

Relative Coverage in the 1980 Census of Puerto Rico  
Robert F. O'Brien, Bureau of the Census

## I. Introduction

The Puerto Rico Labor Force Match Study was intended to identify 1980 census coverage differences in Puerto Rico by age, sex, and certain limited geographies. To meet this objective this study estimated the match rate of the April 1980 Puerto Rico Labor Force Study (LFS) into the 1980 Decennial Census of Puerto Rico.

It should be noted that it was not the intention of this study to quantify the overall coverage rate of the 1980 census in Puerto Rico. This study included no sample from the census to estimate overcoverage in the Puerto Rico census due to duplications and erroneous listings, therefore, the emphasis of this study is on relative coverage differences not total undercoverage.

What we found in this study was the match rate of persons in the LFS into the 1980 census was 92.23 percent. There appears to be some evidence to say that coverage in and around the San Juan and Caguas SCSA and for young adults posed the greatest coverage problems in the census of Puerto Rico.

The 92.23 percent match rate does not indicate that undercoverage in Puerto Rico was 7.77 percent of the population. The 7.77 percent non-match rate more broadly defines the inability of this study to match this portion of the population into the census based on the information available. For example, on the mainland U.S., where the Current Population Survey was matched into the 1980 Census of the United States, the match rate for all persons was estimated at 94.7 percent. It is widely recognized however, that the undercoverage rate in the U.S. census is not in the neighborhood of 6 percent; rather it is estimated to be much lower. This is the result of various assumptions that have to be made about persons who did not match into the census, and the effect of having a sample to estimate census duplications and erroneous enumerations.

## II. Background

### A. Labor Force Survey

The Puerto Rico Labor Force Survey is a multi-stage cluster sample of households used to estimate employment statistics of the non-institutional population in Puerto Rico on a monthly basis. The current size of the household sample consists of approximately 6,300 complete interviews each month, with coverage in almost all "municipios" counties). This means that each household represents a little over 100 households based on 1970 census counts from which the original sample was drawn.

The basic design of the survey consists of 15 strata. These strata are formed geographically by first defining seven non-overlapping zones that cover the island and then subdividing each zone into an urban and rural stratum based on 1970 census geography, making 14 strata overall. The fifteenth stratum is an open stratum formed by accumu-

lating all public and private housing projects established after April 1, 1970.

The basic sampling unit in the LFS is a cluster or segment of households. In an urban stratum the expected cluster size is 6 households and in a rural stratum, 12 households. The selection of clusters was done in three stages using the 1970 census geography. First, within each stratum a systematic sample of enumeration districts was drawn using the number of households as a measure of size. In the second stage, blocks within enumeration districts were selected systematically again using households as a measure of size. In the last stage, segments were systematically selected within blocks after the block had been relisted by LFS enumeration.

To avoid visiting the same household continuously, a rotation plan is used. After making the selection of enumeration districts, these enumeration districts were systematically divided into eight rotation groups or subsamples. Households in a rotation group follow a 4-8-4 pattern of interviewing, that is, they are interviewed for four consecutive months, out of sample for eight months, and interviewed again for four months.

### B. 1980 Census of Puerto Rico

The 1980 Decennial Census of Puerto Rico was a conventional census, that is, all households were enumerated in person by a census enumerator. There were no mail questionnaires in Puerto Rico.

To carry out the census in Puerto Rico, the island was divided into eight district offices or geographic areas for census processing. (See Map ). In each district office there was one processing center or headquarters, through which all work was supervised and all the census questionnaires collected. The enumeration in Puerto Rico took place from April 1980 through August 1980.

### C. Matching the LFS Into the Census

The matching of the LFS households into the 1980 census was an exact match operation performed by census clerks and a small staff of professionals from Census Headquarters. The exact match operation consisted of first attempting to match a LFS household address to an identical address in the census for which there was a census questionnaire, and then attempting to match the household persons listed on the LFS roster to persons listed on the census questionnaire. A person was matched to the census based on name, sex, age and relationship to head of household. In general, a person was considered matched into the census from the LFS if she/he agreed on at least a name basis, plus one or two other personal characteristics.

Persons listed on the LFS household rosters had an opportunity to be matched to the census in three separate matching operations; the Level 1 and Level 2 Match, and the Final Match. The Level 1 and Level 2 matching operations took place in the census district offices right after the census. In the Level 1 match, LFS households were geocoded (assigned)

to the most probable census enumeration districts that contained them. In this phase straightforward matches were resolved. LFS households were linked with census households with the same address and then the persons within the households were attempted to be matched. In the Level 2 matching operation all households that had at least one non-matched person after Level 1 were reviewed and a thorough search of the geocoded enumeration districts was carried out to match the LFS households or non-matched persons to a household or person in the census.

At the end of Level 2 match operations there were about 1,000 of the original 6,000 LFS households that had at least one non-matched person in them. All of these households were interviewed in a follow-up interview in January and February of 1981 to determine if these persons actually resided there on April 1, 1980. All households in this group had a re-response from either a household member or a proxy resident that was accepted.

After the follow-up interview a final match operation took place. In this phase, all the operations were reviewed and the LFS households were re-geocoded to a census enumeration district if necessary. A thorough search was made of the enumeration district, and all contiguous enumeration districts, to match a household or a person in a household to the census. At the end of this phase of matching, all persons in the LFS survey had a final match code assigned.

Throughout the matching operations a person could receive only one of three match codes; match, non-matched or out-of-scope. A person was considered out-of-scope if it was determined from the LFS roster or the follow-up interview that this person was not a resident of the household on April 1, 1980. A person was either matched or non-matched to the census if we had no information contrary to the fact that they were residents of the household as of April 1, 1980.

#### D. Matching Bias

Any type of matching study is open to varying degrees of bias due to matching errors. In this study two of the most obvious components of bias were due to false matching of persons and false non-matching of entire households. The effects of these biases are unmeasurable for this study, but it is believed that they are small and somewhat offsetting.

In Puerto Rico a false person match could occur when a person on the LFS household roster was falsely matched to a person in the census. This is a common error in matching studies. In Puerto Rico this problem is evident in areas where households have non-explicit addresses and many persons in the area have the same surname. In these areas, it is very probable that a small number of individuals were falsely matched to a person with a similar name and personal characteristics. This effect would tend to bias the match rate estimates upwards, but it is believed this type of bias would be very small.

On the other hand, the bias arising from

the false non-matches of entire households would tend to bring the match rate down. These types of matching errors could occur for LFS households that were geocoded to the wrong census enumeration district for searching, thus giving them no chance of being matched. Another reason for these false non-matches could be that the census questionnaire for a specific unit was placed in the wrong enumeration district during census processing. These two types of false matching errors, creating a geocoding bias, would probably be the most common in this study. It is assumed, the decrease in the match rate due to false non-matches is greater than the increase in the match rate due to false matches, and that the combined effect of these errors is within sampling error of the estimates.

#### V. Estimation

The weighted estimates that appear in the tables reflect the application of base-weights (the inverse of the probability of selection) and non-interview adjustment. The base weights were applied within strata and non-interview weights were applied by rotation group.

It should be noted that the weighted estimates of level that appear in the tables are far below census counts for the non-institutional populations of Puerto Rico for 1980. The reason for this is that the base weights reflect the selection of households according to 1970 census counts of housing, with the exception of the new construction stratum that represents only public and private housing projects built and occupied since April 1, 1970. A lack of accurate estimates for April 1980 (independent of the 1980 census) that could be used to ratio adjust the sample by age and sex forced us to use the present estimates. However, since the goal of this study is only to estimate the match rates from the Labor Force Survey into the census in order to identify coverage differences, it is believed that the match rates would not be unduly affected for our purposes.

Sampling errors were estimated using the method of random groups (replicated samples) with the eight rotation groups of the Labor Force Survey used to represent the random groups (replicates) in the standard error computations.

#### VI. Analysis

If we look at the estimates for total population by age and sex (Table 1), we may note that the study shows an overall match rate for the total population of 92.23 percent, with an associated standard error of 0.72 percent. As it is not the object of this study to quantify this estimate, let us look at coverage differences that could have affected this match rate.

In Table 1, we note that by sex there appears to be no apparent coverage difference between males with a 91.99 percent match rate and females with a 92.45 percent match rate. However, if we look at this table by age alone, we see that coverage for persons under 30 is significantly lower, 90.85 percent, than for persons 30 and over, 94.05 percent. In particular, persons in the age group 20-29

have a significantly lower match rate, 88.02 percent over all the other age groups. This type of coverage differential, the younger population versus the older population, is also apparent in the 1980 U.S. Census. Though the differences between age groups appears significant, the magnitude of the difference could be biased upward by a possible bias in the Labor Force Survey to list persons more accurately who are older and in the labor force, since it is concerned with producing employment statistics for persons aged 16 and over.

If we wish to look at coverage differences unique to Puerto Rico, it is best to look at coverage differences generated by geography. At first step we will look at coverage differences in Puerto Rico at the district office level and then look at coverage differences by urban and rural breakdowns and SMSA's (Standard Metropolitan Statistical Areas) and the San Juan-Caguas SCSA (Standard Consolidated Statistical Area). District offices split up the island as evenly as possible by population and housing units, and follow "municipio" (county) boundaries. In observing estimates from district offices it is possible to view the consistency of census coverage in Puerto Rico. (See Map)

At the district office level, district offices 3301 and 3302 tend to have lower match rates than the other district offices, though the data at the district office level do not support any consistent differences. Analyzing data at the district office level is difficult, due to the small sample sizes that may bias the match rates and produce large standard errors. By combining district offices together, we may diffuse some of these problems and sharpen comparisons.

In Table 2 we may view the combined estimates of district offices 3301 and 3302 versus the remainder of the island (district offices 3303-3308). We see that the match rate for total population for district offices 3301 and 3302 combined (89.62 percent) is significantly different from the match rate of the remainder of the island (92.86) at a 95 percent level of significance. The difference in these estimates appears to be primarily due to the lower match rate for the urban area in combined district offices 3301 and 3302. In district offices 3301 and 3302 the combined match rate in urban 1980 areas is 89.39 percent, which is significantly different, at a 95 percent level of significance, from the 92.93 percent match rate for the remainder of the island. At this point we see that the heavily urbanized district offices 3301 and 3302 combined, which contain the central portion of the San Juan and Caguas SCSA, have a lower match rate than the remainder of the island. However, the SCSA of San Juan and Caguas contains more than the area contained in district offices 3301 and 3302, and reaches out into the other district offices. To get a better look at the differences between the San Juan and Caguas areas versus the remainder of the island, some finer breakdowns may be useful.

If we look at the Island of Puerto Rico by

urban and rural as defined in the 1980 census, Table 2, we see that there is no significant difference between urban and rural match rates (91.97 and 92.80 percent, respectively). However, if we look at the urban area alone, we see that there is some evidence of a difference at the 80 percent level of significance that the urban in 1980 and rural in 1970 areas have a lower match rate than the urban in 1980 and 1970 areas (86.86 percent versus 92.66 percent). In a finer breakdown, we may see that this same comparison has no evidence of a difference in the combined estimates of district offices 3301 and 3302, because of the small sample size in the urban 1980, rural 1970 areas; but for the remainder of the island there is some evidence of a difference, at the 80 percent level of significance, where where the urban 1980 and 1970 match rate is 93.61 percent and the urban 1980, rural 1970 match rate is 88.92 percent. Thus, within urban areas, we may conclude that there is some evidence that outside of the central portion of the San Juan and Caguas SCSA there is a coverage differential between urban growth areas and older urban areas.

Table 3 gives estimates for the San Juan and Caguas SCSA, and other SMSA's in Puerto Rico as defined for 1980. In this table we may get a better view of possible coverage differences within urban areas. It may be noted that the SCSA of San Juan and Caguas entirely contains district offices 3301 and 3302, with the exception of one municipio, Aguas Buenas. In Table 3, we note that in the San Juan Caguas SCSA there is some evidence of a coverage difference between the urban 1980 and 1970 portion with a 92.08 percent match rate and the urban 1980, rural 1970 portion with a 82.92 percent match rate, at a 90 percent level of significance. If we look at the Ponce SMSA, we also observe a significant difference in this comparison, but the sample in the urban 1980, rural 1970 area is too small and prone to possible bias to make this comparison meaningful. The other SMSA's of Mayaguez and Arecibo, and the non-SMSA portion of the island show fairly consistent match rates in all areas with little, if any, difference outside of sampling error. Thus, coverage difficulties in urban growth areas appear to be limited to the San Juan and Caguas SCSA.

In reviewing the analysis of Tables 1 through 3, so far we have found the following coverage differences:

- 1) There is evidence to support the fact that persons under 30 are covered at a lower rate than those over 30.
- 2) There is evidence to support the fact that coverage in the central portion of the San Juan and Caguas SCSA, (district offices 3301 and 3302) is lower than the remainder of the island. The remainder of the island seems to be covered at a fairly consistent rate.
- 3) Within urban areas alone there is some evidence to show that there is a coverage difference between urban growth areas (areas urban in 1980 and rural in 1970) and older urban areas (those urban in 1980 and 1970), and that this difference is centered in the

San Juan and Caguas SCSA.

Now that the data have defined to some extent where coverage differences may be in Puerto Rico, another question may be how did these coverage differences occur. In Table 4 there are estimates of the total non-match rate of persons by district office, also broken down by non-matched persons within whole household non-matches and partial household non-matches (households where at least one person was matched and at least one person was non-matched). At the island level it shows that there is some evidence of a difference at the 90 percent significance level, that more persons were non-matched within whole household non-matches than within partial household non-matches. Specifically, an estimated 107,000 persons were non-matched within whole household non-matches which accounted for 4.51 percent of the overall match rate of 7.77 percent, while non-matches within partial household non-matches, totaling 77,000 persons, contributed 3.26 percent to the overall non-match rate. Looking at the district offices separately, it is difficult to determine differences between estimates. However, if we look at the combined estimates from district offices 3301 and 3302, the central portion of the San Juan and Caguas SCSA, we find that a higher percent of the total population is non-matched within whole household non-matches (4.03 percent) than within partial household non-matches (3.11 percent). This difference in the estimates is significant at a 95 percent level of significance. If we look at this comparison in the remainder of the island, there is no evidence of difference between non-matches within whole household non-matches and partial household non-matches. Thus, we seem to find that in the central portion of the San Juan and Caguas SCSA, which already differs in coverage from the remainder of the island, that coverage is more affected by whole household misses more than within household misses.

#### VII. Conclusions

In summary, it appears that the census of Puerto Rico in 1980 had better coverage of persons over 30 than under 30. However, the magnitude of the difference here could be affected by a bias in the Labor Force Survey household roster and also a possible bias in the matching procedures that were discussed earlier.

It also appears that there is some evidence to support the fact that urban growth areas pose a coverage problem in the census, especially the urban growth areas in the San Juan and Caguas SCSA. In the central portion of the San Juan and Caguas SCSA the match rate is lower than the remainder of the island. The tendency here seems to be to miss whole housing units more than persons within housing units, while in the remainder of the island misses due to whole household and within household non-matches are consistent.

In a conventional census, as we had in Puerto Rico, there could be several reasons for this coverage difference between the central portion of the San Juan and Caguas SCSA and the remainder of the island. The

following conjectures are offered:

- 1) Enumerators had greater workloads than was expected in urban areas.
- 2) Census maps were of poor quality in urban areas or they were poorly updated. This would seem to some extent to account for the higher rate of missed housing units in urban areas.
- 3) In urban areas we expect to have more multi-unit structures than single unit structures. The census could be missing units within multi-unit structures; this could be another reason for the higher incidence of missed housing units.

Looking to future censuses in Puerto Rico, it may now be the time to start considering a Prelist/Pre canvass operation in urban areas in conjunction with a conventional census or start a mail return operation in urban areas with an Update List Leave operation. In this way enumerator workloads in the urban areas could be better defined geographically in terms of updated maps. Also, the Prelist/Pre canvass or Update List Leave operations could give a precensus clue to where trouble areas may lie before the census, rather than during the census.

Table 1: Puerto Rico - Total Population by Age and Sex

TOTAL POPULATION			
Age	Total	Matched	% Matched
Total	2,369,588	2,185,432	92.23
00-14	736,500	677,383	91.97
15-19	269,287	246,041	91.37
20-29	341,921	300,971	88.02
30-39	301,370	281,044	93.26
40-49	241,957	230,121	95.11
50-59	206,243	194,742	94.42
60-69	154,244	144,284	93.54
70+	118,067	110,848	93.89
00-29	1,347,708	1,224,395	90.85
30+	1,021,880	961,037	94.05
MALE POPULATION			
Total	1,154,152	1,061,715	91.99
00-14	377,828	347,806	92.05
15-19	135,795	124,600	91.76
20-29	160,120	139,475	87.11
30-39	139,785	128,314	91.79
40-49	110,885	105,025	94.72
50-59	96,714	91,563	94.67
60-69	77,671	72,890	93.84
70+	55,354	52,042	94.02
00-29	673,743	611,881	90.82
30+	480,409	449,834	93.64
FEMALE POPULATION			
Total	1,215,436	1,123,717	92.45
00-14	358,672	329,577	91.89
15-19	133,492	121,441	90.97
20-29	181,802	161,495	88.83
30-39	161,585	152,730	94.52
40-49	131,071	125,096	95.44
50-59	109,529	103,178	94.20
60-69	76,573	71,394	93.24
70+	62,713	58,805	93.77
00-29	673,965	612,513	90.88
30+	541,471	511,203	94.41

Table 2: District Offices 3301 and 3302 Versus the Remainder of the Island - Total Population by Urban & Rural

	Total Population	Urban in 1980			Rural in 1980		
		Total	Urban in 1980 Urban in 1970	Urban in 1980 Rural in 1970	Total	Rural in 1980 Urban in 1970	Rural in 1980 Rural in 1970
<b>Puerto Rico</b>							
Total	2,369,588	1,626,396	1,431,307	195,089	743,192	28,395	714,797
Persons Matched	2,185,432	1,495,781	1,326,320	169,461	689,651	26,430	663,221
% Matched	92.23	91.97	92.66	86.86	92.80	93.08	92.78
<b>DO 3301 and DO 3302 Combined</b>							
Total	462,162	441,171	417,671	23,500	20,991	0	20,991
Persons Matched	414,179	394,361	377,477	16,884	19,818	0	19,818
% Matched	89.62	89.39	90.38	71.85	94.41		94.41
<b>Remaining DO's</b>							
Total	1,907,426	1,185,226	1,013,636	171,590	722,201	28,395	693,805
Persons Matched	1,771,253	1,101,420	948,843	152,577	669,833	26,430	643,403
% Matched	92.86	92.93	93.61	88.92	92.75	93.08	92.74

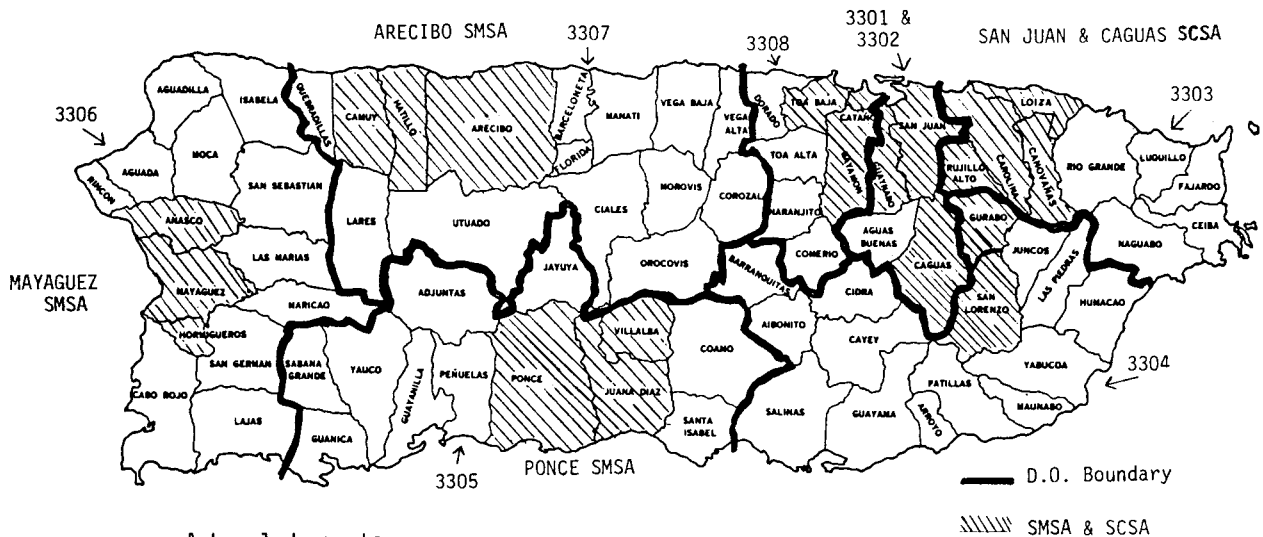
Table 3: 1980 SMSA's - Total Population by Urban and Rural

	Total Population	Urban in 1980			Rural in 1980		
		Total	Urban in 1980 Urban in 1970	Urban in 1980 Rural in 1970	Total	Rural in 1980 Urban in 1970	Rural in 1980 Rural in 1970
<b>SCSA:</b>							
<b>San Juan &amp; Caguas</b>							
Total	929,289	865,154	813,178	51,977	64,134	2,058	62,076
Matched	850,314	791,919	748,813	43,106	58,395	1,862	56,533
% Matched	91.50	91.53	92.08	82.92	91.05	90.48	91.07
<b>SMSA:</b>							
<b>San Juan</b>							
Total	805,415	769,323	737,905	31,418	36,092	2,058	34,034
Matched	738,009	706,818	678,018	28,800	31,191	1,862	29,329
% Matched	91.63	91.88	91.88	91.67	86.42	90.48	86.17
<b>Caguas</b>							
Total	123,874	95,832	75,273	20,559	28,042	0	28,042
Matched	112,305	85,101	70,795	14,306	27,204	0	27,204
% Matched	90.66	88.80	94.05	69.59	97.01	-	97.01
<b>Ponce</b>							
Total	189,742	150,831	137,142	13,689	38,912	1,641	37,271
Matched	173,526	136,111	128,656	7,455	37,415	1,422	35,993
% Matched	91.45	90.24	93.81	54.46	96.15	86.67	96.57
<b>Mayaguez</b>							
Total	103,373	86,839	75,113	11,726	16,534	0	16,534
Matched	96,477	80,774	69,710	11,064	15,703	0	15,703
% Matched	93.33	93.02	92.81	94.35	94.98	-	94.98
<b>Arecibo</b>							
Total	106,890	67,197	44,385	22,812	39,693	0	39,693
Matched	101,967	64,598	42,130	22,468	37,369	0	37,369
% Matched	95.39	96.13	94.92	98.49	94.15	-	94.15
<b>NON-SMSA:</b>							
<b>Remainder of Island</b>							
Total	1,040,295	456,375	361,489	94,886	583,920	24,697	559,223
Matched	963,149	422,379	337,011	85,368	540,770	23,146	517,623
% Matched	92.58	92.55	93.23	89.97	92.61	93.72	92.56

Table 4: Non-Matched Persons in Households by Puerto Rico and District Offices  
(Numbers in Thousands)

Numbers in Thousands (000)	Total Persons	NON - MATCHED PERSONS					
		Total	Percent	Within Whole Household Non-Matches		Within Partial Households Non-Matches	
				Total	Percent	Total	Percent
Puerto Rico	2,370	184	7.77	107	4.51	77	3.26
DO 3301	277	28	10.36	16	5.94	12	4.43
DO 3302	186	19	10.41	13	7.26	6	3.15
DO 3303	292	19	6.55	10	3.46	9	3.09
DO 3304	339	24	6.97	9	2.58	15	4.39
DO 3305	329	30	8.95	21	6.37	9	2.59
DO 3306	308	19	6.30	10	3.38	9	2.92
DO 3307	357	21	5.79	12	3.33	9	2.45
DO 3308	282	24	8.48	15	5.27	9	3.21
DO 3301 & DO 3302 Combined	462	48	10.38	30	6.47	18	3.91
DO 3303 - DO 3308 Combined	1,907	136	7.14	77	4.03	59	3.11

PUERTO RICO



Acknowledgements

I would like to thank all the persons who assisted in this project. Special thanks are due to Jon Clark, Hans Muller and Howard Hogan for their suggestions and assistance in the matching; to John Thompson, Carma Hogue and Deborah Harner for their reviews; and to Jeanne Washington for her patience in typing the many revisions to this paper.