

June 24, 2002

MEMORANDUM FOR: CHESTER E. BOWIE, Chief
Demographic Surveys Division
Bureau of the Census

FROM : KENNETH V. DALTON
Associate Commissioner
Prices and Living Conditions
Bureau of Labor Statistics

SUBJECT : Specifications for the Selection of
CE/CPI Samples in PSUs Based
on the 2000 Census

I. Introduction

The overall design for selecting the Consumer Expenditure Survey (CE) sample from the 2000 Census frame remains unchanged. The eligible population continues to be all civilian noninstitutional persons, including people residing in boarding houses, housing facilities for students and workers, staff units, mobile home parks, permanent-type living quarters in hotels and motels, and eligible group quarters. Appendix I shows the basic survey design for this CE population. Because Core Based Statistical Area (CBSA) definitions will not be final until after CE sample selection has begun, the Bureau of the Census (BOC) will draw a sample in all Primary Sampling Units (PSU) based on the geographic boundaries specified in the data files accompanying this memo and also listed in Appendix II.

PSUs are divided into four size classes: A, X, Y, and Z. The A, X, and Y PSUs are constructed based on preliminary CBSA definitions. The A PSUs were constructed by combining counties using 1990 Metropolitan Statistical Area definitions with adjustment based on preliminary CBSA definitions. The X PSUs are metropolitan CBSAs not contained in any A PSU. The Y PSUs are micropolitan CBSAs not contained in any A PSU. The Z PSUs, which are used only for CE, are defined by the Bureau of Labor Statistics (BLS) in terms of counties not in any CBSA.

Requirements for the Telephone Point of Purchase Survey for the PSU sample based on the 2000 Census will be sent later and are not included in this memorandum.

For the Diary and Interview survey, unclustered systematic samples will be selected within each PSU listed in Appendix II. The reference period for the Diary Survey will continue to be two weeks, and the Interview Survey will continue to have five quarterly interviews. The phase-in period for the 2000-based sample will begin in November 2004. For each survey, sample units will be grouped into 24 representative “sample designations” of equal size. Enough of these sample designations should be assigned sample dates to cover the period 2005 through 2018, with the remaining sample designations set aside as “reserve” sample. The reserve sample allows for sample expansion, methods development, and research.

The Consumer Price Index (CPI) Permit New Construction Housing Sample should be selected to yield a total of 120 permit new construction units per month across all selected CPI PSUs. The housing units should be selected in such a way as to yield an expected four housing units per hit.

The following details are not meant to be all-inclusive, but instead highlight areas of importance in the sample selection process.

II. Self-representing PSUs and the Basic Survey Design

The BLS has selected 28 A (self-representing) PSUs. With the exception of Honolulu and Anchorage, each self-representing PSU has a population greater than 2 million people based on 2000 Census data and preliminary CBSA definitions.

The basic survey design appears in Appendix I. There will be four size classes, A, X, Y, and Z. The X and Y PSUs are metropolitan and micropolitan CBSAs, respectively, that are not included in any A PSU and are not self-representing. The BLS divided the non-CBSA area of the U.S. into the Z PSUs that were formed from adjacent counties with a total population intended to be large enough to support CE sample requirements.

A list of the selected PSUs appears in Appendix II. Samples for CE/CPI should be selected within the geographic boundaries of each PSU shown in Appendix II. A letter and three-digit number identify each selected PSU. The number is even for all nonself-representing PSUs. The Z PSUs are to be used only for CE. Further details are in **Section V, Contents of the Data Files.**

The selection of this PSU sample was done using a stratified sample design with keyfitzing and controlled selection. All nonself-representing PSUs (X, Y and Z PSUs) were stratified. The strata were created in the twelve region-size classes (4 Census regions times 3 size classes (X, Y and Z)) using a methodology similar to that employed for the 1987 and 1998 Revisions. The probabilities of selection for the X, Y and Z PSUs were Keyfitted using a modified version of the 1998 Revision PSU selection Keyfitzing program. One PSU was selected from each stratum using a newly developed three-way controlled selection program with controls on states, strata and overlap. Detailed documentation of the PSU selection process is being written in a paper for the 2002 Joint Statistical Meetings.

III. Diary and Interview Surveys

The basic procedures for selecting the Diary and Interview samples remain unchanged. Important points are summarized below.

- The within-PSU sort order for the systematic sampling of units is left to the discretion of the BOC with the understanding that its research and conclusions will be clearly documented and provided to the BLS.
- Sample selection procedures for the Unit, Area, Permit, and Group Quarters sampling frames should be designed to avoid any clustering.
- The number of interviewed households should remain constant over time. As the sample size increases in the Permit frame, adjustments should be made to the other sampling frames to keep the overall number of interviewed households constant.
- The assignment of placement day for the Diary should be done in such a way as to ensure a representative distribution of sample units within random groups, halvesamples, months, and days of the week. Any cross-classification by these variables should yield a representative sample. Placement days should be strictly adhered to by the field.

Allocating the Nationwide Sample to Individual PSUs

The CE budget allows for 7,700 interviewed households per year in the Diary survey, and 7,700 interviewed households per quarter in the Interview survey (interviews 2-5 only) at the national level.¹ For each survey the nationwide sample of 7,700 interviewed households will be allocated to individual PSUs in two different ways depending on PSU size: 400 interviewed households will be allocated to Z PSUs, and the remaining 7,300 interviewed households will be allocated to A, X, and Y PSUs. We will limit the target number of interviewed households for Z PSUs to 400 to allow more of the sample to be allocated to A, X, and Y PSUs, which are used by the CPI. The nationwide sample should be allocated to individual PSUs as specified below.

The Z PSUs will have 400 interviewed households per year in the Diary survey, and 400 interviewed households per quarter in the Interview survey (interviews 2-5 only). These 400 interviewed households should be allocated to individual Z PSUs directly proportional to the population that each PSU represents. The populations represented by Z PSUs are listed in Appendix III, Table 1.

¹ This is expected to yield 16,170 usable interviews per year in the Diary survey, and 32,340 usable interviews per year in the Interview survey. These estimates come from assuming that 7,700 households contain 8,085 ($=7,700 \times 1.05$) consumer units. Hence there will be approximately 16,170 usable interviews per year in the Diary survey (2 weekly diaries per consumer unit), and 32,340 usable interviews per year in the Interview survey (8,085 consumer units per quarter times 4 quarters per year).

The X and Y PSUs are grouped into 8 region-size classes. They are created by cross-classifying the four regions of the country (Northeast, Midwest, South, West) with the two size classes (X, Y). For allocation these 8 region-size classes are treated just like other self-representing PSUs (A PSUs). Thus for the sample allocation process we define a *CPI Index Area* to be any one of the 28 A PSUs, or any one of the 8 X/Y region-size classes, for a total of 36 CPI Index Areas.

In the 36 CPI Index Areas usable interviews will be collected from 7,300 households per year in the Diary survey, and 7,300 households per quarter in the Interview survey (interviews 2-5 only). The 7,300 interviewed households should be allocated to the 36 CPI Index Areas as close to population proportionality as possible, subject to the constraint that each CPI Index Area contain at least 80 interviewed households per year in the Diary survey, and at least 80 interviewed households per quarter in the Interview survey. The number of interviewed households allocated to the X/Y region-size classes should then be allocated to the individual CE PSUs within the region-size classes directly proportional to the population represented by the PSU. The populations represented by X and Y PSUs are listed in Appendix III, Tables 2 and 3, respectively.

The Phase-In and Phase-Out Period

For the Diary survey, data collection for the 1990-based sample will continue through January 2005. The actual placement of the Diary form and subsequent 14-day reference period is contingent on the “earliest placement date” assigned to the sample case. The last possible “earliest placement date” assigned for the 1990-based Diary sample is December 31, 2004.

For the Diary Survey in continuing and new areas, the 2000-based sample will be introduced in January 2005.

For the Interview Survey, the phase-in period for the 2000-based redesign sample will begin in November 2004. Phase-in/out will occur over a 5-month period (November 2004 – March 2005) in the dropped and new areas, and will occur over a 12-month period (November 2004 – October 2005) in the continuing areas. During the phase-in/out period, the following will occur:

Dropped areas

- Interviews using the 1990-based sample will be collected through March 2005.
- The incoming rotation group (using the 1990-based sample) for the first quarter of 2005 will not be introduced in dropped areas.

Continuing areas

- The final interviews using the 1990-based sample will occur in October 2005.
- The incoming rotation group (using the 2000-based sample) will be introduced beginning in November 2004.

New areas

- The sample in new areas will be introduced beginning in November 2004.
- The incoming rotation groups (using the 2000-based sample) will be introduced beginning in November 2004.
- Sample introduced in January 2005 (using the 2000-based sample) will include only three of the four rotation groups originally introduced in November 2004 because no phase-in interviews will occur in October 2004.

Since phase-in begins in the second month of the last quarter of 2004, only two-thirds of the sample in new and continuing areas will be introduced in that period. After phase-in is completed in October 2005, the rest of the 2000-based sample will be introduced using the ongoing rotation pattern.

IV. CPI Permit New Construction Housing Sample

The sample size for the CPI Permit New Construction Housing Sample is 120 permits per month spread throughout all CPI PSUs. The BOC will use data from the past to project future housing unit permit construction in permit construction areas within the selected PSUs. Several years of data will be used for the the BOC's predictions. The allocation of the sample to PSUs will be proportional to the predicted new construction permit housing units in each PSU.

The BOC will provide the BLS with the sample size and sampling interval in each PSU and the method of sample allocation. The sample should be selected with an expected four housing units per hit from a frame consisting of permits for new housing units constructed since the 2000 Census. In order to help the BLS field staff locate the sample, the BOC should provide the permit office location and permit number of each sample unit, the unit's address including zip code, and the name of the complex in which the unit lies. Urban/Rural status of new construction permits is no longer needed for the CPI Permit New Construction Housing Sample and should not be provided. If no address is sent, then the BOC should provide the BLS with detailed street sketch maps. In addition, the BOC should provide the BLS with an annually updated electronic list of Permit Offices, which includes an accurate telephone number, and the name of the person in charge.

V. Contents of the Data Files

The data files contain records defining the selected PSUs in terms of state-county combinations and the stratum populations and probabilities of selection. The layout of the data files is contained in Appendix IV.

The data file `censout2000cpi.sdf` contains all of the PSUs used for the CPI-U. These PSUs include all of the selected A, X, and Y PSUs. These PSUs are also used for CE. The stratum population is the sum of the population of all PSUs in the same stratum as the selected PSU. The final column is the unconditional probability of selection, which is defined as the PSU population divided by the stratum population.

The data file censout2000ce.sdf contains all Z PSUs, which are used only by CE. Some of the selected Z PSUs had too little population to accommodate the CE Survey. Each of these PSUs was augmented with counties in the same stratum to bring them up to sufficient size. On the data file censout2000ce.sdf, these added counties are not distinguished from the originally selected PSU's counties which contained insufficient population. The column for stratum population contains the total stratum population for all records associated with each Z stratum. The unconditional probability is defined as the augmented Z PSU population divided by the stratum population.

Attachments

OPLC/SMD

WJOHNSON/062402/A0202.6/rm. 3655 PSB/Tele. 691-6921

cc: