WITHIN-PSU SAMPLING FOR THE CENSUS BUREAU'S DEMOGRAPHIC SURVEYS

Edward Kobilarcik, Ronald Statt, and Thomas F. Moore
Bureau of the Census

INTRODUCTION. Major research projects are being conducted at the Census Bureau to redesign the demographic surveys for the 1980s. One project includes research into the method of selecting the sample within Primary Sampling Units (PSUs). It is likely that this method will be used by most or all surveys when the redesigned surveys are introduced. This paper presents the results from an investigation designed to develop a within-PSU sampling procedure which would be efficient for all surveys and simultaneously avoid the selection of any element for more than one survey.

The method involves multivariate techniques to cluster similar census Enumeration Districts (EDs). The EDs are first sorted within the clusters and then the clusters are sorted before the sample selection. The variables used to cluster and sort EDs will be presented. The relative efficiency of several alternatives are also discussed.

SELECTION OF THE PRESENT SAMPLES. The present method used by the demographic surveys was originally developed primarily for the Current Population Survey (CPS). The samples of households are selected by means of a multistage stratified sampling of the U.S. population. Once the sample PSUs (normally counties or groups of counties) are identified, the addresses of sample housing units are selected in two steps. The first step is the selection of a sample of census EDs. These are administrative units designated in the 1970 census. They cover the entire country and contain, on the average, about 350 housing units. Computer summary records of the EDs are sorted by geographic variables. Then a systematic sample of EDs is drawn with the result that the sample EDs are spread over the entire PSU. The probability of selection of any ED is proportionate to its 1970 population.

Next a cluster, or Ultimate Sampling Unit (USU), of about four addresses, geographically contiguous, if possible, is selected from each designated ED. Where possible, this selection is done from the list of addresses for the EDs compiled during the 1970 census. This list sample is supplemented by a selection of housing units constructed since the census. If a significant proportion of addresses within an ED are incomplete or inadequate, area sampling methods are used. The within-PSU sampling interval is determined so that the overall probability of selection of each household in the frame is equal, i.e., the samples are self-weighting. (NOTE: During the 1970s, several expansions to CPS to improve the reliability of estimates for states and substate areas resulted in a sample that is no longer self-weighting.)

Several samples are required for CPS over the life of the design since each sample is gradually replaced by another sample with the same specifications as the original. This is achieved by associating the USUs adjacent to the initial selections with the subsequent samples. In this way, a large number of matching samples (i.e., sets of adjacent USUs) can be designated. Samples for other surveys are identified by assigning them to all or a subset of as many matching samples as necessary. More details on the overall design may be found in [1].

PURPOSE OF RESEARCH. The redesign provides an opportunity to develop a new procedure for within-PSU sampling which would take into account the specific needs of all the individual surveys and simultaneously avoid sample duplication.

Other surveys which will share the basic design with the CPS include the Annual Housing Survey (AHS), the National Crime Survey (NCS), the Health Interview Survey (HIS), and a General Purpose Sample (GPS).

Even modest variance reductions from the use of an improved within-PSU sampling procedure represent substantial savings over the decade compared to the cost of obtaining similar reductions by increasing sample sizes.

Two questions for which answers are needed in order to develop a new procedure are:

1. What intermediate stages, if any, of within-PSU sampling should be used; i.e., should EDs continue to be used in selecting the sample or would something else such as census blocks, block groups, housing units, etc. give better results?
2. How should the units used in the intermediate stage within PSUs be arranged before sampling?

The second question was the major objective of our research and, as such, most of our efforts were put to use in answering this question. Time limitations and practical considerations hindered our investigation of the first question.

To answer these questions, we considered potential procedures with respect to several factors including sampling and field costs, variance, implementation time, complexity, and the impact of the procedures on small PSUs and on future sample selection.

SORT UNIT. It was assumed that the choice of geographic unit for use in the intermediate stage of sampling within PSUs would influence our eventual recommendation. The following units were considered as possible choices for use in this stage of the sample selection: tract, ED, block group, block, block face, address, and housing unit.

An initial study was performed which investigated the feasibility of using each unit. [2] After this study, block face, housing unit, and address were not given further consideration because of added complexity and cost of survey and sampling operation that would be incurred if these were selected.

Preliminary variance computations for some items showed that a large percentage of the within-PSU variance may be due to the variability within EDs. Thus, there was little interest in examining units larger than EDs such as census tracts. Since the within-ED variance is relatively large, smaller units would probably be more effective in reducing the variance.

The only units besides EDs which were given further consideration were blocks and block groups. It did not seem appropriate to use block groups over EDs because these units appear to be relatively similar in nature to EDs. Also EDs would be more convenient because they provide entire coverage of the U.S.

(Initially, time constraints prevented a thorough study of the use of blocks. However, the use of blocks is being investigated at this time as the original time constraints have been relaxed.)

For the above reasons, we have tentatively decided that the ED should remain the unit used in the intermediate stage of within-PSU sampling.

DETERMINATION OF VARIABLES. It would have been desirable to arrange the EDs within PSU based on the items estimated by a particular survey, for example, in the CPS, using labor force characteristics. However, the sample data which includes some relevant labor force data from the 1980 census were not expected to be available for use at the time of the redesign according to the schedule in effect when the research began. For this reason, it was necessary to determine a set of 100-percent census variables correlated to several key items. In the determination of these variables, a large number of socio-economic characteristics were considered in addition to the present variables which are mostly geographic.

For the most part, these variables were selected by analyzing simple, partial and multiple correlations at the ED-level between 1970 census 100-percent variables and variables of interest. Other variables were chosen because they were believed to be related to survey objectives.

Some of the variables felt to be correlated with labor force variables included the number of persons 16+ for total and minority population and housing characteristics; for health, housing and age characteristics were chosen; for crime, housing and urban characteristics were chosen; and for housing, various housing and urban characteristics were chosen. Many of the variables were useful for more than one type of estimate. Although the research used 1970 census data, the actual sample selection will be based on 1980 data.

METHODOLOGY

Two Techniques. In order to maximize the advantage of systematic sampling within PSUs, it is requisite to arrange EDs with similar characteristics adjacent to each other. Two techniques of arranging EDs were investigated. The first technique used a modification of the Friedman-Rubin algorithm to form clusters of EDs. [3] and [4] The second, a more straightforward technique, rank-ordered EDs according to combinations of census variables that were highly correlated to survey variables of interest. Both of the techniques were examined to determine their effectiveness in reducing the systematic sampling variance of an estimator of total for many types of demographic characteristics.

The first technique strived to form homogeneous clusters of EDs within PSUs such that a criterion function was reduced. The modified Friedman-Rubin clustering algorithm used in the technique was based on the variance formula for sampling with probability proportionate to size (pps) with replacement. As such, the functions did not measure the systematic sampling variance; however, we feel the use of this function was still highly effective in creating homogeneous clusters of EDs. This criterion function F for estimates of levels for a single variable is given by the following formula:

$$F = \frac{1}{n} \sum_{h=1}^{g} \sum_{i=1}^{t_h} \frac{M_{hi}}{M_h} \left(\frac{M_h}{M_{hi}} x_{hi} - x_h \right)^2$$

where g = the number of clusters (fixed)

t_h = the number of EDs in the h^{th} cluster

M_{hi} = the number of USUs in the i^{th} ED of the h^{th} cluster

$M_h = \sum_{i=1}^{t_h} M_{hi}$ = the number of USUs in the h^{th} cluster

x_{hi} = the total number of persons with the characteristic of interest in the i^{th} ED of the h^{th} cluster (100 percent census variable)

$x_h = \sum_{i=1}^{t_h} x_{hi}$ = the total number of persons with the characteristic of interest in the h^{th} cluster (100 percent census variable)

$n = n_h$ = the number of EDs sampled from each cluster.

The final step of this technique was to form clusters of EDs such that the sum of F was minimized for several variables for a given number of clusters. Although g is not known, its desired value depends on the number of EDs within sample PSUs.

The modified Friedman-Rubin clustering algorithm was used as the basis in attempting to accomplish the above goal. The main part of their algorithm used in this investigation was the procedure referred to as the "hill-climbing" pass. Before implementation of this procedure, an initial allocation of EDs to clusters were formed so that their 1970 population sizes were approximately equal. The hill-climbing pass examined the reduction to the sum of the function F over all variables as each ED was moved one at a time to a different cluster. The ED was placed in that cluster where the reduction to the sum of F was the greatest. If no reduction occurred, the ED remained in its current cluster. The procedure terminated when an entire pass of the EDs produced no further reassignment of EDs to a different cluster. Upon completion of the hill-climbing pass, the clusters and the EDs within clusters were rank-ordered.

In an attempt to prevent the clustering algorithm from concentrating on reducing the criterion function for those variables for which the initial value of F was the greatest, the variables used to determine the cluster were standardized by dividing all observations of the j^{th} variable by its standard deviation. In this way, the variables used to determine the clusters were considered equal in importance.

The second technique of arranging EDs simply rank-ordered EDs with respect to two or more variablesthat were correlated to the survey characteristics of interest. The procedures in this technique were called sorts. The present procedure is a version of this technique. The EDs were sort-ordered based on $p(p \geq 2)$ variables in which p-1 of the variables were designated to group EDs and the remaining variable used to rank-order EDs within each of the predefined groups. Variables designated to group the EDs were measurable, such as the proportion black, or qualitative such as the urban/rural location. Groups under a measurable variable were either formed by:

(1) clustering EDs having the same value of the measurable variable, or
(2) clustering EDs based on the value of the measurable variable within intervals defined by cutpoints.

Once the EDs were arranged according to a specific technique, the EDs were sampled systematically with probability proportionate to size.

Evaluation. The evaluation of the two within-PSU sampling techniques was performed in two steps. In the first step, the effects of the two different techniques on variables of interest were analyzed separately in an attempt to determine the optimal procedure for each survey. The second step consisted of a coordinated approach to select procedures that could be utilized by all surveys at the time of the redesign.

To aid in the research, two computer programs were developed. One program was used to simulate the present sort as well as alternative sorts. The second program clustered the EDs based on the modified Friedman-Rubin algorithm.

For a given PSU, the EDs would first be arranged according to a specific technique and variable set. When the arrangement was complete, a number of systematic samples of EDs were selected with probability proportionate to size using a given sampling interval (SI). The number of USUs selected from the i^{th} sample ED was dependent on M_i, the total number of USUs in the i^{th} ED. For each sample ED, zero or one USUs were selected if $M_i \leq SI$ and one or more USUs were selected if $M_i > SI$. Since detailed survey characteristics were not available at the USU level, x_i/Mi (see below) was used to approximate this data. If two USUs were selected within an ED, the USU estimate x_i/Mi was used twice when computing the sample estimate. For a given sample k, the estimate $\hat{Y}_k$ was computed for each variable of interest using the 1970 census countat the ED level by the following formula:

$$\hat{Y}_k = SI \sum_{i=1}^{T} m_{ik}(x_i/M_i) \ .$$

where x_i = the total number of persons with the characteristic of interest in the i^{th} ED

M_i = the total number of USUs in the i^{th} ED

m_{ik} = the total number of sample USUs in the i^{th} ED for the k^{th} sample

T = the total number of EDs in the given PSU

SI = the sampling interval of the given PSU

The systematic sampling variance of the estimated total was computed as:

$$\hat{V}_{SYS}(\hat{Y}) = \begin{cases} \frac{1}{K}\sum_{k=1}^{K}(\hat{Y}_k-\hat{Y})^2 & \text{for } K=SI \text{ (all possible systematic samples were selected)} \\ \frac{1}{K-1}\sum_{k=1}^{K}(\hat{Y}_k-\hat{Y})^2 & \text{for } K<SI \text{ (less than all possible systematic samples were selected)} \end{cases}$$

where K = the number of samples selected

$\hat{Y}$ = the estimate of the number of persons with the characteristic of interest in the given PSU

$$\hat{Y} = \frac{1}{K}\sum_{k=1}^{K}\hat{Y}_k$$

It should be noted that since each USU in the i^{th} ED was assigned the value x_i/M_i, the systematic sampling variance only explained the variance between EDs, not within EDs.

This variance estimate was the key statistic used to evaluate the various procedures under the two techniques. The optimum procedure was deemed the one which produced the lowest systematic sampling variance for most items of interest in most of the areas examined.

Since the variances for the items were very different, we needed a reference point for each characteristic for evaluating the variances obtained under the two techniques. To fix this reference point, we computed the expected variance for each characteristic over all possible sorts. The following formula, developed by Hartley and Rao in[5], gives the expected value of the variance using a systematic sample, assuming a random arrangement of units with unequal measures of size. The formula was applied using the ED as the unit and the ED mean as the value for the unit. The Hartley/Rao formula assumes that $M_i < SI$ for all units. Since for some EDs $M_i > SI$, it became necessary to modify the formula. This modification consisted of replacing x_i by x_i' as defined below. This variance is estimated by:

$$\hat{V}_{RSYS}(\hat{Y}) \doteq \frac{SI}{(M_o')^2}\sum_{i=1}^{T} M_i'\left[1-\left(\frac{M_o'}{SI}-1\right)\frac{M_i'}{M_o'}\right]\left[\frac{M_o'}{M_i'}x_i'-Y'\right]^2$$

where:

$x_i' = \frac{M_i'}{M_i}x_i$ where $M_i' = M_i - SIq_i$ where q_i is the greatest positive integer such that $SIq_i \nmid M_i$,

$$\hat{Y} = \sum_{i=1}^{T} x_i' \ ,$$

$$M_o' = \sum_{i=1}^{T} M_i' \ ,$$

and M_i, SI, x_i, and T have been previously defined.

To provide some measure of efficiency of each procedure, the ratio of the systematic sampling variance to the systematic sampling variance assuming a random arrangement of EDs was computed:

$$\hat{V}_{SYS}(\hat{Y})/\hat{V}_{RSYS}(\hat{Y}) \ .$$

Due to the large number of variables that were being examined in the coordination phase, it was difficult to assess the effectiveness of any one procedure for a given area by examining each individual estimate. For this reason, a number of summary statistics were computed.

Geometric means were computed from the variance ratios for a select group of variables for each survey. These means or indices were a useful tool in determining the effectiveness of a procedure in meeting the specific needs of each survey.

The PSUs examined were a fairly representative sample of the U.S. They were chosen based on the type and size of the PSU. Also included were particular types of PSUs that possessed certain characteristics considered to be important, such as the minority and poverty status. This process might have caused some types of areas to be over-represented, especially areas with atypical distributions of minorities.

Initially, a set of procedures based on the rank-ordering of EDs (sorts) were selected by PSU. These were to be used for comparison with the Friedman-Rubin procedure for the same PSUs. Various combinations of variables were examined in pilot PSUs with the inferior ones eliminated from further consideration. These were eventually narrowed down to six candidate sorts which were applied to other PSUs. We looked at some or all of the sorts in determining the best sorts for each type of PSU. The size and characteristics of the PSU were used as guides to decide which sorts were to be examined in which PSUs.

It was not always possible to reach a consensus when deciding which of the procedures was better in a PSU. The index of variables was not always an adequate indicator of whether a particular procedure was effective. At times the index was misleading because it disguised the effects of a procedure on some of the more important items. For example, one procedure may have given generally good results for labor force data but poor results for estimates of unemployment, in particular. Thus, the indices were only used as a guide; individual items were also examined before making a final decision.

For the final recommendation, the PSUs were eventually grouped into the following size categories:

1) PSUs with total population 500,000 or more;
2) PSUs with total population between 125,000 and 500,000;
3) PSUs with total population less than 125,000.

RESULTS

Most Effective Sort-Order Techniques. A small number of sort-orders were found to be generally more effective than others for key items. The effectiveness of a sort depended on the type of PSU examined. Different sorts gave better results depending on the total population and the minority status of the PSU. The preferred sort for a particular PSU was the sort that produced the lowest variance for most items of interest. The following sorts were selected for further testing and comparisons in our final evaluation of this technique. The variables in the sorts are listed in the hierarchial order in which the EDs were grouped and sorted.

SORT A : CBUR code
Place size
ED number

This is the sort procedure used in the current design. The EDs are first sorted by CBUR code, a geographical classification of EDs, so that EDs of the same classification are placed together where:

C - represents the central city of an SMSA
B - represents an urbanized area not in category C
U - represents an urban place not an urbanized area and not in category C
R - represents all other EDs (mainly rural)

Within each of these categories, EDs from the largest place sizes are assembled first, followed by those from smaller places. (Place refers to a concentration of population such as a city, town, village, and borough). The EDs are finally sorted by ED number, a geographic code which tends to place geographically contiguous EDs together.

SORT D : CBUR code
ED number

SORT F : Median value of owner-occupied housing units
Percent Black or Spanish
ED number

SORT F′: Median value of owner-occupied housing units
Percent both Black and Spanish
ED number

The sorts found to be most effective for each of the size categories of PSUs were the following:

1) For PSUs with total 1970 population 500,000 or more, sorts A, D, F, and F′ were selected.
2) For PSUs with total 1970 population between 125,000 and 500,000, sorts D and F were selected.
3) For PSUs with total 1970 population less than 125,000, sorts A and F were selected.

It was reassuring to find that the current sort-order (Sort A) was generally highly effective in reducing the between-ED variance for most items examined. In fact, in some PSUs, the present sort was selected as the preferred sort among all others examined. Still, if our recommendation had been based solely on the examination of this technique, we could have expected improvements over the current within-PSU sampling procedure using different sorts for different types of PSUs.

Most Effective Modified Friedman-Rubin (MFR) Techniques. A few methods based on the modified Friedman-Rubin (MFR) technique were found to be more effective than others, depending on the population size of the PSU. Subsequently, a single MFR technique was selected for each of the size categories of PSUs. The variables used to form clusters, to rank-order the clusters, and to order the EDs within clusters are shown below. The variables were selected based on their effectiveness in the sort-order technique.

1. Formation of Clusters. The variables used to form the clusters, which are identical for each of the PSU size categories, are the following: 1) median value of owner-occupied housing units, 2) Spanish population (if percent Spanish > 5%), 3) Black population (if percent Black > 5%), 4) housing units occupied by owner, 5) CBUR code (as defined in previous approach, 6) number of one-room housing units, 7) population under 6 years old, 8) population 65 years old and over, 9) number of mobile homes or trailers, 10) number of units lacking some or all plumbing facilities, 11) number of owner-occupied housing units of value < $15,000, 12) number of renter-occupied units with rent < $100/month.

2. Rank-Ordering of Clusters. The manner in which the clusters are rank-ordered differentiates the MFR techniques for the PSU size categories.

- For PSUs with 1970 total population greater than 500,000: rank-order clusters based on the median value of owner-occupied housing units (MFR-A).
- For PSUs with 1970 total population between 125,000 and 500,000: rank-order clusters based on CBUR code and clusters within CBUR categories by median value of owner-occupied housing units (MFR-B).
- For PSUs with 1970 total population less than 125,000: rank-order clusters based on number of housing units occupied by owner (MFR-C).

3. Rank-Ordering Within Clusters. For each of the PSU size categories, the EDs within clusters are rank-ordered based on ED number.

Tables in the index display the survey indices obtained from the current sort (A), the preferred sort, and the MFR technique for selected PSUs within each of the size of categories of PSUs.

RECOMMENDATION AND CONCLUSIONS. Based on the results from our investigation, it was tentatively recommended that the following versions of the Modified Friedman-Rubin algorithm, which depend on the type of PSU to be used to cluster and sort EDs within sample PSUs before systematic sample selection.

- In PSUs with total population 500,000 or more, use MFR-A.
- In PSUs with total population of 125,000 or more, but less than 500,000, use MFR-B.
- In PSUs with total population less than 125,000, use MFR-C.

Further investigation is needed to determine the best procedure for the smaller PSUs. Our tentative recommendation to use MFR-C may be reconsidered since some of the sorts also worked well. Since the MFR technique had been generally superior for the other types of PSUs, we initially preferred to choose among the MFR procedures for these types of PSUs when making our recommendation.

Compared to the sort-order technique, the Friedman-Rubin technique produced more consistent results throughout all types of PSUs. In addition to producing superior results with respect to variances, the choice of the Friedman-Rubin was satisfying for other reasons. Use of this technique added a level of flexibility in allowing for differences between PSUs. Only a small number of groups of PSUs could be formed for determining different procedures because of the relatively small number of areas examined. The alternative to Friedman-Rubin was a fixed sort for each group. However, within each group, the Friedman-Rubin clustering algorithm treated each PSU differently. If one item used for clustering were more variable from ED to ED than another item it would have a greater impact on the final arrangement of EDs for that PSU. In other words the Friedman-Rubin technique was more tailored to characteristics of individual PSUs than were the sorts. The Friedman-Rubin technique also has the capability of using more information than the sorts.

LIMITATIONS OF THE STUDY. There were several limitations to this study; still, we feel that the improvements offered by the Friedman-Rubin technique warrant its use. Some of the limitations apply to the current procedure as well.

It is possible that as the characteristics of various areas change over a 10-year period, variances of estimates using some other procedure would become better than those from the recommended one. We have no evidence of this but it would be desirable to compare estimates from different procedures over time using past data.

Given more time, we could have examined the effect of different sampling intervals on the systematic sampling variances. Sometimes a small change in the sampling interval in a PSU caused a large change in the variance for some items. We could have tried several sampling intervals in each area but we decided to use our resources on more areas rather than more sampling intervals. We feel that we looked at enough areas and items that we would not have made any major changes to our final recommendation had we looked at more sampling intervals.

Most of the survey variables used for evaluating the techniques were subject to sampling in the 1970 census. We did not try to estimate the effect, if any, on our comparisons of different procedures.

There are other areas where the recommended procedure might be improved or refined. Adding to or deleting variables might yield better results, for example.

The 5 percent cutoff necessary for including minority variables in the clustering algorithm and the population sizes for the different versions of the procedure were determined somewhat subjectively and might be refined.

Finally, we would like to have tried using sample data from the census, but because of the additional cost, and especially, the likely unavailability of data from the 1980 census in time to be useful, this was given a secondary priority.

REFERENCES

[1] U.S. Department of Commerce, Bureau of the Census. "The Current Population Survey: Design and Methodology," by Robert Hanson. Technical Paper 40. Superintendent of Documents No.: C2.212:40; January 1978.

[2] Diskin, David H. and Baer, Leonard R. "Unit of Stratification of Within-PSU Sampling." Bureau of the Census internal memorandum for the Within-PSU Work Group, December 15, 1980.

[3] Friedman, H. P. and Rubin, J. (1967). "On Some Invariant Criteria for Grouping Data." Journal of the American Statistical Association, 62, 1159-1178.

[4] Kostanich, Donna, et al. "Modification of Friedman-Rubin's Clustering Algorithm for Use in Stratified PPS Sampling." Paper presented at the 1981 American Statistical Association Meetings in Detroit.

[5] Hartley, H.O. and J.N.K. Rao (1962). "Sampling with Unequal Probabilities Without Replacement." Annals of Mathematical Statistics, 33, 350-374.

TABLE 1: PSUs With 1970 Census Population > 500,000

Survey Indices Based on the Current Sort (A), the Preferred Sort, and the Modified Friedman-Rubin Technique (MFR-A)

	PITTSBURGH PA			SEATTLE WA			SAN DIEGO CA			MIAMI FL			DENVER CO			JERSEY CITY NJ		
	SORT		MFR	SORT		MFR	SORT		MFR	SORT		MFR	SORT		MFR	SORT		MFR
	A	F	A	A	D	A	A	F	A	A	F'	A	A	F'	A	A	F	A
INDEX:																		
CPS	0.91	0.50	0.47	0.66	0.59	0.68	0.49	1.09	0.40	0.53	0.37	0.40	0.75	0.57	0.62	0.74	0.71	0.65
NCS	0.81	0.67	0.48	0.81	0.59	0.54	0.50	0.94	0.53	0.69	0.39	0.37	0.58	0.76	0.40	0.59	0.80	0.74
HIS	0.54	0.57	0.49	0.74	0.67	0.54	0.49	0.83	0.43	0.51	0.40	0.37	0.55	0.61	0.39	0.58	0.74	0.61
AHS	0.54	0.67	0.40	0.53	0.58	0.37	0.61	0.70	0.44	0.49	0.53	0.37	0.48	0.62	0.40	0.63	0.95	0.59

TABLE 2: PSUs With 1970 Census Population Between 125,000 and 500,000

Survey Indices Based on the Current Sort (A), the Preferred Sort, and the Modified Friedman-Rubin Technique (MFR-B)

	MONTGOMERY AL			SAVANNAH GA			WACO TX		
	SORT		MFR	SORT		MFR	SORT		MFR
	A	F	B	A	F	B	A	F	B
INDEX:									
CPS	0.54	0.55	0.59	0.96	0.93	0.96	1.00	0.72	0.64
NCS	0.64	0.49	0.54	1.09	0.68	1.09	0.88	0.79	0.80
HIS	0.63	0.55	0.57	0.95	0.66	0.95	0.80	0.74	0.68
AHS	0.71	0.80	0.75	0.63	0.74	0.63	0.71	0.89	0.64

TABLE 3: PSUs With 1970 Census Population < 125,000

Survey Indices Based on the Current Sort (A), the Preferred Sort, and the Modified Friedman-Rubin Technique (MFR-C)

	LYCOMING PA		CLEARFIELD PA		PUTNAM IL		GADSDEN FL			WALTON GA			DONA ANA NM			TALLAPOOSA AL		
	SORT	MFR	SORT	MFR	SORT	MFR	SORT		MFR	SORT		MFR	SORT		MFR	SORT		MFR
	A	C	A	C	A	C	A	F	C	A	F	C	A	F	C	A	F	C
INDEX:																		
CPS	0.61	0.78	0.67	0.82	0.97	0.57	0.81	0.38	1.09	0.64	0.46	0.86	0.83	0.75	1.32	0.91	0.30	0.49
NCS	0.60	0.76	0.64	0.68	0.65	0.54	0.66	0.60	0.77	0.72	0.86	0.81	0.57	0.77	0.67	0.80	0.52	0.47
HIS	0.67	0.65	0.80	0.76	0.70	0.68	0.73	0.76	0.85	0.90	0.81	0.95	0.81	0.70	0.87	0.85	0.37	0.53
AHS	0.64	0.42	0.71	0.67	0.55	0.69	0.53	0.82	0.48	0.81	0.75	0.83	0.72	0.63	0.76	0.63	0.72	0.57