

### **3. ALTERNATIVE MEASURES OF SIZE**

#### **3.1 Overview of Alternative Measures of Size**

The 1988 MCBS PSU sample was based on the 1980 Westat Master sample, which used 1980 Census population counts as the measure of size (MOS). At that time, these data were the most recently available. Since that time, of course, the overall population count in the U.S. has increased; furthermore, the population distribution changed, with declining populations in the Northeast and increasing populations in the South and other areas.

In this chapter, we will examine the impact of these changes on the original MCBS sample design. In our analysis, we compare the existing MOS for MCBS PSUs with two alternatives. One alternative would be 1990 U.S. Census population counts. The other would be 1999 counts of Medicare beneficiaries in the U.S. provided to us by HCFA. We will also evaluate the effects of weighting the HCFA counts as would be done in the MCBS to equalize workloads for the seven study age groups.

In general, HCFA counts of Medicare beneficiaries would seem to be the best choice of the MOS for MCBS because one might expect the Medicare population to be distributed differently across geographic regions as compared with overall Census population counts. The migration of the Medicare population may be uneven across regions because retirees tend to move to retirement communities, less expensive areas, etc. and these areas are unevenly distributed across U.S. As the goal of MCBS is to study data on Medicare beneficiaries, the HCFA MOS should better capture these changes. On the other hand, the vast majority of Medicare beneficiaries "age in place" or move to homes in nearby communities, so the difference may not be great.

#### **3.2 Descriptive Statistics for Alternative MOS**

##### **Overview**

We compared the original 1980 PSU MOS with the two alternative MOS, 1990 U.S. Census population counts and 1999 HCFA Medicare beneficiaries counts, for the 1,396 PSUs on the MCBS sampling frame.

PSUs in Puerto Rico were treated slightly differently with respect to these comparisons. When the MCBS sample was drawn, 1990 Census population counts were used for the MOS for Puerto Rico. Thus Puerto Rico PSUs are excluded from the comparison of 1980 MOS with 1990 MOS. These PSUs are included in the comparison of the original MCBS MOS with HCFA MOS.

In designing future surveys, the Medicare population in Puerto Rico is large enough to constitute a separate stratum. Thus a separate sample of Puerto Rico PSUs could be drawn, guaranteeing their representation in the MCBS.

### Univariate Distributions and Correlations

The results of the comparison of original PSU MOS with the two alternatives with respect to the basic descriptive statistics are presented in Table 3-1. The comparison of 1980 and 1990 population counts in Table 3-1 reflects the increased population over that decade through mean MOS and all quantile statistics. The mean number of Medicare beneficiaries is approximately 15 percent of the mean 1990 population count.

Table 3-1. Comparison of descriptive statistics for three measures of size

MOS	Mean	Median	25 <sup>th</sup> pctl.	75 <sup>th</sup> pctl.	Min	Max
1980 population counts	163,758	52,643	33,041	104,171	5,530	7,477,503
1990 population counts	179,784	54,753	33,190	112,074	5,817	8,863,164
1999 beneficiary counts	27,619	9,985	5,936	18,928	866	969,522

Table 3-2. Correlation coefficients for three measures of size

	1990 population counts	1999 beneficiary counts
1980 population counts	0.99367	0.97912
1990 population counts	,	0.97621

However, the alternative measures of size are closely correlated, as shown in Table 3-2. The correlation between 1980 and 1990 population counts is 99.4 percent, indicating that the growth of population from 1980 to 1990 was roughly proportional across PSUs. The correlation between the population counts and HCFA MOS was approximately 98 percent, again indicating a strong linear relationship between either population count and Medicare beneficiaries. After dropping the largest outliers, correlations among the variables still exceeded 97 percent.

### Scatterplots and Ratio of MOS

Figures 3-1 and 3-2 give scatterplots of 1980 MOS with the two alternative MOS. In both plots, there are several outliers from the trend line. These outliers are more pronounced for the comparison of 1980 population counts with 1999 HCFA counts.

To provide another analysis of disproportionate growth or decline in some PSUs, we looked at the ratio of 1980 MOS to each of the two alternative PSU MOS. If MOS changes proportionately, then the ratio statistic is expected to have similar values across PSUs. Table 3-3 displays descriptive statistics for the two ratios. Again, the ratio statistics indicate more extreme differences between the 1980 population counts and the HCFA counts than between the 1980 and 1990 population counts.

Table 3-3. Descriptive statistics for the ratio of original MOS to 1990 MOS and HCFA MOS

Ratio	Mean	Median	25 <sup>th</sup> pctl.	75 <sup>th</sup> pctl.	Min	Max
1990 MOS ÷ 1980 MOS	1.0438	1.0211	0.9581	1.0936	0.7246	2.2738
HCFA MOS ÷ 1980 MOS (standardized)	1.0000	0.9519	0.8606	1.0571	0.4732	4.9816

Table 3-4 lists the largest outliers based on the ratio of MOS values. These outliers are also indicated on Figures 3-1 and 3-2. The first panel shows the outliers for the 1980 versus 1990 population counts, while the second panel gives the outliers for the 1980 counts versus HCFA counts.

Table 3-4. Comparison of original 1980 PSU MOS with two alternative MOS

Comparison of 1980 MOS and 1990 MOS				
Stratum / PSU	County, State	1980 MOS	1990 MOS	RATIO
A410 / 001	Los Angeles, CA	7,477,503	8,863,164	1.1853
A340 / 001	Dade, Palm Beach, Broward, FL	3,220,844	4,056,100	1.2593
A320 / 001	Collin, Dallas, Kaufman, Ellis, Rockwall, Denton, Tarrant, Johnson, Hood, Wise, Parker, TX	2,974,805	3,949,075	1.3275
B460 / 005	San Bernardino, Riverside, CA	1,558,182	2,540,000	1.6614
Comparison of 1980 MOS and HCFA MOS				
Stratum / PSU	County, State	1980 MOS	HCFA MOS	Std. Ratio
A410 / 001	Los Angeles, CA	7,477,503	969,552	0.7011
A210 / 001	McHenry, Will, Du Page, Cook, Lake, Kane, IL	7,103,624	926,476	0.7052
A340 / 001	Dade, Palm Beach, Broward, FL	3,220,844	791,892	1.3294
B330 / 001	Pinellas, Hillsborough, Pasco, FL	1,569,134	422,889	1.4573
B460 / 005	San Bernardino, Riverside, CA	1,558,182	359,886	1.2489
B460 / 001	Maricopa, AZ	1,509,052	354,963	1.2719

In comparing 1980 and 1990 population counts, the Los Angeles, Miami (Dade County), Dallas/Ft. Worth (Tarrant County), and San Bernardino PSUs grew at greater rates than most PSUs. This can also be seen in Figure 3-1, when the coordinates for Los Angeles, Miami, Dallas/Ft. Worth, and San Bernardino fall above the trendline.

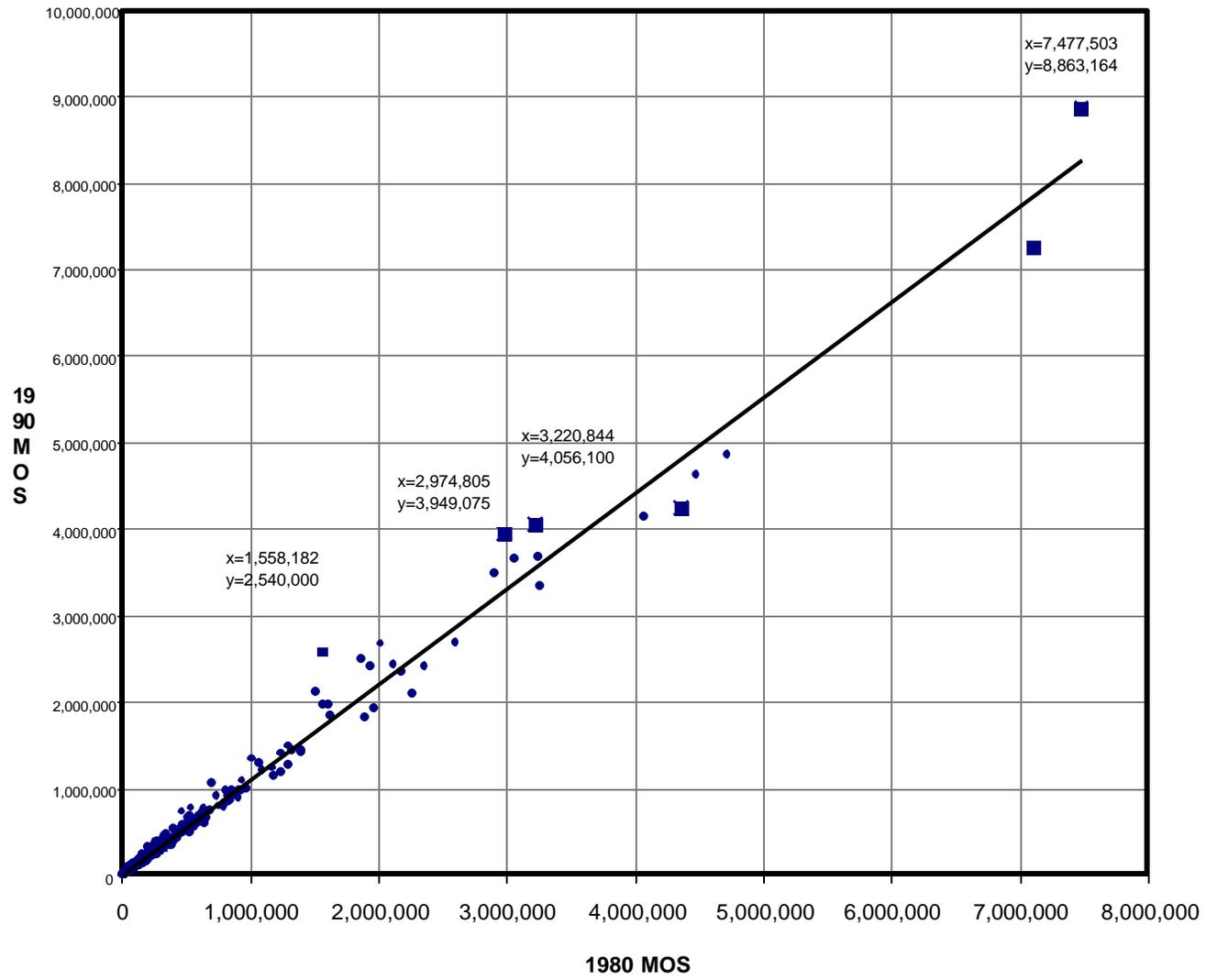


Figure 3-1. Comparison of 1980 and 1990 MOS

Comparing the 1980 population counts with 1999 counts of Medicare beneficiaries, the Tampa (Pinellas County), Miami (Dade County), Phoenix (Maricopa County), and San Bernadino PSUs had more beneficiaries than would be expected based on 1980 population counts. The Los Angeles and Chicago (Cook County) PSUs had fewer beneficiaries than would be expected based on the 1980 population counts. These relationships can be seen in Figure 3-2 as well.

### **Weighted HCFA MOS**

In practice, measures of size are often weighted by study group sizes so that study group sizes are achieved while controlling workloads within PSUs. The weighted MOS is a weighted sum of the number of Medicare beneficiaries in each of the seven age groups, using a weight that reflects the sampling rate for the group. The weighted sum is computed for each PSU. Because this weighted measure of size could be used in a redesigned survey, we have done an analysis comparing it with the 1980 MOS.

As might be expected, the weighted HCFA MOS and the unweighted HCFA MOS are very closely related. For example, the mean and median of the weighted HCFA MOS are 27,773 and 10,143, respectively, which are both well within 2% of the corresponding values for the unweighted HCFA MOS. As would be expected, the weighted HCFA MOS is strongly correlated with 1980 MOS, with a correlation of 0.9798.

Figure 3-3 gives a plot of the weighted HCFA MOS versus the 1980 MOS. The outliers in this plot are the same as the ones in Figure 3-2; that is, the outliers using the HCFA MOS are the same regardless of whether the MOS is weighted or not.

All of these results indicate that using a weighted HCFA MOS could be substituted for the unweighted HCFA MOS.

### **3.4 Number of Certainties with Alternative MOS**

Since we observed some differences between the original 1980 PSU MOS and the two alternative PSU MOS, it will be of interest to see how the definition of certainty PSUs is affected by alternative MOS.

The general method for determining the PSUs to be selected with certainty can be described as follows. First, the PSUs are sorted from largest to smallest. For the first PSU (i.e., the one with the largest MOS), we computed the ratio of the MOS for the PSU to the total MOS, and multiplied this ratio by PSU sample size. If the resulting number is at least 1.0, then the first PSU is selected with certainty. For the next largest PSU, the MOS for this PSU was divided by the total MOS minus the MOS of the last PSU; this ratio is then multiplied by the PSU sample size minus one. As before, the PSU is selected with certainty if the resulting number is at least 1.0. This process is repeated until the calculated value is less than some cutoff value, which may range from 0.50 to 1.0; in their classic sampling text, Hansen et al. suggest a cutoff of 0.67 (Hansen et al., 1953).

As noted earlier, the MCBS sample is based on the 1980 Westat Master sample. This sample, which used 1980 population counts as the MOS, had 20 certainties. In modifying this sample for the MCBS, 13 certainty PSUs were added to make 33. The 1990 Master Sample (based on 1990 population counts) has 22 certainties. With a measure of size based on Medicare beneficiaries, we estimate that between 26 and 39 certainties would be designated, depending on the certainty cutoff used.

We worked out three different scenarios for defining certainty PSUs based on the original 1980 MOS, 1990 MOS, and HCFA MOS. For each of the three MOS we listed all PSUs in the continental U.S. and Puerto Rico in descending order for the given MOS. MCBS calls for selecting 107 PSUs.

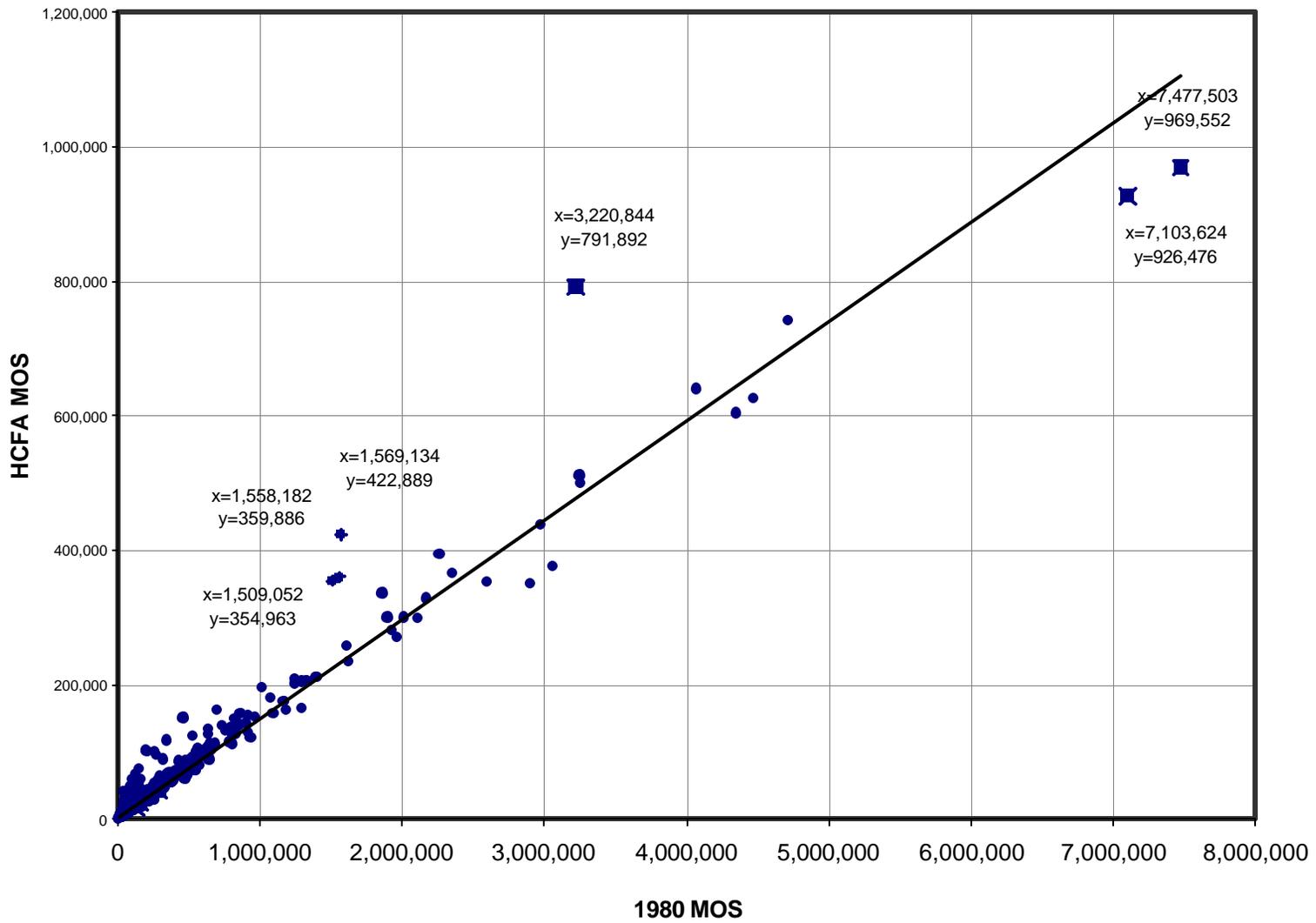


Figure 3-2. Comparison of 1980 MOS and HCFA MOS

Table 3-5 compares the certainty PSUs for the original 1980 PSU MOS and two alternatives 1990 Census counts and 1999 HCFA MOS. For comparison purposes, we also provide a listing of certainty PSUs that were used as certainty PSUs in the MCBS design.

The issue of the certainty cutoff deserves some discussion here. For the 1980 master sample, the certainty cutoff was 50 percent; in 1990, it was 100 percent. For the HCFA MOS, we considered 50 percent, 75 percent, and 100 percent, with 75 percent being used in Table 3-5. Prior to designing a new sample for MCBS, further research would be required to establish the most efficient cut-off.

The comparison of certainty PSUs with respect to the three PSU MOS shows that most of the 1980 certainty PSUs are identified as certainties regardless of MOS. However, neither the 1980 MOS, the 1990, nor the HCFA MOS contains as many certainties as the MCBS design. However, as described in Section 2.2.2, the designation of certainties for the MCBS had several constraints that would not be present in a redesign of the survey. Finally, we note that if the certainty cutoff were extended to 50 percent, the number of certainties for the HCFA MOS would increase to 39 PSUs. As noted earlier, the optimum certainty cutoff for a redesigned MCBS would require further study.

### **3.5 Impact of Redesign on PSU Workloads**

One of the reasons for selecting a sample of PSUs with probability proportional to size is to equalize workloads for the noncertainty PSUs. The more accurate the measure of size, the more balanced the workloads are.

Using Medicare beneficiaries as a PSU measure of size would result in almost perfectly balanced workloads in the new sample PSUs first few years following the redesign. However, even this measure of size will deteriorate over time if the population of beneficiaries shifts among PSUs. Consequently, the workloads may become unbalanced if the number of beneficiaries shifts away from the values used to design the sample.

### **3.6 Tracking Future Changes in MOS**

As indicated in the previous section, even the best measure of size can deteriorate if the distribution of the population changes. In this section we propose two methods for monitoring changes in the population that will indicate a need for redesigning the survey.

The first measure of changing MOS is variation in the workloads in noncertainty PSUs. This can be monitored either through the coefficient of variation of the sample sizes (by panel, say) or through the minimum and maximum sample sizes. When the variation in sample workloads begins to affect cost-effectiveness of survey operations, then a redesign should be considered.

A second measure of a need for survey redesign is the related to the coefficient of variation of the sampling weights. As the measure of size for the noncertainty PSUs deteriorates, the variation in sample baseweights increases, causing a corresponding increase in the variance of survey estimates. The degree of inflation of the variance can be measured by the factor  $(1 + CV^2)$ , where “CV” indicates the coefficient of variation of the sample baseweights. Using data from the 1999 Round 25 panel, this approximation gives  $(1 + CV^2) \approx 1.08$ , indicating an inflation of the variances by about 8% due to unequal sampling weights.

Table 3-5. Certainty PSUs based on 1980, 1990, and HCFA MOS versus MCBS Certainty PSUs

Count	Stratum	PSU	1980 Sample <sup>1</sup>	1990 Sample <sup>2</sup>	HCFA MOS <sup>3</sup>	MCBS Sample <sup>4</sup>
1	A111	001	1	1	1	1
2	A112	001	1	1	1	1
3	A113	001	1	1	1	1
4	A120	001	1	1	1	1
5	A130	001	1	1	1	1
6	A140	001	1	1	1	1
7	A150	001	1		1	1
8	A210	001	1	1	1	1
9	A220	001	1	1	1	1
10	A230	001	1	1	1	1
11	A240	001	1		1	1
12	A250	001	1	1	1	1
13	A310	001	1	1	1	1
14	A320	001	1	1	1	1
15	A330	001	1	1	1	1
16	A340	001	1	1	1	1
17	A350	001	1	1	1	1
18	A360	001	1	1	1	1
19	A410	001	1	1	1	1
20	A420	001	1	1	1	1
21	B320	002				1
22	B330	005			1	1
23	B350	001				1
24	B350	002				1
25	B410	001				1
26	B410	002			1	1
27	B420	002		1	1	1
28	B420	020		1	1	1
29	B440	002				1
30	B450	012				1
31	B460	001		1	1	1
32	B460	005		1	1	1
33	PR1	001				1
Sum			20	22	26	33

<sup>1</sup> The 1980 Master Sample was based on 1980 population counts with a certainty cut-off of 0.50.

<sup>2</sup> The 1990 Master Sample was based on 1990 population counts with a certainty cut-off of 1.00.

<sup>3</sup> The certainties for the HCFA MOS were based on a cut-off of 0.75.

<sup>4</sup> The MCBS certainties consisted of the 1980 Master Sample, modified as discussed in Section 2.2.2.