OVERVIEW OF CURRENT POPULATION SURVEY SAMPLE DESIGN

Thomas F. Moore, Paul Bettin, Donna Kostanich, and Gary M. Shapiro
Bureau of the Census

I. INTRODUCTION

This paper briefly describes the sample design of the Current Population Survey (CPS). A companion paper in this session, "Overview of Current Population Survey Operations and Estimation," by Schooley, et al., describes the other aspects of CPS.

The CPS is conducted monthly by the Bureau of the Census to provide estimates of unemployment and other labor force characteristics. It was initially designed primarily for a specific purpose, but is now used for a wide variety of purposes, providing a large amount of detail on the economic status and activities of the population. A monthly survey to measure unemployment began in March 1940 by the Works Projects Administration. The Bureau of the Census took over responsibility for the survey in August 1942. An October 1943 sample revision resulted in about 21,000 assigned interviews per month in 68 primary sampling units (PSU's) comprising 125 counties and independent cities. A listing of the most important survey changes prior to July 1975 follows: July 1945, major revision of CPS schedule; February 1952, introduction of document sensing; September 1953, conversion of tabulations to high-speed electronic equipment; February 1954, major sample redesign to 230 sample PSU's; May 1956, major sample redesign to 330 sample PSU's and expansion to about 35,000 assigned units; June 1957, significant initial use of seasonal adjustment; July 1959, transfer of responsibility for planning, analysis, and publication of the labor force statistics from the Bureau of the Census to the Bureau of Labor Statistics; October 1961, conversion from document sensing to the FOSDIC system; March 1963, selection of sample from census lists for the first time; January 1967, major sample redesign to 449 sample PSU's and expansion to about 52,500 assigned units.

From July 1975 to January 1980, there have been three expansions of the CPS. Earlier expansions had been intended to improve the reliability of national data and were distributed nationwide proportionate to population. The recent expansions, however, have had the purpose of improving reliability for State and substate areas and have been made for selected State and substate areas to meet the needs for better subnational data, particularly for the distribution of Federal funds under the Comprehensive Employment and Training Act (CETA).

The second section of this paper describes the basic CPS sample design, specifically as it was prior to the first State supplementation. Section III describes the first supplementation and the resulting overall sample design; Section IV describes the second supplementation and the resulting overall sample design ; and Section V describes the planned third supplementation. Table 1 below contains a brief description of each of these designs (and of one design which was never implemented), and the timing of their introduction. Section VI discusses future plans briefly.

The Current Population Survey: Design and Methodology, Bureau of the Census Technical Paper 40, 1978, written by Robert H. Hanson, contains substantially more detail on the CPS design than does this paper. However, that publication does not include the supplementations described here in Sections IV and V.

III. DESIGN OF THE CPS NATIONAL SAMPLE

A. Background

In the discussion which follows, the design of the CPS national sample is described as it existed after the major design revisions undertaken to incorporate new information from the 1970 census. These revisions were largely completed as of March 1973.

The specifications characterizing the present version of the CPS national sample are summarized as follows: The CPS is a probability sample. As a consequence of this specification, it is possible to estimate most of the components of the survey error from data produced by the sample. The sample is designed primarily to produce estimates of the major components of the labor force, in general, estimated levels with minimum variance for a fixed cost. The problem of maximizing the reliability for a wide range of other demographic statistics and for other tabulation areas is recognized as an important although secondary requirement. The monthly survey covers the civilian noninstitutional population.

B. Sample Design

The sample design used for the CPS is a multistage stratified sample of the U.S. population and is based to a large extent on the distribution of the population reported in the most recent decennial census. As the time since the last census increases, the efficiency of the sample declines and, although the estimates continue to be unbiased, the sampling errors tend to increase. This increase occurs because the population distribution does not remain constant. By March of 1973, the sample was located in 461 sample areas comprising 924 counties and independent cities, with coverage in every State and the District of Columbia. These areas were selected by dividing the entire area of the United States, consisting of 3,146 counties and independent cities, into 1,931 primary sampling units. With some minor exceptions, a primary sampling unit (PSU) consists of a county or a number of contiguous counties. Each of the standard metropolitan statistical areas (SMSA's) constitutes a separate PSU. Outside SMSA's, counties normally are combined, except where the geographic area of the single county is excessive. More heterogeneous PSU's tend to produce more reliable sample results. Greater heterogeneity can be accomplished by combining a large number of counties and by combining, in particular, different types of counties. However, an important consideration is to keep the PSU sufficiently compact in area so that a small sample, spread throughout, can be efficiently canvassed by one interviewer. This is important for both travel costs and interview time.

Following identification, the 1,931 PSU's were grouped into 376 strata. The largest 156 PSU's are strata by themselves. Other strata were formed by combining PSU's which are similar in such characteristics as SMSA/Non-SMSA status, geographic region, population density, rate of growth in the 1960-1970 decade, proportion nonwhite, principal industry, number of farms, retail sales per capita, and relative number of hotels and motels. A study was made of characteristics that could be used as a basis of stratification, and these were regarded as most appropriate and as having the greatest im-

pact on the data collected in the survey. The more successful this endeavor, the better the reliability of the sample estimates. Except for the large PSU's that form strata by themselves, the strata were established so that their 1970 populations are approximately equal.

The more strata there are, the more sample PSU's there are, and thus the more reliable the data. However, when the strata become too small, the workload per PSU gets so small that the cost per interview becomes much higher. To create the strata for this CPS design the strata that existed for the previous design were taken as the starting point and changes made as required to meet the objectives discussed above. The figure of 376 was not predetermined, but occurred as the result of these changes, and represents an approximate upper bound to the number of strata that can be formed before the cost per interview becomes excessive. The only modifications in PSU definitions that were made were due to changes in SMSA definitions.

The CPS sample consists of two independent national samples called the A and C samples. The full sample is used for the regular monthly series of estimates of labor force and related statistics while the A and C subsamples can be used for other national investigations which do not require the full sample. Furthermore, the use of the A and C samples makes it possible to provide better estimates of some of the survey errors.

Each of the A and C samples was selected in stages. The first stage of selection for the A sample consisted of the 156 largest PSU's and one PSU selected in each of the remaining 220 strata. PSU's within each strata were selected with probability proportionate to size as determined by 1970 census population. The selection of addresses within the sample PSU's is described in subsequent sections of this paper. For the C sample, the 156 largest PSU's were again selected. The remaining 220 strata were grouped into 110 pairs with one stratum selected from each pair at random. From each selected stratum, one PSU was selected in a manner similar to and independent of the A sample selection. This independent selection resulted in 25 PSU's being selected for both the A and C samples and created a total of 461 PSU's for the combined A and C samples.

Within these PSU's, about 55,000 sample addresses were selected. In an average month about 8,000 of these addresses do not contain any occupants or contain occupants who are not eligible for interview. Also for about 2,000 households which are eligible for interview, an interview is not obtained because the occupants could not be found at home after repeated calls, refused the interview, or were unavailable for other reasons. This results in about 45,000 households being interviewed per month.

The initial overall sampling ratio or probability of selection was approximately 1 in 1,300. The sampling ratio is automatically modified over time so that the size of the sample is held relatively constant despite the overall growth of the population. The within-PSU sampling ratio is determined in such a way that the overall sampling rate for each household in the frame is equal.

Two stages of sampling are used in selecting the units in each PSU. The first step is the selection of a sample of census enumeration districts (ED's). These are administrative units designated in the 1970 Census and contain, on the average, about 300 households. ED's are selected systematically from a geographically arranged listing, so that the sample ED's are spread over the entire PSU. The probability of selection of any one ED is proportionate to its 1970 population.

The next step is to select a cluster of approximately four addresses within each designated ED. Whenever possible, the four units, comprising a "segment," are geographically contiguous. The selection of clusters of living quarters is done, wherever possible, from the list of addresses for the ED compiled during the 1970 Census. If the addresses are incomplete or inadequate area sampling methods are used. The address lists are used in about two-thirds of the ED's; these are primarily in urban areas. Area sampling is applied in the remaining ED's.

The list sample is supplemented by a selection of the appropriate proportion of units newly constructed in the PSU since the census date. The sample of new construction units is mainly selected from records of building permits in permit-issuing offices in the areas. Area sampling methods are used in areas where there is no adequate system of building permits. A special procedure of updating parts of the census lists is used to reflect units missed in the census.

In those enumeration districts where area sampling methods are used--mainly rural areas--the ED's are subdivided into small land areas with well-defined boundaries. Generally, these area segments have an expected "size" of eight to twelve housing units or other living quarters. For each subdivided enumeration district, one area segment is designated for the sample, with the probability of selection proportionate to the estimated "size" of the segment. When available advance information indicates that a selected segment contains about four addresses, all units within the segment are included in the sample. In cases where the advance information indicates the "size" of a segment is several times four units, several subsamples are systematically drawn, so as to achieve the equivalent of a four household cluster in each subsample taken from the segment. In both the list and area segments, a set of living quarters is the sample unit and the interview is conducted with the occupants even if they change while the unit is in sample.

Part of the sample is changed each month. A primary reason for rotating the sample is to avoid the poor cooperation which may result from interviewing a constant panel indefinitely. Another reason for replacing households is to reduce the cumulative effect of biases in response which are sometimes observed when the same persons are interviewed indefinitely. For each sample, eight systematic subsamples or rotation groups are identified. Segments in a given rotation group are interviewed for a total of eight months, divided into two equal periods. The segment is in the sample for 4 consecutive months 1 year, out of sample for the following 8 months, and then returns for the same 4 calendar months the next year. In any 1 month, one-eighth of the sample segments are in their first month of enumeration, another 1/8 are in their second month, and so on, with the last 1/8 in for the eighth time. Under this system, 75 percent of the sample is common from month to month and 50 percent from year to year. This procedure provides a substantial amount of month-to-month and year-to-year overlap in the panel, thereby,

reducing discontinuities in the series of data without burdening any specific group of households with an unduly long period of inquiry.

III. STATE ANNUAL SAMPLE SUPPLEMENTATION

A. Purpose

Since 1976, the Census Bureau has been providing estimates of annual averages of labor force data for the 50 States and the District of Columbia. Under the Comprehensive Employment and Training Act of 1973, federal funds are distributed to thousands of prime sponsors based on the relative numbers of unemployed within the prime sponsor areas. The State annual, or D, sample was designed to supplement the national sample in 27 States and the District of Columbia in order to meet the reliability standard specified by the Department of Labor: Provide estimated annual averages of the number of monthly unemployed with a coefficient of variation of 10 percent or less assuming an unemployment rate of 6 percent. (At first, a 5 percent rate was assumed. When the assumed rate was raised to 6 percent, the size of the D sample was reduced and eliminated entirely in three States). The national sample was sufficient to provide reliable estimates in the remaining States.

B. Sample Design

Alternative designs that could be used to supplement the national sample within the States were evaluated on the basis of reliability, cost and the need for consistency in the estimates over all States. Three principles guided the development of the State annual sample design.

1. The D sample, when combined with the existing A/C sample within the State, must meet the specified reliability requirement.

2. The sample design and survey procedures for the D sample must be compatible with the corresponding features of the A/C sample to allow for the use of the interviewer staff, training, sampling, processing and estimation procedures for both samples.

3. The D sample should use the same rotation system as the A/C sample. Experience has shown that panels of households, interviewed at monthly intervals, have reported levels of unemployment in the first month which differ from those reported in succeeding months. Estimates based on a sample with one-eighth of the households interviewed for the first time would, therefore, have a different expected value than a sample with all first month interviews and the expected value of a State estimate would be related to the amount of supplemental sample in the State. Since States have different amounts of supplementation, an inequitable distribution of funds could result.

As in the national sample, a two-stage sample design was used. In the first stage, PSU's were stratified such that strata were defined entirely within the State, unlike the national sample where strata may cross State boundaries. Thus, in unsupplemented States, strata may continue to cross State boundaries. Additional sample PSU's were selected from the new strata.

The stratification and selection of PSU's was accomplished by a computer process. First a preliminary estimate of the number of households needed monthly to meet the annual reliability requirement was made. This estimate and the total number of housing units in the State determined a sampling fraction for a self-weighting sample. The sampling fraction then lead to an ideal stratum size consistent with an interviewer's monthly workload of about 55 households. PSU's with populations larger than this stratum size were made self-representing in the State sample.

All national sample PSU's were retained for the State sample. The stratum represented by such a PSU was that portion of the national stratum within the State unless the size of the new stratum would have resulted in too large a workload size for a single interviewer. In that case, the stratum was subdivided so that the part containing the sample PSU yielded an acceptable workload for one interviewer. The PSU's in the other part, together with all other PSU's not in a stratum containing a sample PSU within the State, were pooled to be grouped into strata.

The computer regressed various 1970 census social characteristics and economic data for the PSU's to derive equations for predicting unemployment rates, sorted the PSU's by these predicted rates and divided the PSU's into strata of approximately equal size. The predicted unemployment rates were usually quite close to the 1970 census unemployment rates.

The components of the sampling error were then estimated. The variance arising from the sampling of PSU's, or between-PSU variance, was calculated by considering all possible samples of PSU's, using 1970 census populations for probabilities and 1960 census unemployment (to allow for deterioration of the stratification over 10 years). The within-PSU component, due to sampling within PSU's, was estimated by computing the variance of a simple random sample of persons 16 years old and over and multiplying by a design effect of 1.4 to reflect the impact of the CPS design and estimation.

If the variance suggested that the coefficient of variation of the annual average unemployment rate would be larger than or substantially less than 10 percent, the process was repeated beginning with the assumption of a new total number of households needed to meet a 10 percent coefficient of variation. See [1] for more details on this stratification procedure.

After the new State strata were defined, a sample PSU was selected from each one with probability proportionate to 1970 population.

All the sample in PSU's already selected for the national sample was retained for the State annual sample. Consequently, the sample in a State was not necessarily self-weighting. If a larger sample was present in some PSU than required, the weights were adjusted accordingly. When the sample present was too small, an additional sample was selected. In a new PSU, the sample to be drawn was determined by the State sampling fraction.

The weight of a unit in the State sample reflected its probability of selection in the combined A/C/D sample.

C. National Estimates

As stated above, the original purpose of the State sample was to provide annual average State estimates. Beginning in January of 1978, the State sample was incorporated into the monthly national estimates as well.

IV. STATE QUARTERLY SAMPLE SUPPLEMENTATION

A. Purpose

To satisfy increasing demand for more frequent estimates for States as reliable as the annual estimates, the State quarterly, or B, sample supplement was designed to provide quarterly averages of labor

force data at the State level. Specifically, the B sample supplement, together with the A/C and D samples, was to produce quarterly State estimates of the number of unemployed with a coefficient of variation of 10 percent assuming a 6 percent unemployment rate. A total of 33 States were supplemented. After the B supplementation, the distribution of States by samples present and type of stratification was as shown in table 2.

B. Sample Design

The requirements of the State quarterly sample were similar to those of the State annual sample: the combined samples should meet the reliability requirement; the sample design and procedures should be compatible with the corresponding features of the A/C and D samples; and the same rotation pattern should be used.

Two methods were considered for increasing the reliability for quarterly State estimates. One sample size was computed assuming the existing stratification would be maintained. If this sample size was less than the current size, the State already met the reliability requirement. If not, a second sample size was computed assuming the State would be restratified in a manner similar to the D sample restratification. Comparing the two sample sizes lead to a decision on which method should be used. Generally, if two workloads, about 100 households, could be saved by restratifying, then this method was used. Only three of the thirty-three States were restratified. In States which had already been restratified for the D sample, no attempted restratification made efficient use of existing sample.

In general, the same procedures were used for both expansions. One difference already mentioned was in the method of supplementation. In the B sample some States which were never restratified for either the B or D samples were supplemented. In the D sample every supplemented State was restratified unless all the PSU's were already in sample. This meant that in the D sample, the current and new sample PSU's could be considered as having been drawn from the State strata. In the eight A/C sample States supplemented but not restratified for the B sample, no new PSU's were selected; the existing sample PSU's were drawn from the national strata which frequently cross State boundaries.

For the D sample, 1970 census figures for civilian labor force, population and housing units were used in the estimates of required sample sizes. Estimates from 1976, the most recent year for which these figures were available, were used in the B sample to account for changes in population and labor force participation rates.

When States were restratified for the B sample, a change was made to the stratification program. Experimentation had shown that there was little difference between stratum definitions based on predicted unemployment and those using actual 1970 census rates. Therefore, the regression routine was deleted from the program and the strata were formed in a simpler and more direct manner using actual unemployment rates.

Research conducted after the D sample supplementation led to the use of an improved estimate of the within PSU component of variance for monthly estimates.

One other difference did not result from a change in procedures, but rather from the period of time covered by the estimates. The monthly within PSU variance was mutliplied by a factor of 0.5 to obtain the variance of a quarterly average; the factor for annual estimates was 0.2. This factor accounts for the interaction of the number of months being averaged and the partial overlap of the monthly samples.

C. Estimates

The B sample was introduced over a period from November 1978 through April 1979. It is planned to use the additional sample in national estimates beginning in January of 1980. State estimation including the B sample should begin in the period July–September 1979.

V. SMSA SAMPLE SUPPLEMENTATION

A. Purpose

The most recent supplemental sample to the CPS, still in the process of being implemented, is the E sample expansion, also known as the SMSA expansion. The purpose of the E sample is to provide more reliable annual average estimates for specified substate areas when combined with the CPS A/C, D and B samples. The reliability standard required that the estimated annual average number of unemployed persons be provided for each of the 67 substate areas shown in table 4 with at most a ten percent coefficient of variation, assuming an unemployment rate of six percent. The 67 areas are contained in 21 States. None of the substate areas in Massachusetts required supplementation; therefore, the E sample is contained in 20 States. The distribution of the States by samples present and type of stratification is given in table 3.

B. Sample Design

The design requirements of the E sample are identical to those of the B sample (see Section IV.B.).

The differences between the E sample design and the designs of the previous two expansions are largely due to the types of areas to be supplemented. The purpose of both the D and B samples was to provide more reliable estimates for all States while the E sample was designed to provide reliable estimates for 67 substate areas--Central Cities, SMSA's, Balances of SMSA's and Balances of States. Some sample PSU's, known as split PSU's, were in two substate areas; this resulted in differing amounts of supplementation for the parts of these PSU's.

Split PSU's also occurred due to the additional requirement that estimates for SMSA's be produced according to 1973 SMSA definitions; SMSA PSU's had previously been defined consistent with 1970 SMSA definitions. Furthermore, new sample areas had to be introduced in order to account for the new definitions. These new areas were in sample with certainty even though their sizes may not have required it.

For SMSA's, central cities and balances of SMSA's, essentially only the method of supplementation which assumes the existing stratification was considered (see Section IV.B.). There exists a slight difference in the application of this method for the SMSA expansion: the new areas added to SMSA's due to 1973 definitions were supplemented even if the specified variance requirements were met without supplementation, in order to reduce possible biases form unsampled portions of SMSA's.

Both methods of supplementation which had been used for the State quarterly expansion were considered for each of the 21 State balances. The only differences were in the application of the restrati-

fication method:
  1. For the SMSA expansion only a substate area was restratified rather than an entire State.
  2. It was more efficient to restratify some State balances even if these areas had previously been restratified for either of the two State expansions.
  Generally, the methods used for the SMSA expansion were more similar to those used for the State quarterly expansion than to the State annual expansion.
  Estimates from 1977 of civilian labor force, population and housing units were used in determining the amount of supplementation needed, just as 1976 estimates were used for the State quarterly expansion.
  Restratification of State balances for the E sample was based on the actual percent of the population unemployed according to the 1970 census as it was for the B sample.
  The monthly within PSU variance was multiplied by a factor of 0.24 to obtain the variance of an annual average. The 0.24 factor was computed using newly available data [3] , and thus is slightly higher than the factor used in the D sample.
  As had been done for the two State expansions, current sample PSU's were retained for the E sample and their weights were adjusted if more sample was present than was needed for a self-weighting sample.
  C. Estimates
    The SMSA expansion is scheduled to be introduced in September and October of 1979. It has not yet been established if and when the published national estimates would begin including the E sample.

VI. FUTURE PLANS
    We anticipate the next modification to the CPS to occur in the 1983-1984 time period. Another expansion at that time is likely, the objectives for which are discussed in the paper by Robert C. Stein in this session "Implementing the Recommendations of the National Commission of Employment and Unemployment Statistics." In addition to an expansion will be a major redesign of the survey. Early in each decade the Census Bureau has revised the strata definitions, reselected the sample PSU's, and changed from sampling using the previously conducted Census materials to the most recent Census materials. In addition to performing these updating operations for 1983-1984, the Census Bureau has begun an extensive research program to reconsider almost every aspect of the survey design. Early redesign planning is discussed in the paper in this session, "Planning for the Current Population Survey Redesign Effort," by Thomas Plewes and one aspect of the research is discussed in the paper by Charles D. Cowan, et al., in this session, "A Test of Data Collection Methodologies: The Methods Test." Among other things, consideration is being given to major changes in the way PSU's are defined, special stratification and sampling techniques for efficiently improving the reliability for minority statistics are being investigated, different rotation patterns are being compared, and the possibility of sampling from unemployment insurance rolls (in addition to sampling from the general population) is being considered. Although in most of these areas we may decide that the present procedures are best, there will likely be at least one or two major design changes in 1983-1984.

Table 1. Summaries of Each Supplementation to Current Population Survey, 1975-1979

Supplementation	Sample Designation	# of New Sample PSU's in Supplementation	# of PSU's in National Design Resulting	Supplementation Objective	Time Period of 1st Interview	Time Introduced into National Tabs	# of Assigned Units in Supplementation	# of Interviewed Units in Supplementation	# of Assigned Units in Full National Design	# of Interviewed Units in Full National Design
State Annual	D	151	614	10% CV Annually for States	7/75-9/75	1/78	10,500	8,500	65,000	53,000
Substate¹ Annual		215²	829	10% CV Annually for SMSA's and other Substate Areas	(N.A)	(N.A.)	12,500	(N.A.)	78,000	(N.A.)
State Quarterly	B	15	629	10% CV Quarterly for States	11/78-4/79	1/80	10,000	7,900	75,000	61,000
SMSA	E	14²	643	10% CV Annually for SMSA's and other Substate Areas	9/79-10/79	3	6,500	5,200	81,500	66,000

¹This supplementation was planned for 8/77-10/77 but never actually implemented.
²New areas were also added to seven existing PSU's.
³Unknown at this time.

Table 2:	Type of Stratification	
	State	National
Samples AC	-	15
AC,D	3	-
AC,B	3	8
AC,D,B	22*	-

* Includes two States where all the PSU's were in the national sample.

Table 3:	Type of Stratification	
	State	National
Samples: AC	-	2
AC,D	1	-
AC,B	2	5
AC,E	-	13
AC,D,B	21*	-
AC,D,E	2	-
AC,B,E	2	2
AC,D,B,E	1	-

* Includes two States where all the PSU's were in the national sample.

Table 4: Substate Areas in the E Sample Expansion
State and Substate Area (Supplementation Required)

California
Los Angeles SMSA (No), San Francisco SMSA (No), Anaheim SMSA (No), San Diego SMSA (Yes), San Bernardino SMSA (Yes), San Jose SMSA (Yes), Balance of State (No)

Colorado
Denver SMSA (Yes)[3], Balance of State[2](Yes)

District of Columbia
D.C. Central City (No)

Florida
Miami SMSA (Yes), Balance of State (No)

Georgia
Atlanta SMSA (Yes)[3], Balance of State (No)

Illinois
Chicago Central City (No), Balance of Chicago SMSA (No), Balance of St. Louis SMSA (Ill. pt.) (Yes), Balance of State (No)

Indiana
Indianapolis SMSA (Yes), Balance of State (No)

Indiana & Kentucky[1]
Cincinnati SMSA (Ind. & Ky. pts.) (Yes)

Kansas
Kansas City SMSA (Kn. pt.) (Yes), Balance of State (No)

Kentucky
Balance of State (No)

Maryland
Baltimore Central City (Yes), Balance of Baltimore SMSA (Yes), Balance of D.C. SMSA (Md. pt.) (Yes), Balance of State[2] (Yes)

Massachusetts
Boston SMSA (No), Balance of State (No)

Michigan
Detroit Central City (Yes), Balance of Detroit SMSA (Yes)[3], Balance of State (No)

Minnesota
Minneapolis SMSA (Yes)[3], Balance of State[2](Yes)

Missouri
St. Louis Central City (Yes), Balance of St. Louis SMSA (Mo. pt.)(Yes), Kansas City SMSA (Mo. pt.)(Yes), Balance of State (Yes)

New Jersey
Newark SMSA (No), Balance of Philadelphia SMSA (N.J. pt.)(Yes), Balance of State (No)

New York
New York Central City (No), Balance of New York LMA (Yes), Nassau-Suffolk SMSA (No), Buffalo SMSA (Yes), Balance of State (No)

Ohio
Cleveland Central City (Yes), Balance of Cleveland SMSA (Yes), Cincinnati SMSA (Ohio pt.)(Yes), Balance of State (No)

Pennsylvania
Philadelphia Central City (Yes), Balance of Philadelphia SMSA (Pa. pt.)(No), Pittsburgh SMSA (No), Balance of State (No)

Texas
Houston Central City (Yes), Balance of Houston SMSA (Yes), Dallas Central City (Yes), Balance of Dallas-Ft. Worth SMSA (Yes)[3], Balance of State (No)

Virginia
Balance of D.C. SMSA (Va.pt.) (Yes), Balance of State (No)

Washington State
Seattle SMSA (No), Balance of State[2] (Yes)

Wisconsin
Milwaukee Central City (Yes), Balance of Milwaukee SMSA (Yes), Balance of State (No)

[1]The Indiana and Kentucky portion of the Cincinnati SMSA is the only Substate Area crossing state boundaries.
[2]These Substate areas were restratified for the E-sample.
[3]Only the new areas being added to SMSA's due to the 1973 SMSA definitional changes require supplementation.

REFERENCES
[1] Dippo, Cathryn, "Expansion of CPS to Provide Reliable State Estimates of Unemployment." Proceedings of the Social Statistics Section, Annual Meeting of the American Statistical Association, Atlanta, Ga., Aug. 25-28, 1975. Washington, D.C., American Statistical Association, 1976, pp. 387-394.

[2] U.S. Department of Commerce, Bureau of the Census. The Current Population Survey: Design and Methodology, by Robert H. Hanson. Technical Paper 40. Washington, D.C., Government Printing Office, 1978.

[3] Train, George; Cahoon, Larry; Makens, Paul; "The Current Population Survey Variances, Inter-Relationships, and Design Effects." Proceedings of the Section on Survey Research Methods, Annual Meeting of the American Statistical Association, San Diego, CA., Aug. 14-17, 1978. Washington, D.C.: American Statistical Association, 1978, pp.443-448.