. Unpublished data from Passel and Word's earlier work suggest that these 715 "Spanish" surnames represent 83 percent of the Spanish surname population.

Tables 6A, 6B, and 7 provide "householder" data on proportion Hispanic for those 8,614 surnames.

TABLE 6A—CATEGORIZING FREQUENTLY OCCURRING SPANISH SURNAMES (1980 LIST) BY PROPORTION HISPANIC

Total Surnames = 715

Heavily Hispanic (over 75 percent)	93.1	
More than 95 percent		43.4
More than 90 percent		73.1
Generally Hispanic (50 to 75 percent)	6.0	
Moderately Hispanic (25 to 50 percent)	0.7	
Occasionally Hispanic (5 to 25 percent)	0.1	
Rarely Hispanic (less than 5 percent)	0.0	

From the information appearing in Table 6A and Table 7, it is evident that the Bayesian approach used to create the 1980 Spanish Surname List was quite successful. The vast majority (93.1 percent) of these 715 names fell into the Heavily Hispanic category, and nearly three-fourths of those surnames (73.1 percent) were Hispanic 90 percent of the time.

In our 1990 SOR File, we found only 5 instances where a "frequently" occurring 1980 "Spanish" surname fell into the Moderate classification (FELIX, PASCUAL, MIGUEL, JUAN, and TOLENTINO). And there is only a single instance (DECASTRO) where a surname appearing on the 1980

Spanish list would be classified as Occasionally Hispanic based on data in the SOR file. No surname appearing on the 1980 Spanish surname list occurring 25 or more times falls into the Rarely Hispanic category.

We now turn to the 7,899 surnames occurring at least 25 times in the SOR file that do not appear on the 1980 Spanish surname list.

TABLE 6B—CATEGORIZING FREQUENTLY OCCURRING NON-SPANISH SURNAMES (1980 LIST) BY PROPORTION HISPANIC

(Total Surnames = 7,899)

Rarely Hispanic (less than 5 percent)	96.3	
Less than 2 percent		84.3
Occasionally Hispanic (5 to 25 percent)	3.0	
Moderately Hispanic (25 to 50 percent)	0.5	
Generally Hispanic (50 to 75 percent)	0.3	
Heavily Hispanic (over 75 percent)	0.0	

Based on results from the SOR sample, not one of the 7,899 most frequently occurring "non-Spanish surnames" would now be assigned to the Heavily Hispanic category. There are, however, 20 surnames categorized as Generally Hispanic based on the SOR sample. They are, in order of Hispanic occurrence: (1) SILVA, (2) ROMAN, (3) MACHADO, (4) VENTURA, (5) PIMENTEL, (6) PALMA, (7) AQUINO, (8) BELLO, (9) ARAUJO, (10) CHAVES, (11) LEMOS, (12) VALERIO, (13) MANZO, (14) MATTA, (15) SALVADOR, (16) MACEDO, (17) VICTORIA, (18) BARBOZA, (19) REAL, and (20) LOMAS

Table 7 provides a numerical assessment of the Hispanic classification for the 8,614 surnames which appear 25 or more times in the SOR file. When Passel and Word created their 1980 Spanish surname list, they did not have the luxury of using the General or Moderate classification where most of the inconsistencies lie. As might be expected many of the surnames falling into those two categories were considered "close calls" by Word and Passel when they developed the 1980 Spanish surname list.

TABLE 7—HISPANIC CLASSIFICATION FOR SURNAMES OCCURRING 25 OR MORE TIMES ON THE SOR FILE

(On List: surname classified as Spanish in 1980)

	On List	Not on List
Heavily Hispanic (75% and over)	666	0
Generally Hispanic (50-75%)	43	20
Moderately Hispanic (25-50%)	5	42
Occasionally Hispanic (5-25%)	1	234
Rarely Hispanic (less than 5%)	0	7603
TOTAL	715	7899

Summary: The most frequent 8,614 surnames (715 + 7899) in the SOR file are exceedingly efficient for differentiating the Hispanic and Non-Hispanic populations. All of the 666 names which are over 75 percent Hispanic in the SOR file were identified as Spanish surnames in 1980. There are 7,603 surnames, none previously categorized as "Spanish", where fewer than 5 percent of respondents indicated that they are Hispanic. Note the paucity of surnames falling into the General and Moderate categories.

5.2 STATISTICAL PROPERTIES FOR INFREQUENTLY OCCURRING SURNAMES

Even though the 8,614 most frequently occurring surnames in the SOR file contain 70 percent of the total population and 83 percent of the Spanish surname population, they represent a very small proportion of all surnames or all surnames designated as "Spanish". The information appearing in Table 8 demonstrates that the correspondence between surnames classified as Spanish in 1980 and 1990 becomes somewhat weaker as the SOR sample thins. Nevertheless, the correspondence between surname and ethnicity for surnames occurring as few as 5 to 9 times in the SOR "householder" sample is still strong.

TABLE 8—CLASSIFYING SURNAMES ON THE 1980 SPANISH SURNAME LIST ACCORDING TO NUMBER OF OBSERVATIONS ON THE SOR FILE (householder only)

Group I,	25 or More Observations	n = 715
Group II,	10 to 24 Observation	n = 605
Group III,	5 to 9 Observations	n = 776

	Group I n = 715	Group II n = 605	Group III n = 776
Heavily Hispanic	93.1	84.3	78.4
Generally Hispanic	6.0	10.4	11.1
Moderately Hispanic	0.7	3.3	6.1
Occasionally Hispanic	0.1	1.6	2.6
Rarely Hispanic	0.0	0.3	1.9

Again referring to Passel and Word's unpublished data, the most frequent 1320 (those occurring 10 or more times) Spanish surnames on their 1980 list cover 90.6 percent of the Spanish surnamed population. When we extend the universe to the most frequent 2096 Spanish surnames (those occurring 5 or more times in the SOR sample), we reach 93.6 percent of the 1980 Spanish surnamed population.

Table 9, following, is similar to Table 7 but is confined to surnames appearing 5 to 24 times in the SOR file.

TABLE 9—1990 HISPANIC CLASSIFICATION OF SURNAMES OCCURRING 5 TO 24 TIMES IN THE SOR FILE BASED ON HISPANIC CLASSIFICATION IN 1980

	10 to 24 Observations		5 to 9 Observations	
1990 Hispanic Classification	On 1980 List	Not On 1980 List	On 1980 List	Not On 1980 List
Heavily Hispanic	510	9	600	58
Generally Hispanic	63	22	94	53
Moderately Hispanic	20	79	50	151
Occasionally Hispanic	10	893	17	1005
Rarely Hispanic	2	9033	15	15345
TOTAL	605	10036	776	16612

As before, the terms “On” and “Not On” refer to whether the surname does or does not appear on the 1980 Spanish surname list. There are 1381 (605+776) different surnames on the 1980 Spanish surname list which appear 5 to 24 times in the SOR sample file. Only 44 (10 + 2 + 17 + 15) of those surnames will be reclassified as either Occasionally or Rarely Hispanic based on the 1990 analysis.

Again referring to Table 9, we find that there are 26,648 (10,036 + 16,612) different surnames occurring 5 to 24 times on the SOR file that do not appear on the 1980 Spanish surname list. Only 67 (9+58) of those names are now classified as Heavily Hispanic. An additional 75 names (22+53) fall into the Generally Hispanic category.

Summary: Of the 605 Spanish names on the 1980 list occurring 10 to 24 times, 95 percent fall into the Heavy or General classifications, and only 2 names fall into the Rarely Hispanic group. For 776 names that occurred 5 to 9 times, almost 90 percent continue to be classified as Heavily or Generally Spanish. Fifteen surnames previously classified as Hispanic are now Rarely Hispanic.

6.0 LIMITATIONS

The data presented in Tables 3 through 9 are derived from a sample—albeit a very large one. The 5,609,592 matchable SOR records contain 597,533 individuals who reported themselves to be Hispanic in the 1990 Census. The proportion Hispanic (10.7 percent) within the SOR sample is higher than the Hispanic proportion (9.0 percent) enumerated in the 1990 Census. This finding is not unexpected as there was a conscious effort to oversample Hispanics in the PES. If we were using unweighted responses to estimate the total proportion of population with Spanish surnames, we would certainly overstate that ratio. But this analysis does not attempt to estimate population totals; rather, our goal is to estimate (on a name by name basis) the proportion of persons who are Hispanic. With this goal in mind there is no inherent reason against using unweighted observations.

Another limitation is response variance. We must accept the individuals census designation as to his or her origin. For most census question such as sex and age, a respondent will provide answers that are consistent over time. Based on the 1990 Decennial Census Content Reinterview Survey (McKenney et al, 1993), about 7 percent of persons saying that they were Hispanic origin in the 1990 Census decided that they were non-Spanish at the later date. And 11 percent of persons saying that they were Hispanic origin in the reinterview, indicated that they were non-Spanish on their 1990 Census forms. This recent finding on lack of consistency for Hispanic origin response reinforce previous findings from reinterview surveys.

Finally, we have errors in measurement due to random sampling. When 90 persons out of 100 with a particular name in the SOR sample answer the Spanish origin question affirmatively, we say that 90 percent of persons with that surname are Hispanic. But, there is an error associated with that estimate. Using the normal approximation to the binomial, the standard error of that estimate is approximately $\sqrt{p * (1 - p)/(n}$. Here p = 0.9 and n = 100. Table 10 below displays values of sampling errors associated with two choices of "p" and three values of "n".

TABLE 10—STANDARD ERRORS IN PROPORTION HISPANIC ARISING FROM A SAMPLE

N	X	P	S_p
300	270	90.0	1.7
100	90	90.0	3.0
30	27	90.0	5.5*
300	210	70.0	2.6
100	70	70.0	4.6
30	21	70.0	8.4

In Table 10, N = observations;
X = Hispanics;
P = Proportion Hispanic (x/n)
S_p = Standard error of p in percent

* When x or (n-x) drops below 5, the values of the normal distribution are no longer appropriate. For this row, the two sigma upper and lower limits are 97.5 and 73.7 percent.

7.0 RARELY OCCURRING SURNAMES: OR WHEN DO STATISTICS END AND WHEN DOES COMMON SENSE TAKE OVER?

To this point we have confined our comments to surnames appearing 5 or more times in our data set. Those 34,000 surnames encompass 85 percent of the householder population in the SOR file but less than 15 percent of the **number** of different surnames appearing in that file. Our goal is to classify every surname appearing on the SOR file; but for names appearing less than five times the proportion Hispanic should not and will not be the sole criterion for classification. In this section, we outline the thought process used in classifying infrequently occurring surnames. The exact details are found in Appendix Section 10.2 on page 21.

The 7.2 million record SOR file is a reasonably representative national sample (almost 3 percent) of persons enumerated in the 1990 Census. In general terms, it is quite possible to designate a surname as being Heavily Hispanic or Rarely Hispanic from samples of three or possibly even two surnames; but samples of this size are inappropriate for separating Generally Hispanic from Moderately Hispanic or Moderately Hispanic from Occasionally Hispanic. Table 11 presents data demonstrating why it is difficult to badly misclassify the ethnicity of a surname when 5 independent observations of that surname exist.

Assume that we are trying to categorize three separate surnames, and that five independent observations exist for each name. We also happen to know that among **all** Americans, surname "H" (Heavily) is 90 percent Hispanic; surname "M" (Midway) is 50 percent Hispanic and surname "R" (Rarely) is 2 percent Hispanic. Table 11 provides binomial probabilities (in percent) of getting 0, 1, 2, 3, 4, and 5 persons identifying as Hispanic for each of these three surnames.

TABLE 11—PROBABILITY OF FINDING "X" HISPANICS FROM 5 INDEPENDENT OBSERVATIONS

(Numbers in percent)

X	Name "H" (90%)	Name "M" (50%)	Name "R" (2%)
0	0.0	3.1	90.4
1	0.1	15.6	9.2
2	0.8	31.3	0.4
3	7.3	31.3	0.0
4	32.8	15.6	0.0
5	59.1	3.1	0.0

Armed with this knowledge, it is evident that for Heavily Hispanic (”H”) or Rarely Hispanic (”R”) surnames there is little chance of misclassifying a surname that occurs 5 times. If our five observation sample were to yield three Hispanics, we might be tempted to classify the surname as "H" when it should have been "M" or vice versa, but there is little chance that a type "R" name could provide 3 Hispanics in a sample of 5 independent observations.

7.1.1 Classification of 1980 Spanish Surnames Occurring 4 or Fewer Times on the SOR Sample. Table 12 presents data on the number of "householders" with Spanish surnames (1980 definition) whose surname surfaced four or fewer times on the SOR file.

TABLE 12—SURNAMES INCLUDED ON THE 1980 SPANISH SURNAME LIST WHICH APPEAR 4 OR FEWER TIMES ON THE SOR FILE

		Number of Hispanics				
Distinct Surnames	Appearances	4	3	2	1	0
424	4	**273**	91	30	14	16
594	3		**401**	100	53	40
1143	2			**790**	229	124
2358	1				**1784**	574
5882	0					

To aid in interpreting Table 12, the 1143 different surnames appearing exactly 2 times on the SOR sample represent 2286 (2 x 1143) householders. In 790 instances both householders having those particular surnames identified as Hispanic; in 229 cases one householder with the surname was Hispanic and one was not; in 124 cases neither householder with that surname said they were Hispanic. Overall, 74.8 percent of Spanish surnamed (1980 list) householders with names appearing exactly two times on the SOR file self-identified as Hispanic in the 1990 file.

It is especially enlightening to note that nearly one-half (5882) of the 12,497 surnames on the 1980 Spanish surname list did not even occur in the SOR file. For those 5882 names we **can not** make any judgement as to whether those names are associated with persons who are Hispanic origin. There are two reasons why the SOR file did not capture those 5,882 surnames: (1) Many of these 1980 names may have themselves been the result of miskeying (e.g., RODRIGUF); (2) The data base used in assembling the 1980 list consisted of 80 million observations; this sample uses only 1.8 million records. In any case, the length (number of names) of a surname list has little correlation on its effectiveness.

Table 13 presents data on the "householders" whose surname occurs 4 or fewer times on the SOR file and that surname **did not appear** on the 1980 Spanish surname list.

TABLE 13—SURNAMES THAT ARE NOT INCLUDED ON THE 1980 SPANISH SURNAME LIST AND APPEAR 4 OR FEWER TIMES ON THE SOR FILE

		Hispanic Responses				
Distinct Surnames	Appearances	4	3	2	1	0
9,056	4	48	34	57	362	8,555
16,115	3		180	142	543	15,250
37,073	2			740	1,146	35,187
165,407	1				9,849	155,558

Since none of the entries appearing in Table 13 was previously (1980 surname list) classified as Hispanic, we would never consider reclassifying surnames included in the far right column of Table 13 into any positive Hispanic category. The names appearing in the remaining cells in Table 13 will be categorized by more subjective measures described in the Appendix. One possible yardstick for classifying surnames might have been to extend the binomial expansion appearing in Table 11 to lesser numbers of sample observations. For example, the probability that 4 independent readings on a truly Spanish surname (90 percent successful in identifying Hispanics) would yield 1 or 0 Hispanics is 0.3 and 0.0 percent respectively. But we decided against employing the binomial because we have additional data at our disposal for classifying ethnicity of surnames.

There is a natural predilection to retain any surname appearing on the existing 1980 Spanish surname list unless the evidence for removal is strong. And we don't want to add additional surnames to the 1990 list unless there is overriding evidence for doing so. For surnames occurring often, we feel that the probability of misclassification is minimal, but the chance of misclassifying ethnicity based only on probabilities rises sharply as the sample shrinks. To aid us in our classification of surnames we turn to:

7.1.2 Orthographic Structure of Surname and Hispanic Status of Surname in 1980. For names occurring 4, 3, or even 2 times the entries on the binomial expansion can be of some guidance. But for surnames with single observations, the binomial expansion is useless. For that reason, we have assembled two additional items of information to guide us on the classification of surnames. They are (1) orthographic structure of surnames and (2) whether that surname appeared among the 12,497 surnames on the 1980 Spanish surname list.

7.1.3 Orthographic Structure of Surnames. Linguists, particularly the late Robert W. Buechley (Buechley, 1961, 1967, 1971, 1976), have observed that certain letter combinations are common amongst Spanish surnames. The two letter ending EZ as in MARTINEZ, RODRIGUEZ and LOPEZ is almost always indicative of a Spanish surname. But of even greater importance for Spanish surname classification is the fact that certain letter formations never or almost never occur among Spanish surnames.

We initially parsed all surnames appearing 5 or more times in the SOR file by the Hispanic classifications described previously. We discovered (not surprisingly) that no surname falling into Heavily, Generally, or Moderately category contained either a K or a W. Based on that finding, it would be logical to assume that any surname containing the letter K or W should not be classified Hispanic regardless of its performance in the SOR sample.

In addition to checking for the appearance of a K and/or W anywhere in the surname we also analyzed opening three letter and closing three letter combinations. The letters SMI as in SMITH and

JOH as in JOHNSON never initiated surnames falling into the first 3 Hispanic categories and ITH is not a Hispanic ending among frequently occurring SOR names. Buechley had previously determined that there are 1465 valid 3 letter starts and 1114 valid 3 letter endings among Spanish surnames. (More information on starts and endings appear in the technical Appendix.)

A third orthographic finding is that double letters excepting R and L just don't occur. The notable exceptions are S **AA** VEDRA, JA **SS** O, DELO **SS** ANTOS, and CO **TT** O. Thus a surname containing a double letter excepting RR and LL should not be classified as Spanish regardless of the proportion of householders with that surname who are Hispanic in the SOR file.

7.1.4 Hispanic Status of Surname in 1980. A second and final auxiliary item of information used in determining Hispanic classification for low occurrence surnames in the SOR was the 1980 status. We felt that the previous research was sound and the knowledge of whether a surname was or was not Spanish on the previous list was a piece of information to be used in categorizing surnames.

Summary—For frequently occurring surnames (e.g., 5 or more times in the SOR file), we believe that proportion Hispanic should be the sole means for classifying a surname. For rarely occurring surnames, there are three indicators used in classifying. They are, listed in importance: (1) proportion Hispanic, (2) orthographic structure, and (3) appearance on 1980 surnames list. See Section 10.2 in the Appendix for additional details on how these three criteria fit into a point value system.

8.0 CONCLUSION

The authors hope that the evidence presented here convinces the reader that a well constructed Spanish surname list is a useful alternative for identifying persons of Hispanic origin when Hispanic origin is not known. In some instances (estimating rate of change in the Hispanic origin population) defining Spanish origin solely through the use of surname may be preferable to self-designated Hispanic origin because surname provides a "consistent" response.

With very few exceptions every frequently occurring surname is either Heavily Hispanic or Rarely Hispanic and there is no middle ground. This finding is the determining factor why Spanish surname is such an excellent proxy for identifying Hispanics within the United States. Based on the analysis of the SOR file, fewer than 1000 surnames are sufficient for capturing 80 percent of the Hispanic population in the United States. Moreover, householders with those surnames are Hispanic 95 percent of the time.

The Census Bureau has released Spanish surnames following the Censuses of 1950, 1960, 1970, and 1980. This 1990 edition is only another station on an ongoing research journey, but this 1990 product does differ significantly from its predecessors. Each of the 25,277 individual surnames appearing on the electronic file that supplements this report contain auxiliary information allowing prospective users the flexibility to construct their own Spanish surname list if necessary. For example, we provide data on the surnames **SMITH, JONES,** and **ROBINSON** as well as **GARCIA, GOMEZ,** and **SILVA**. Granted, it is unlikely that any one would use this auxiliary information to conclude that **SMITH** is a Spanish surname. In theory, we are not providing a Spanish surname "list". Rather, we provide auxiliary data for each surname that can be sorted into a continuum allowing the prospective user to determine his or her own criteria as to what is or is not a Spanish surname.

If the SOR sample universe was doubled or even tripled (we had 1.9 million households in the SOR sample), we might have a better measure for classifying surnames that now appear 3 to 5 times. But a larger sample would also double or triple the number of persons named **SMITH** and **GARCIA** where the current sample size is already sufficient for classifying Hispanic status. Moreover, surnames that do not occur in this sample might appear 1 or 2 times in the larger sample and the problems with infrequently occurring surnames would still remain; only the infrequent surnames would be different.

9.0 REFERENCES

1. Word, David L., Jeffrey S. Passel, Beverly D. Causey, and Edward F. Fernandez , "Determining a List of Spanish Surnames by Analysis of Geographical Distributions." Unpublished paper delivered at annual meeting of Southern Regional Demographic Group, San Antonio Texas, October 1978

2a. Hogan, Howard, "The 1990 Post-Enumeration Survey: Operations and Results," The Journal of the American Statistical Association, 88:423, pp. 1047-1060, 1993.

2b. Hogan, Howard, "The 1990 Post-Enumeration Survey: An Overview", The American Statistician, 46:4, pp. 291-269, 1992.

3. Perkins, R. Colby, "Evaluating the Passel-Word Spanish Surname List: 1990 Decennial Census Post Enumeration Survey Results.", Population Estimates and Projections Technical Working Paper Series, August 1993

4. Passel, Jeffrey S. and David L. Word, "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, 1980.

4. Word, David L, "Population Estimates by Race and Hispanic Origin for States, Metropolitan Areas, and Selected Counties: 1980 to 1985.", Current Population Reports, Series P-25, No 1040 RD-1, Bureau of the Census, May 1989.

5. McKenney, Nampeo, Claudette Bennett, Roderick Harrison, and Jorge del Pinal, "Evaluating Racial and Ethnic Reporting in the 1990 Census", American Statistical Association, Proceedings of the Section on Survey Research Methods, 1993.

6. Social Security Administration, "Report of Distribution of Surnames in the Social Security Number File September 1, 1984", 1984.

7a. Buechley, Robert W., 1961. "A Reproducible Method of Counting Persons of Spanish Surname", Journal of the American Statistical Association 56 (March 1961)

7b. Buechley, Robert W., 1967. "Characteristic Name Sets of Spanish Populations", Names 15 (1, March 1967): 53-69.

7c. Buechley, Robert W., 1971. "Spanish Surnames Among the 2,000 Most Common United States Surnames", Names 19, (2, June 1971)

7d. Buechley, Robert W., 1976. "Generally Useful Ethnic Search System: GUESS", mimeographed paper, Cancer Research and Treatment Center, University of New Mexico, Albuquerque, New Mexico, November 1976.

10.0 APPENDIX

A significant portion of the Appendix is written for persons requiring electronic access to individual surname data. Consequently, persons with only a casual interest in Spanish surnames can be adequately served by reading section 10.3 and browsing the contents of Appendix Table A.

10.1 SERVING OUR CUSTOMERS

From talking to prospective customers of Spanish surname data, we conclude that we are serving two or perhaps even three classes of customers. The three classes include:

10.1.1 Persons who are satisfied with a minimal number of surnames (preferably on a piece of paper) that adequately cover a large proportion of the Hispanic origin/surnamed population within the United States. For these persons, we provide 639 Heavily Hispanic Spanish surnames arranged in alphabetic order in Appendix Table A. Persons with those surnames represent more than two-thirds of the Hispanic origin population and approximately 80 percent of the Spanish surnamed population (see Section 5.1 of the main text). The 639 surnames share two characteristics:

(1) For each surname appearing in Appendix Table A, at least 25 SOR “householders” provided positive responses to the Spanish origin question on their 1990 Census forms.

(2) Each of the 639 surnames listed in Appendix Table A qualify as heavily (75 percent) Hispanic. Overall, 94 percent of the householders in the United States with those surnames answered the 1990 Hispanic origin question affirmatively.

Note that these criteria do not precisely produce the tabulations appearing in Table 6A. There, we tabulated responses from 715 surnames that **both** occurred 25 or more times in the SOR file **and** appeared on the 1980 Spanish surname list. None of those 715 surnames were subjected to a minimum standard for percent Hispanic. In fact, one of those 715 surnames (DECASTRO) is now classified as occasional Hispanic.

For a surname to appear in Appendix Table A, we require 25 positive responses in the SOR file and a minimum Hispanic “hit rate” of 75 percent. Thus a 1980 Spanish surname that appeared 27 times in the SOR file with 24 positive Hispanic entries would be an entry in Table 6A but not in Appendix Table A.

For many purposes, this abridged 639 surname list is sufficient for making a reasonably accurate assessment on the number or proportion Hispanic within a group. Consider an organization of 100 persons. Twenty of the organization’s members have surnames that match the abbreviated 639 entry surname list. Armed with this information one can reasonably conclude that between 20 and 30 members are Hispanic. The number 30 is derived by dividing matched members (20) by 2/3—the proportion of the Hispanic population with these 639 surnames. For many/most uses an approximation with this level of accuracy suffices as a “ball park” estimator.

10.1.2 Persons who need surname data in electronic form and want the flexibility of customizing their own Spanish surname lists. The authors have arbitrarily categorized a surname to be Heavily Hispanic if more than 75 percent of householders with that name are Hispanic. Some users of Spanish surname data might wish to construct a surname base of Heavily Hispanic names where the criteria for Heavily is 90 percent, or 60 percent or some intermediate value. These customers will receive a flat file of 25,276 surnames arranged in nine data fields.

For purposes of illustration, we provide the contents for four individual names.

Field 1	Field 2	Field 3	Field 4	Field 5	Field 6	Field 7	Field 8	Field 9
0225	SILVA	0	2	710	499	407	344	0.441
0105	FEBUS	0	-2	8	5	7	5	1.875
0325	FELIX	1	2	187	132	88	78	-0.160
5500	BROOKS	0	-6	1714	587	5	4	-2.987

SILVA's category—0225—indicates that the surname is Generally Hispanic with more than 25 positive occurrences. The name did not appear on the 1980 list, but it does pass the Buechley test. The surname is much more likely (344/499) to be Hispanic in Hispanic states than non-Hispanic states (63/211).

FEBUS's, 0105 classification signifies that the surname is Heavily Hispanic with between 5 and 9 positive occurrences. The surname was not on the 1980 Spanish surname list. The final three letters in the surname (BUS) do not match the Buechley "Ends". Of the 8 householders with the name FEBUS, 7 are Hispanic. All 5 householders living in "Spanish States" are Hispanic.

FELIX is similar to SILVA except that the surname FELIX did appear on the 1980 Spanish surname list. It's category 0325 indicates that the surname is classified as Moderately Hispanic and there are more than 25 positive replies to the Hispanic question in the SOR sample.

BROOKS appears on the electronic file because it had at least one (actually 5) positive responses on the SOR file. The category 5500 indicates that the surname is Rarely Hispanic and that there are at least 500 negative responses for that surname. BROOKS (as expected) was not on the 1980 Spanish surname list. The score of -6 for Buechley occurs because of the existence of the letter K, the ending (OKS), and the double OO in the middle of the name.

Field 1 A numeric descriptor (located in positions 1-4) that provides both a Hispanic classification and a frequency grouping. Each of the 25,276 surnames appearing in these files falls into one and only one of 28 mutually exclusive categories. Appendix Table B (Spanish Surname Categories) define these 28 groupings.

Field 2 The surname itself—limited to 13 characters and appearing in positions 6 through 18.

Field 3 A "1" or a "0" appearing in column 20. A "1" signifies that this particular surname appears on the 1980 Spanish surname list; a "0" indicates that it did not.

Field 4 A positive "2" in column 24 or a negative even number appearing in columns 22 through 24. A "2" in column 24 signifies that the particular surname passes all the Buechley criteria. (See section 7.1.3 in main text for reference to Robert A. Buechley) A negative 2, 4, 6, 8, or 10 indicates whether the surname violates 1, 2, 3, 4, or even 5 Buechley rules.

Buechley Rule 1 — the letter K anywhere in name
Buechley Rule 2 — the letter W anywhere in name
Buechley Rule 3 — starts (initial 3 letters)
Buechley Rule 4 — ends (final 3 letters)
Buechley Rule 5 — double letters (excepting rr and $$)

Field 5 Total number of householders in the SOR File possessing the surname appearing in Field 2. Columns 25 through 30.

Field 6 Number of householders in the SOR file residing in one of the 11 states with large numbers of Hispanics. Columns 31 through 35.

We define the following 11 states to contain a large number of Hispanics: 1. Arizona, 2. California, 3. Colorado, 4. Connecticut, 5. Florida, 6. Illinois, 7. New Jersey, 8. New Mexico, 9. New York, 10. Pennsylvania, and 11. Texas.

Field 7 Total householders (national) with this surname who provide a positive response to the Spanish origin question. Columns 36 through 40. The ratio of the entry in Field 7 to the entry in Field 5 generates national Hispanic proportions for that particular surname.

Field 8 Hispanic householders in 11 States with large numbers of Hispanics. Columns 41 through 45. The ratio of the entry in Field 8 to the entry in Field 6 yields the Hispanic proportion for those 11 States.

Field 9 “Point Value of Surname” An integer (possibly preceded by a negative sign), decimal point, followed by three digits appears in columns 47 through 52. Although each and every one of the 25,276 surnames appearing in the electronic file is assigned a point value, that point value is only germane for classifying surnames when the number of positive and negative responses is fewer than 5.

10.1.3 Customers who want surname data in electronic form, but are willing to accept census “Hispanic” classifications. For those customers, we provide a file of surnames arranged in strict alphabetic order with the same 9 data fields described above. The major difference is that the number of surnames is limited to the 12,215 names which are classified as Heavily Hispanic. In addition to the surname data described above, we also furnish two additional tables which are:

(2) Electronic Table 3—STARTS is a file of 1465 three letter combination which start Spanish surname.

(3) Electronic Table 4—ENDS is a file of Buechley’s 1114 three letter combinations which end Spanish surname.

The entries appearing in STARTS and ENDS are primarily a product of Buechley’s research; but Passel and Word uncovered some inconsistencies which were relayed to Buechley in 1978. This version of STARTS and ENDS does not incorporate those additions to Buechley’s original work.

10.2 POINT VALUES FOR INFREQUENTLY OCCURRING SURNAMES

In Section 7.0 of this paper (Rarely Occurring Surnames: or Where Do Statistics End and When Does Common Sense Take Over?) we allude to the fact that proportion Hispanic would not and could not be the sole determinant for whether a prospective surname is Spanish and to which of the five categories (Heavily, Generally, Moderately, Occasionally, and Rarely) the surname is assigned.

From rereading the description of Field 9 in Section 10.1.2, it is immediately clear that any surname appearing 9 or more times is classified solely on the basis of proportion Spanish and any surname with fewer than 5 householder occurrences will be classified on the basis of point value. Some names appearing 5 to 9 times in the SOR file are assigned a Hispanic category based on proportion Hispanic while other surnames with 5 to 9 SOR appearances are classified only on point value.

As described in Section 7.0 there are three characteristics that can be used to classify a surname. These characteristics are:

(1) proportion of times possessor of surname is Spanish, (2) whether or not the surname follows acceptable Spanish language constructions, and (3) whether or not the 1980 research assigned that surname to be Spanish. We assigned points for each of these three attributes, with the assignment following the order described below:

1. For “householders” with a given surname captured in the SOR sample, how often does the possessor of that surname provide a positive Hispanic response? Give each Hispanic response a value of +3 and each non-Hispanic response a value of negative 3.

2. Does the surname adhere to or violate "orthographic correctness?" If the surname follows all 5 orthographic rules assign the surname a value of +2; assign a value of -2 for each violation.

> For example, DAVIS (which could be pronounced Dah Vees) violates no orthographic precepts. The starting three letters D A V appear in DAVILLA, the ending three letters V I S occur in OROVIS. DAVIS contains no W's, no K's, nor does it contain a double letter. All five American surnames occurring more frequently than DAVIS (eg. SMITH, JOHNSON, WILLIAMS, BROWN, and JONES) violate at least one of the orthographic rules which typify "Spanish" surnames.

3. Did the surname appear on the Census Bureau's 1980 Spanish Surname List? Give the surname a value of +1 if yes, and a value of -1 if no.

The point value of the surname is defined to be total points divided by total occurrences. If a name occurs only once, it could have a value as high as +6.00, and a theoretical low of -14.00. For example, the surname WEEKS receives -10 points on the orthographic variable alone. For frequently occurring surnames, the number of points awarded for orthographics and appearance on the 1980 Spanish surname list has very little weight. We illustrate this point with a surname occurring 100 times and a success rate of 95 percent.

AN ILLUSTRATION OF POINT SCORE CALCULATION:

Based on 100 observations

	Answers		Points Awarded		
	Yes	No	Yes	No	Total
(1) Response to Spanish origin question	95	5	285	-15	270
(2) Orthographics	1		2		2
(3) Appearance on 1980 List	1		1		1
Total Points			288	-15	273
Point Score					2.73

A frequently occurring Heavily Hispanic surname will achieve a point value ranging between 1.5 and 3.0. Point values of 2.5 to 2.7 are typical. The Heavily Hispanic standard for **infrequently** occurring surnames is set at equal to or greater than 2.00. It is possible for a surname appearing exactly one time on the SOR file with a single positive Spanish response to fall in the Heavily Hispanic category even though the surname did not appear on the 1980 Spanish surname list. But that surname **must** satisfy all five orthographic principles to receive the Heavily Hispanic designation.

The point values for Generally Hispanic were set at +1.00 to +1.99. The bounds for Moderately Hispanic were pegged from -0.50 to +0.99. As might be expected, the point values used in classifying infrequently occurring surnames parallel the values for frequently occurring surnames. We decided that it was virtually impossible to make an Occasionally Hispanic determination for infrequently occurring surnames. For that reason Spanish categories 0401 and 0402 (Appendix Table B) do not exist.

10.3 COMPARING HEAVILY HISPANIC WITH RARELY HISPANIC SURNAMES

Here we compare attributes of surnames for category 125—surnames with at least 25 Hispanic responses that are more than 75 percent Hispanic with category 5500 (surnames with more than 500 non-Hispanic responses that are less than 5 percent Hispanic). Data for the remaining 26 categories can be found in Appendix Table C.

Category	**125**	**5500**
Number of Surnames	639	353
Number of Observations	115,526	522,614
Percent Hispanic	94.2	0.7
Percent residing in Spanish States	86.3	37.2
Percent Passing Buechley	99.8	21.8
Percent on 1980 List	100.0	0.0

The analytic data associated with these most diverse categories of surnames aptly illustrate the points that we have made throughout the text.

1. Nearly 95 percent (94.2) of the male householder population with commonly “acknowledged” Spanish surnames identified themselves as Hispanic in the 1990 Census. Less than 1 percent of male householders with the most frequently occurring “non-Spanish” surname identified as Hispanic in the 1990 Census.

2. 86.3 percent of the persons possessing commonly “acknowledged” Spanish surnames reside in 11 states. The 1990 Census found 87.7 percent of the Hispanic origin population living in those same 11 states. By contrast, only 37 percent of persons with Anglo surnames reside in those same 11 states.

3. For the 639 surnames appearing in Appendix Table A, there are 638 surnames (99.8 percent) adhering to the Buechley rules. The one exception (COTTO) contains a double T. Although Buechley’s rules reject all doubletons except RR and LL, Spanish surnames containing a double T have been found in the SOR file.

4. Finally, all of the 639 most frequently occurring Spanish surnames were previously (1980) classified as Spanish. Not one of the 353 frequently occurring “Anglo” names were ever candidates for inclusion on a Spanish surname list.

APPENDIX TABLE A: 639 MOST FREQUENTLY OCCURRING HEAVILY HISPANIC SURNAMES

(Number to right of surname indicates relative ranking among Spanish surnames)

Abeyta	476	Baca	157	Carrion	340	Dominguez	63	Guardado	587
Abrego	534	Badillo	515	Carvajal	478	Dominquez	448	Guerra	85
Abreu	416	Baez	193	Casanova	419	Duarte	201	Guerrero	54
Acevedo	112	Baeza	456	Casares	600	Duenas	499	Guevara	211
Acosta	60	Bahena	616	Casarez	458	Duran	76	Guillen	311
Acuna	370	Balderas	359	Casas	341	Echevarria	394	Gurule	539
Adame	326	Ballesteros	552	Casillas	271	Elizondo	379	Gutierrez	24
Adorno	549	Banda	339	Castaneda	123	Enriquez	173	Guzman	43
Agosto	597	Banuelos	378	Castellanos	261	Escalante	349	Haro	471
Aguayo	409	Barajas	220	Castillo	25	Escamilla	275	Henriquez	480
Aguilar	45	Barela	405	Castro	37	Escobar	139	Heredia	336
Aguilera	243	Barragan	526	Cavazos	228	Escobedo	244	Hernadez	528
Aguirre	104	Barraza	381	Cazares	406	Esparza	169	Hernandes	520
Alanis	598	Barrera	111	Ceballos	498	Espinal	500	Hernandez	5
Alaniz	267	Barreto	497	Cedillo	571	Espino	469	Herrera	33
Alarcon	364	Barrientos	432	Ceja	410	Espinosa	143	Hidalgo	282
Alba	404	Barrios	200	Centeno	459	Espinoza	68	Hinojosa	229
Alcala	424	Batista	418	Cepeda	467	Esquibel	460	Holguin	372
Alcantar	567	Becerra	226	Cerda	296	Esquivel	231	Huerta	188
Alcaraz	599	Beltran	158	Cervantes	99	Estevez	619	Hurtado	253
Alejandro	550	Benavides	208	Cervantez	479	Estrada	52	Ibarra	114
Aleman	347	Benavidez	310	Chacon	213	Fajardo	382	Iglesias	489
Alfaro	207	Benitez	172	Chapa	247	Farias	428	Irizarry	233
Alicea	303	Bermudez	227	Chavarria	306	Feliciano	205	Jaime	442
Almanza	387	Bernal	168	Chavez	22	Fernandez	29	Jaimes	588
Almaraz	551	Berrios	299	Cintron	348	Ferrer	360	Jaquez	553
Almonte	614	Betancourt	290	Cisneros	135	Fierro	395	Jaramillo	171
Alonso	238	Blanco	163	Collado	536	Figueroa	59	Jasso	472
Alonzo	264	Bonilla	153	Collazo	318	Flores	13	Jimenez	35
Altamirano	466	Borrego	398	Colon	53	Florez	429	Jiminez	490
Alva	568	Botello	516	Colunga	434	Fonseca	335	Juarez	78
Alvarado	56	Bravo	194	Concepcion	426	Franco	116	Jurado	603
Alvarez	27	Briones	457	Contreras	71	Frias	461	Laboy	540
Amador	281	Briseno	433	Cordero	180	Fuentes	97	Lara	94
Amaya	265	Brito	333	Cordova	142	Gaitan	573	Laureano	604
Anaya	195	Bueno	316	Cornejo	441	Galarza	449	Leal	176
Anguiano	477	Burgos	209	Corona	186	Galindo	179	Lebron	400
Angulo	438	Bustamante	274	Coronado	221	Gallardo	232	Ledesma	300
Aparicio	535	Bustos	399	Corral	353	Gallegos	73	Leiva	622
Apodaca	273	Caballero	268	Corrales	601	Galvan	125	Lemus	297
Aponte	236	Caban	439	Correa	159	Galvez	307	Leon	95
Aragon	230	Cabrera	105	Cortes	175	Gamboa	354	Lerma	322
Arana	581	Cadena	440	Cortez	64	Gamez	302	Leyva	258
Aranda	285	Caldera	582	Cotto	468	Gaona	501	Limon	383
Arce	288	Calderon	107	Covarrubias	518	Garay	538	Linares	368
Archuleta	289	Calvillo	617	Crespo	278	Garcia	1	Lira	401
Arellano	190	Camacho	98	Cruz	17	Garibay	527	Llamas	554
Arenas	525	Camarillo	425	Cuellar	246	Garica	620	Loera	412
Arevalo	321	Campos	84	Curiel	572	Garrido	430	Lomeli	555
Arguello	569	Canales	260	Davila	129	Garza	26	Longoria	192
Arias	166	Candelaria	366	Deanda	584	Gastelum	586	Lopez	4
Armas	615	Cano	167	Dejesus	131	Gaytan	462	Lovato	502
Armendariz	447	Cantu	102	Delacruz	151	Gil	262	Loya	420
Armenta	417	Caraballo	317	Delafuente	585	Giron	411	Lozada	541
Armijo	377	Carbajal	367	Delagarza	371	Godinez	388	Lozano	122
Arredondo	212	Cardenas	106	Delao	602	Godoy	621	Lucero	124
Arreola	365	Cardona	214	Delapaz	537	Gomez	15	Lucio	481
Arriaga	397	Carmona	252	Delarosa	164	Gonzales	12	Luevano	491
Arroyo	132	Carranza	269	Delatorre	237	Gonzalez	6	Lugo	137
Arteaga	332	Carrasco	210	Deleon	81	Gracia	389	Lujan	215
Atencio	496	Carrasquillo	570	Delgadillo	427	Granado	519	Luna	66
Avalos	250	Carreon	583	Delgado	46	Granados	350	Macias	115
Avila	86	Carrera	517	Delrio	393	Griego	435	Madera	542
Aviles	245	Carrero	618	Delvalle	334	Grijalva	470	Madrid	185
Ayala	65	Carrillo	77	Diaz	14	Guajardo	308	Madrigal	270

APPENDIX TABLE A: 639 MOST FREQUENTLY OCCURRING HEAVILY HISPANIC SURNAMES

(Number to right of surname indicates relative ranking among Spanish surnames)

Maestas	304	Nazario	545	Posada	593	Salcedo	532	Vaca	636
Magana	248	Negrete	324	Prado	294	Salcido	309	Valadez	330
Malave	521	Negron	216	Preciado	531	Saldana	219	Valdes	240
Maldonado	51	Nevarez	369	Prieto	313	Saldivar	445	Valdez	47
Manzanares	623	Nieto	251	Puente	358	Salgado	184	Valdivia	524
Mares	402	Nieves	120	Puga	609	Salinas	80	Valencia	127
Marin	177	Nino	626	Pulido	444	Samaniego	511	Valentin	257
Marquez	61	Noriega	344	Quesada	484	Sanabria	454	Valenzuela	110
Marrero	178	Nunez	58	Quezada	292	Sanches	431	Valladares	577
Marroquin	312	Ocampo	355	Quinones	146	Sanchez	8	Valle	235
Martinez	2	Ocasio	361	Quinonez	413	Sandoval	55	Vallejo	386
Mascarenas	589	Ochoa	91	Quintana	140	Santacruz	631	Valles	396
Mata	138	Ojeda	255	Quintanilla	277	Santana	117	Valverde	548
Mateo	503	Olivares	272	Quintero	162	Santiago	41	Vanegas	637
Matias	529	Olivarez	305	Quiroz	218	Santillan	562	Varela	223
Matos	202	Olivas	291	Rael	463	Sarabia	632	Vargas	36
Maya	556	Olivera	558	Ramirez	10	Sauceda	512	Vasquez	23
Mayorga	605	Olivo	475	Ramon	407	Saucedo	239	Vazquez	62
Medina	30	Olmos	507	Ramos	20	Sedillo	594	Vega	49
Medrano	191	Olvera	276	Rangel	133	Segovia	523	Vela	182
Mejia	93	Ontiveros	301	Rascon	610	Segura	241	Velasco	293
Melendez	109	Oquendo	530	Raya	561	Sepulveda	280	Velasquez	96
Melgar	624	Ordonez	421	Razo	492	Serna	249	Velazquez	130
Mena	323	Orellana	443	Regalado	403	Serrano	89	Velez	83
Menchaca	482	Ornelas	283	Rendon	287	Serrato	612	Veliz	578
Mendez	39	Orosco	452	Renteria	256	Sevilla	613	Venegas	375
Mendoza	32	Orozco	147	Resendez	485	Sierra	187	Vera	197
Menendez	337	Orta	436	Reyes	19	Sisneros	563	Verdugo	579
Meraz	543	Ortega	50	Reyna	149	Solano	315	Verduzco	638
Mercado	103	Ortiz	16	Reynoso	325	Solis	90	Vergara	495
Merino	557	Osorio	338	Rico	295	Soliz	385	Viera	415
Mesa	342	Otero	174	Rincon	522	Solorio	446	Vigil	136
Meza	156	Ozuna	559	Riojas	574	Solorzano	564	Villa	134
Miramontes	606	Pabon	590	Rios	48	Soria	437	Villagomez	465
Miranda	79	Pacheco	92	Rivas	88	Sosa	118	Villalobos	225
Mireles	298	Padilla	57	Rivera	9	Sotelo	328	Villalpando	596
Mojica	343	Padron	508	Rivero	373	Soto	34	Villanueva	145
Molina	67	Paez	607	Robledo	509	Suarez	101	Villareal	423
Mondragon	450	Pagan	148	Robles	82	Tafoya	455	Villarreal	87
Monroy	544	Palacios	181	Rocha	121	Tamayo	414	Villasenor	392
Montalvo	254	Palomino	627	Rodarte	493	Tamez	595	Villegas	165
Montanez	286	Palomo	591	Rodrigez	629	Tapia	141	Yanez	266
Montano	203	Pantoja	356	Rodriguez	3	Tejada	513	Ybarra	189
Montemayor	504	Paredes	357	Rodriquez	38	Tejeda	464	Zambrano	488
Montenegro	505	Parra	217	Rojas	74	Tellez	352	Zamora	108
Montero	351	Partida	453	Rojo	510	Tello	565	Zamudio	639
Montes	154	Patino	345	Roldan	391	Teran	633	Zapata	224
Montez	451	Paz	327	Rolon	611	Terrazas	533	Zaragoza	376
Montoya	70	Pedraza	592	Romero	28	Tijerina	362	Zarate	331
Mora	119	Pedroza	422	Romo	222	Tirado	329	Zavala	170
Morales	18	Pelayo	546	Roque	486	Toledo	363	Zayas	514
Moreno	31	Pena	42	Rosado	144	Toro	346	Zelaya	580
Mota	483	Perales	384	Rosales	113	Torres	11	Zepeda	234
Moya	279	Peralta	263	Rosario	126	Torrez	242	Zuniga	155
Munguia	506	Perea	390	Rosas	152	Tovar	204		
Muniz	160	Peres	560	Roybal	408	Trejo	206		
Munoz	40	Perez	7	Rubio	128	Trevino	72		
Murillo	183	Pichardo	608	Ruelas	630	Trujillo	69		
Muro	625	Pina	196	Ruiz	21	Ulibarri	566		
Najera	319	Pineda	161	Ruvalcaba	575	Ulloa	494		
Naranjo	473	Pizarro	628	Saavedra	314	Urbina	374		
Narvaez	474	Polanco	320	Saenz	199	Urena	634		
Nava	198	Ponce	150	Saiz	487	Urias	576		
Navarrete	380	Porras	547	Salas	100	Uribe	284		
Navarro	75	Portillo	259	Salazar	44	Urrutia	635		

APPENDIX TABLE B: SPANISH SURNAME CATEGORIES

In Section 10.1.2 we described the file layout of the nine data fields associated with each surname. Now we concentrate on data field 1. The first two characters in field 1 denote Hispanic classification (01 for Heavily, 02 for Generally, 03 for Moderately, 04 for Occasionally and 05 for Rarely). The 3rd and 4th characters represent a frequency indicator.

When the frequency indicator (positions 3 and 4) takes on numerical values 05 through 25 (05, 10, 15, 25), Hispanic classification (Heavily, Generally, etc.) is determined strictly on the basis of proportion Hispanic as described in Section 5 of the text. When the frequency indicators are 01 or 02, (those names with 4 or fewer positive or negative) responses), we need to be more innovative. See Point Values for Infrequently Occurring Surnames. (Section 10.2 of this Appendix.)

Heavily Hispanic Surnames

Category	Entries	Description
0125	639	Surnames that are Heavily Hispanic with at least 25 positive Hispanic responses.
0115	251	Surnames that are Heavily Hispanic with at least 15 but no more than 24 positive responses.
0110	263	Surnames that are Heavily Hispanic with at least 10 but no more than 14 positive responses.
0105	625	Surnames that are Heavily Hispanic with at least 5 but no more than 9 positive responses.
0102	2463	Surnames that are Heavily Hispanic with at least 2 but no more than 4 positive responses.
0101	7974	Surnames that are Heavily Hispanic with exactly 1 positive Hispanic response.

Generally Hispanic Surnames

Category	Entries	Description
0225	39	Surnames that are Generally Hispanic with at least 25 positive Hispanic responses.
0215	25	Surnames that are Generally Hispanic with at least 15 but no more than 24 positive responses.
0210	25	Surnames that are Generally Hispanic with at least 10 but no more than 14 positive responses.
0205	106	Surnames that are Generally Hispanic with at least 5 but no more than 9 positive responses.
0202	354	Surnames that are Generally Hispanic with at least 2 but no more than 4 positive responses.
0201	218	Surnames that are Generally Hispanic with exactly 1 positive Hispanic response.

Moderately Hispanic Surnames

Category	Entries	Description
0325	11	Surnames that are Moderately Hispanic with at least 25 positive Hispanic responses.
0315	10	Surnames that are Moderately Hispanic with at least 15 but no more than 24 positive responses.
0310	21	Surnames that are Moderately Hispanic with at least 10 but no more than 14 positive responses.
0305	68	Surnames that are Moderately Hispanic with at least 5 but no more than 9 positive responses.
0302	260	Surnames that are Moderately Hispanic with at least 2 but no more than 4 positive responses.
0301	3611	Surnames that are Moderately Hispanic with exactly 1 positive Hispanic response.

Appendix Table B (continued)

For reasons cited in “Point Values for Infrequently Occurring Surnames”, Hispanic surname categories 0401 and 0402 do not exist.

Occasionally Hispanic Surnames

Category	Entries	Description
0425	5	Surnames that are Occasionally Hispanic with at least 25 positive Hispanic responses.
0415	13	Surnames that are Occasionally Hispanic with at least 15 but no more than 24 positive responses.
0410	16	Surnames that are Occasionally Hispanic with at least 10 but no more than 14 positive responses.
0405	65	Surnames that are Occasionally Hispanic with at least 5 but no more than 9 positive Hispanic responses.

Rarely Hispanic Surnames

Category	Entries	Description
5500	353	Surnames that are Rarely Hispanic with at least 500 negative responses and 1 or more positive Hispanic responses.
5100	1141	Surnames that are Rarely Hispanic with at least 100 but no more than 499 negative responses and 1 or more positive responses.
5025	1411	Surnames that are Rarely Hispanic with at least 25 but no more than 99 negative responses and 1 or more positive responses.
5010	986	Surnames that are Rarely Hispanic with at least 10 but no more than 24 negative responses and at least 1 but no more than 4 positive responses.
5005	969	Surnames that are Rarely Hispanic with at least 5 but no more than 9 negative responses and at least 1 positive response.
5001	3354	Surnames that are Rarely Hispanic with at least 1 but no more than 4 negative responses and at least 1 positive Hispanic response.

Category 5001 may include some surnames with 0 positive responses (and 1 to 4 negative responses) provided that that surname exists on the 1980 Spanish surname list.

The careful reader may have already realized that the 28 categories listed here do not encompass every surname appearing on the SOR file. For example a surname with 2 positive Hispanic responses and 50 negative responses would be tabulated in category 5025. Another surname with 0 (zero) positive responses and 50 negative responses would not be tabulated in any of the 28 categories. In fact, no surname with zero positive Hispanic responses in the SOR file (excepting surnames classified as Spanish in 1980) appear in Appendix Table B.

Because of this convention, the summary tabulations shown in Appendix Table C tend to overstate the proportion Hispanic within the Rarely Hispanic Classification. This phenomena is most noticeable with infrequently occurring surnames.

APPENDIX TABLE C: SELECTED SUMMARY STATISTICS FOR SPANISH SURNAMES

Heavily Hispanic

Category	101	102	105	110	115	125
Number of Names	7974	2463	625	263	251	639
Occurrences	7974	6626	4300	3295	5080	115526
Percent Hispanic	100.0	96.1	94.8	94.6	93.5	94.2
Percent in Spanish State	82.9	86.2	85.9	86.6	86.2	86.3
Percent Buechley-Yes	99.4	97.1	98.4	99.2	100.0	99.8
Percent on 1980 List	22.3	69.2	93.0	97.3	100.0	100.0

Generally Hispanic

Category	201	202	205	210	215	225
Number of Names	218	354	106	25	25	39
Occurrences	436	1041	1046	449	726	4038
Percent Hispanic	50.0	77.9	64.8	64.6	63.8	64.0
Percent in Spanish State	76.1	78.6	78.4	77.3	75.5	73.8
Percent Buechley-Yes	100.0	50.6	92.5	100.0	100.0	97.4
Percent on 1980 List	100.0	14.1	71.7	68.0	68.0	66.7

Moderately Hispanic

Category	301	302	305	310	315	325
Number of Names	3611	260	68	21	10	11
Occurrences	4288	1345	1187	640	522	1190
Percent Hispanic	71.4	49.7	37.2	39.2	38.1	39.6
Percent in Spanish State	75.2	69.2	65.9	65.6	60.7	61.7
Percent Buechley-Yes	32.2	82.7	94.1	90.5	100.0	100.0
Percent on 1980 List	17.0	34.6	25.0	14.3	10.0	9.1

Occasionally Hispanic

Category			405	410	415	425
Number of Names			65	16	13	5
Occurrences			3265	1445	2253	1375
Percent Hispanic			12.6	12.1	11.5	17.7
Percent in Spanish State			53.7	51.9	56.3	39.1
Percent Buechley-Yes			72.3	87.5	100.0	80.0
Percent on 1980 List			1.5	0.0	0.0	0.0

Rarely Hispanic

Category	5001	5005	5010	5025	5100	5500
Number of Names	3354	969	986	1411	1141	353
Occurrences	7940	7642	16689	74881	249666	522614
Percent Hispanic	41.5	15.6	7.7	2.5	1.0	0.7
Percent in Spanish State	62.4	54.6	48.2	41.0	38.4	37.2
Percent Buechley-Yes	22.9	44.6	39.1	31.1	24.8	21.8
Percent on 1980 List	7.0	3.2	1.0	0.0	0.0	0.0

It is important to note the low proportion of surnames in categories 102 (69.2 percent) and 101 (22.3 percent) that were classified as Hispanic in 1980. The evidence (proportion Hispanic, a pass on Buechley, and residence in 11 states where most Hispanic reside) suggests that the majority of persons possessing these names are borne by persons of Hispanic origin. But an examination of those surnames on a case by case basis suggests that the precise spelling of many of the names is incorrect. In other words, the sizeable number of surnames recorded as **VILLANVEVA** are almost assuredly a misinterpretation of **VILLANUEVA.**

POPULATION DIVISION WORKING PAPER SERIES

NO. 1 - "The Census Bureau Approach for Allocating International Migration to States, Counties, and Places: 1981-1991." David L. Word. October 1992.

NO. 2 - "Geographic Coding of Administrative Records—Past Experience and Current Research." Douglas K. Sater. April 1993.

NO. 3 - "Postcensal Population Estimates: States, Counties, and Places." John F. Long. August 1993.

NO. 4 - "Evaluating the Passel-Word Spanish Surname List: 1990 Decennial Census Post Enumeration Survey Results." R. Colby Perkins. August 1993.

NO. 5 - "Evaluation of Postcensal County Estimates for the 1980s." Sam T. Davis. March 1994.

NO. 6 - "Metropolitan Growth and Expansion in the 1980s." Richard L. Forstall and James D. Fitzsimmons. April 1993.

NO. 7 - "Geographic Coding of Administrative Records — Current Research in ZIP/Sector-to-County Coding Process." Douglas K. Sater. June 1994.

NO. 8 - "Illustrative Ranges of the Distribution of Undocumented Immigrants by State." Edward W. Fernandez & J. Gregory Robinson. October 1994.

NO. 9 - "Estimates of Emigration of the Foreign-Born Population: 1980-1990." Bashir Ahmed and J. Gregory Robinson. December 1994.

NO. 10 - "Estimation of the Annual Emigration of U.S. Born Persons by Using Foreign Censuses and Selected Administrative Data: Circa 1980." Edward W. Fernandez. January 1995.

NO. 11 - "Using Analytic Techniques to Evaluate the 1990 Census Coverage of Young Hispanics." Edward W. Fernandez. May 1995.

NO. 12 - "Metropolitan and Nonmetropolitan Areas: New Approaches to Geographical Definition." Donald C. Dahmann and James D. Fitzsimmons. October 1995.

NO. 13 - "Building a Spanish Surname List for the 1990's—A New Approach to An Old Problem." David L. Word and R. Colby Perkins, Jr. February 1996.

For copies of these Working Papers, please contact author at: Population Division, Bureau of the Census, Washington, DC 20233.