Current Population Survey

Technical Paper 63RV

Design and Methodology

TP63RV

Acknowledgments

Current Population Survey

Technical Paper 63RV

Design and Methodology

Foreword

The Current Population Survey (CPS) is one of the oldest, largest, and most well-recognized surveys in the United States. It is immensely important, providing information on many of the things that define us as individuals and as a society—our work, our earnings, our education. It is also immensely complex. Staff of the Census Bureau and the Bureau of Labor Statistics have attempted, in this publication, to provide data users with a thorough description of the design and methodology used in the CPS. The preparation of this technical paper was a major undertaking, spanning several years and involving dozens of statisticians, economists, and others from the two agencies.

This paper is the first major update of CPS documentation in more than two decades, and, while the basic approach to collecting labor force and other data through the CPS has remained intact over the intervening years, much has changed. In particular, a redesigned CPS was introduced in January 1994, centered around the survey's first use of a computerized survey instrument by field interviewers. The questionnaire itself was rewritten to better communicate CPS concepts to the respondent, and to take advantage of computerization.

This document describes the design and methodology that existed for the CPS as of December 1995. Some of the appendices cover updates that have been made to the survey since then.

Users of CPS data should have access to up-to-date information about the survey's methodology. The advent of the Internet allows us to provide updates to the material contained in this report on a more timely basis. Please visit our CPS web site at http://www.bls.census.gov/cps, where updated survey information will be made available. Also, we welcome comments from users about the value of this document and ways that it could be improved.

Kenneth Prewitt
Director
U.S. Census Bureau

March 2000

Katharine G. Abraham
Commissioner
Bureau of Labor Statistics

March 2000

Contents

Chapter 8. Transmitting the Interview Results

Chapter 9. Data Preparation

Chapter 10. Estimation Procedures for Labor Force Data

Chapter 11. Current Population Survey Supplemental Inquiries

Chapter 12. Data Products From the Current Population Survey

Chapter 13. Overview of Data Quality Concepts

Chapter 14. Estimation of Variance

Chapter 15. Sources and Controls on Nonsampling Error

Chapter 16. Quality Indicators of Nonsampling Errors

Figures

Illustrations

Illustrations—Con.

Tables

Tables—Con.

Appendix A. Maximizing Primary Sampling Unit (PSU) Overlap

Appendix B. Sample Preparation Materials

Appendix C. Maintaining the Desired Sample Size

Appendix D. Derivation of Independent Population Controls

Summary of Changes

(Changes made to Current Population Survey Technical Paper 63 to Produce Technical Paper 63RV, March 2002)

Chapter 1.   Background
- page 1-1, left column, fourth paragraph:[1] added a footnote about the sample size increase detailed in Appendix J.

Chapter 2.   History of the Current Population Survey
- page 2-3, third paragraph of *December 1971-March 1973:* changed **1992** to **1972.**[2]
- page 2-4, *April 1984:* changed **1995** to **1985**.
- page 2-5: added a section ***January 1998*** describing a two-step composite estimation method.
- page 2-5: added a section ***July 2001*** describing the SCHIP.

Chapter 3.   Design of the Current Population Survey Sample
- page 3-1, chapter heading: added a reference to Appendix J.
- page 3-1, after second paragraph of *INTRODUCTION:* added text on the SCHIP.
- page 3-11, Table 3-3a: corrected sorts.

Chapter 9.   Data Preparation
- page 9-1, first paragraph of *INDUSTRY AND OCCUPATION (I&O) CODING:* added a footnote about the increase of cases for coding because of the SCHIP.

Chapter 10.   Estimation Procedures for Labor Force Data
- page 10-1, last paragraph of *INTRODUCTION*: added text about a new compositing procedure detailed in Appendix I.
- page 10-5, Table 10-1: changed **1999** in heading to **1990.**
- page 10-6, Table 10-3: inserted missing age category **55-59.**
- page 10-11, second paragraph of *Estimates for States*: added a footnote that all states are now based on a model.

Chapter 11.   Current Population Survey Supplemental Inquiries
- pages 11-1–11-3: made clarifications and revisions to improve readability.
- pages 11-3–11-8, *Annual Demographic Supplement (March Supplement)*: major revisions due to more current methodologies and the inclusion of the SCHIP.

Chapter 14.   Estimation of Variance
- page 14-5, Table 14-1: added a footnote that refers to Appendix H.

Chapter 16.   Quality Indicators of Nonsampling Errors
- page 16-1, chapter heading: added a reference to a BLS internet site.

Appendix E.   State Model-Based Labor Force Estimation
- page E-1, Table E-1: added a footnote about the SCHIP.

Appendix F.   Organization and Training of the Data Collection Staff
- pages F-1 – F3, all sections: updated several of the numbers/percentages with 2001 data.
- page F-3, Figure F-2: changed to Table F-1 and updated it.
- page F-3, fourth paragraph of *FIELD REPRESENTATIVE PERFORMANCE GUIDELINES:* changed **CARMIN** to **CARMN**.

---

[1]Page number, column, and paragraph number are those from TP63. They indicate where the change begins.
[2]Boldface indicates a change or addition from TP63 to TP63RV.

Appendix G. **Reinterview: Design and Methodology**
- pages G-1–G-3, all sections: major revisions/deletions due to more current methodologies.

Appendix H. **Sample Design Changes of the Current Population Survey: January 1996**
- page H-1, chapter heading: added a reference to Appendix J.

Appendix J. **Changes to the Current Population Survey Sample in July 2001**
- This is an entirely new appendix, focusing on the changes that are collectively known as the SCHIP sample expansion.

Chapter 1.
Background

The Current Population Survey (CPS), sponsored jointly by the U.S. Census Bureau and the U.S. Bureau of Labor Statistics (BLS), is the Nation's primary source of labor force statistics for the entire population. The CPS is the source of numerous high-profile economic statistics including the Nation's unemployment rate and provides data on a wide range of issues relating to employment and earnings. The CPS also collects extensive demographic data which complement and enhance our understanding of labor market conditions in the Nation overall, among many different population groups, the various states, and even substate areas.

Although the labor market information is central to the CPS, the survey provides a wealth of other social and economic data that are widely used by social scientists in both the public and private sectors. In addition, because of its long history, the CPS has been a model for other household surveys, both in the United States and in many other countries.

Thus, the CPS is a source of information for both social science research and the study of survey methodology. This report aims to provide all users of the CPS with a comprehensive guide to the survey. The report focuses on labor force data because the timely and accurate collection of those data remains the principal purpose of the survey.

The CPS is administered by the Census Bureau using a scientifically selected sample of some 50,000 occupied households.[1] The fieldwork is conducted during the calendar week that includes the 19th of the month. The questions refer to activities during the prior week; that is, the week that includes the 12th of the month.[2] Households from all 50 states and the District of Columbia are in the survey for 4 consecutive months, out for 8, and then return for another 4 months before leaving the sample permanently. This design ensures a high degree of continuity from 1 month to the next (as well as over the year). The 4-8-4 sampling scheme has the added benefit of allowing for the constant replenishment of the sample without excessive response burden.

To be eligible to participate in the CPS, individuals must be 15 years of age or over and not in the Armed Forces. Persons in institutions, such as prisons, long-term care hospitals, and nursing homes are, by definition, ineligible to

be interviewed in the CPS. In general, the BLS publishes labor force data only for persons age 16 and over, since those under 16 are substantially limited in their labor market activities by compulsory schooling and child labor laws. No upper age limit is used, and full-time students are treated the same as nonstudents. One person generally responds for all eligible members of the household. The person who responds is called the "reference person" and usually is the person who either owns or rents the housing unit. If the reference person is not knowledgeable about the employment status of the others in the household, attempts are made to contact those individuals directly.

Within 2 weeks of the completion of these interviews, the BLS releases the major results of the survey. Also included in BLS's analysis of labor market conditions are data from a survey of nearly 400,000 employers (the Current Employment Statistics (CES) survey, conducted concurrently with the CPS). These two surveys are complementary in many ways. The CPS focuses on the labor force status (employed, unemployed, not in labor force) of the working-age population and the demographic characteristics of workers and nonworkers. The CES focuses on aggregate estimates of employment, hours, and earnings for several hundred industries that would be impossible to obtain with the same precision through a household survey. The CPS reports on individuals not covered in the CES, such as the self employed, agricultural workers, and unpaid workers in a family business. Information also is collected in the CPS about persons who are not working.

In addition to the regular labor force questions, the CPS often includes supplemental questions on subjects of interest to labor market analysts. These include annual work activity and income, veteran status, school enrollment, contingent employment, worker displacement, and job tenure, among other topics. Because of the survey's large sample size and broad population coverage, a wide range of sponsors use CPS supplements to collect data on topics as diverse as expectation of family size, tobacco use, computer use, and voting patterns. The supplements are described in greater detail in Chapter 11.

The CPS questionnaire is a completely computerized document that is administered by Census Bureau field

---

[1]Beginning with July 2001, the sample size increased to 60,000 occupied households. (See Appendix J for details.)

[2]In the month of December, the survey is often conducted 1 week earlier to avoid conflicting with the holiday season.

representatives across the country through both personal and telephone interviews. Additional telephone interviewing also is conducted from the Census Bureau's three centralized collection facilities in Hagerstown, Maryland; Jeffersonville, Indiana; and Tucson, Arizona.

The labor force concepts and definitions used in the CPS have undergone only slight modification since the survey's inception in 1940. Those concepts and definitions are discussed in Chapter 5.

Chapter 2.
History of the Current Population Survey

INTRODUCTION

The Current Population Survey (CPS) has its origin in a program established to provide direct measurement of unemployment each month on a sample basis. There were several earlier attempts to estimate the number of unemployed using various devices ranging from guesses to enumerative counts. The problem of measuring unemployment became especially acute during the Economic Depression of the 1930s.

The Enumerative Check Census, taken as part of the 1937 unemployment registration, was the first attempt to estimate unemployment on a nationwide basis using probability sampling. During the latter half of the 1930s, the Work Projects Administration (WPA) developed techniques for measuring unemployment, first on a local area basis and later on a national basis. This research combined with the experience from the Enumerative Check Census led to the Sample Survey of Unemployment which was started in March 1940 as a monthly activity by the WPA.

MAJOR CHANGES IN THE SURVEY: A CHRONOLOGY

In August 1942, responsibility for the Sample Survey of Unemployment was transferred to the Bureau of the Census, and in October 1943, the sample was thoroughly revised. At that time, the use of probability sampling was expanded to cover the entire sample, and new sampling theory and principles were developed and applied to increase the efficiency of the design. The households in the revised sample were in 68 Primary Sampling Units (PSUs) (see Chapter 3), comprising 125 counties and independent cities. By 1945, about 25,000 housing units were designated for the sample, of which about 21,000 contained interviewed households.

One of the most important changes in the CPS sample design took place in 1954 when, for the same total budget, the number of PSUs was expanded from 68 to 230, without any change in the number of sample households. The redesign resulted in a more efficient system of field organization and supervision and provided more information per unit of cost. Thus the accuracy of published statistics improved as did the reliability of some regional as well as national estimates.

Since the mid-1950s, the CPS's sample has undergone major revision after every decennial census. The following list chronicles the important modifications to the CPS starting in the mid-1940s:

- **July 1945.** The CPS questionnaire was revised. The revision consisted of the introduction of four basic employment status questions. Methodological studies showed that the previous questionnaire produced results that misclassified large numbers of part-time and intermittent workers, particularly unpaid family workers. These groups were erroneously reported as not active in the labor force.

- **August 1947.** The selection method was revised. The method of selecting sample units within a sample area was changed so that each unit selected would have the same basic weight. This change simplified tabulations and estimation procedures.

- **July 1949.** Previously excluded dwelling places were now covered. The sample was extended to cover special dwelling places—hotels, motels, trailer camps, etc. This led to improvements in the statistics, (i.e., reduced bias) since residents of these places often have characteristics different from the rest of the population.

- **February 1952.** Document sensing procedures were introduced in the survey process. The CPS questionnaire was printed on a document-sensing card. In this procedure, responses were recorded by drawing a line through the oval representing the correct answer using an electrographic lead pencil. Punch cards were automatically prepared from the questionnaire by document-sensing equipment.

- **January 1953.** Ratio estimates now used data from the 1950 population census. Starting in January 1953, population data from the 1950 census were introduced into the CPS estimation procedure. Prior to that date, the ratio estimates had been based on 1940 census relationships for the first-stage ratio estimate, and 1940 population data were used to adjust for births, deaths, etc., for the second-stage ratio estimate. In September 1953, a question on "color" was added and the question on "veteran status" was deleted in the second-stage ratio estimate. This change made it feasible to publish separate, absolute numbers for persons by race; whereas, only the percentage of distributions had previously been possible.

- **July 1953.** The 4-8-4 rotation system was introduced. This sample rotation system was adopted to improve measurement over time. In this system households are interviewed for 4 consecutive months 1 year, leave the

sample for 8 months, and return for the same period of 4 months the following year. In the previous system, households were interviewed for 6 months and then replaced. The 4-8-4 system provides some year-to-year overlap, thus improving estimate of change on both a month-to-month and year-to-year basis.

- **September 1953.** High speed electronic equipment was introduced for tabulations. The introduction of electronic calculation greatly increased timeliness and led to other improvements in estimation methods. Other benefits included the substantial expansion of the scope and content of the tabulations and the computation of sampling variability. The shift to modern computers was made in 1959. Keeping abreast of modern computing has proved a continuous process, and to this day, the Census Bureau is still updating and replacing its computer environment.

- **February 1954.** The number of PSUs was expanded to 230. The number of PSUs was increased from 68 to 230 while retaining the overall sample size of 25,000 designated housing units. The 230 PSUs consisted of 453 counties and independent cities. At the same time, a substantially improved estimation procedure (See Chapter 10, Composite Estimation) was introduced.

  Composite estimation took advantage of the large overlap in the sample from month-to-month. These two changes improved the reliability of most of the major statistics by a magnitude that could otherwise be achieved only by doubling the sample size.

- **May 1955.** Monthly questions on part-time workers were added. Monthly questions exploring the reasons for part-time work were added to the standard set of employment status items. In the past, this information had been collected quarterly or less frequently and was found to be valuable in studying labor market trends.

- **July 1955.** Survey week was moved. The CPS survey week was moved to the calendar week containing the 12th day of the month to align the CPS time reference with that of other employment statistics. Previously, the survey week had been the calendar week containing the 8th day of the month.

- **May 1956.** The number of PSUs was expanded to 330. The number of PSUs was expanded from 230 to 330. The overall sample size also increased by roughly two-thirds to a total of about 40,000 households units (about 35,000 occupied units). The expanded sample covered 638 counties and independent cities.

  All of the former 230 PSUs were also included in the expanded sample.

  The expansion increased the reliability of the major statistics by around 20 percent and made it possible to publish more detailed statistics.

- **January 1957.** Employment status definition was changed. Two relatively small groups of persons, both formerly classified as employed "with a job but not at work," were assigned to new classifications. The reassigned groups were (1) persons on layoff with definite instructions to return to work within 30 days of the layoff date and (2) persons waiting to start new wage and salary jobs within 30 days of the interview. Most of the persons in these two groups were shifted to the unemployed classification. The only exception was the small subgroup in school during the survey week who were waiting to start new jobs; these persons were transferred to "not in labor force." This change in definition did not affect the basic question or the enumeration procedures.

- **June 1957.** Seasonal adjustment was introduced. Some seasonally adjusted unemployment data were introduced early in 1955. An extension of the data—using more refined seasonal adjustment methods programmed on electronic computers—was introduced in July 1957. The new data included a seasonally adjusted rate of unemployment and trends of seasonally adjusted total employment and unemployment. Significant improvements in methodology emerged from research conducted at the Bureau of Labor Statistics and the Census Bureau in the ensuing years.

- **July 1959.** Responsibility for CPS was moved between agencies. Responsibility for the planning, analysis, and publication of the labor force statistics from the CPS was transferred to the BLS as part of a large exchange of statistical functions between the Commerce and Labor Departments. The Census Bureau continued to have (and still has) responsibility for the collection and computer processing of these statistics, for maintenance of the CPS sample, and for related methodological research. Interagency review of CPS policy and technical issues continues under the aegis of the Statistical Policy Division, Office of Management and Budget.

- **January 1960.** Alaska and Hawaii were added to the population estimates and the CPS sample. Upon achieving statehood, Alaska and Hawaii were included in the independent population estimates and in the sample survey. This increased the number of sample PSUs from 330 to 333. The addition of these two states affected the comparability of population and labor force data with previous years. Another result was in an increase of about 500,000 in the noninstitutional population of working age and about 300,000 in the labor force, four-fifths of this in nonagricultural employment. The levels of other labor force categories were not appreciably changed.

- **October 1961.** Conversion to the Film Optical Sensing Device for Input to the Computer (FOSDIC) system. The CPS questionnaire was converted to the FOSDIC type used by the 1960 census. Entries were made by filling in small circles with an ordinary lead pencil. The questionnaires were photographed to microfilm. The microfilms were then scanned by a reading device which transferred the information directly to computer tape. This system permitted a larger form and a more flexible arrangement

of items than the previous document-sensing procedure and did not require the preparation of punch cards. This data entry system was used through December 1993.

- **January 1963.** New descriptive information was made available. In response to recommendations of a review committee, two new items were added to the monthly questionnaire. The first was an item, formerly carried out only intermittently, on whether the unemployed were seeking full- or part-time work. The second was an expanded item on household relationships, formerly included only annually, to provide greater detail on the marital status and household relationship of unemployed persons.

- **March 1963.** The sample and population data used in ratio estimates were revised. From December 1961 to March 1963, the CPS sample was gradually revised. This revision reflected the changes in both population size and distribution as established by the 1960 census. Other demographic changes, such as the industrial mix between areas, were also taken into account. The overall sample size remained the same, but the number of PSUs increased slightly to 357 to provide greater coverage of the fast growing portions of the country. For most of the sample, census lists replaced the traditional area sampling. These lists were developed in the 1960 census. These changes resulted in further gains in reliability of about 5 percent for most statistics. The census-based updated population information was used in April 1962 for first- and second-stage ratio estimates.

- **January 1967.** The sample was expanded to 449 PSUs. The CPS sample was expanded from 357 to 449 PSUs. An increase in total budget allowed the overall sample size to increase by roughly 50 percent to a total of about 60,000 housing units (52,500 occupied units). The expanded sample had households in 863 counties and independent cities with at least some coverage in every state.

  This expansion increased the reliability of the major statistics by about 20 percent and made it possible to publish more detailed statistics.

  The concepts of employment and unemployment were modified. In line with the basic recommendations of the President's Committee to Appraise Employment and Unemployment Statistics (Eckler, 1972), a several-year study was conducted to develop and test proposed changes in the labor force concepts. The principal research results were implemented in January 1967. The changes included a revised age cutoff in defining the labor force; and new questions to improve the information on hours of work, the duration of unemployment, and the self-employed. The definition of unemployment was also revised slightly. The revised definition of unemployment led to small differences in the estimates of level and month-to-month change.

- **March 1968.** Separate age/sex ratio estimation cells were introduced for Negro and other races. Previously, the second-stage ratio estimation used non-White and White race categories by age groups and sex. The revised procedures allowed for separate ratio estimates for Negro and Other[1] race categories.

  This change amounts essentially to an increase in the number of ratio estimation cells from 68 to 116.

- **January 1971 and January 1972.** 1970 census occupational classification was introduced. The questions on occupation were made more comparable to those used in the 1970 census by adding a question on major activities or duties of current job. The new classification was introduced into the CPS coding procedures in January 1971. Tabulated data were produced in the revised version beginning in January 1972.

- **December 1971-March 1973.** Sample was expanded to 461 PSUs and data used in ratio estimation were updated. From December 1971 to March 1973, the CPS sample was revised gradually to reflect the changes in population size and distribution as described by the 1970 census. As part of an overall sample optimization, the sample size was reduced slightly (from 60,000 to 58,000 housing units), but the number of PSUs increased to 461. Also, the cluster design was changed from six nearby (but not contiguous) to four usually contiguous households. This change was undertaken after research found that smaller cluster sizes would increase sample efficiency.

  Even with the reduction in sample size, this change led to a small gain in reliability for most characteristics. The noninterview adjustment and first stage ratio estimate adjustment were also modified to improve the reliability of estimates for central cities and the rest of the Standard Metropolitan Statistical Areas (SMSAs).

  In January 1972, the population estimates used in the second-stage ratio estimation were updated to the 1970 census base.

- **January 1974.** Inflation-deflation method was introduced for deriving independent estimates of the population. The derivation of independent estimates of the civilian non-institutional population by age, race, and sex used in second-stage ratio estimation in preparing the monthly labor force estimates now used the inflation-deflation method (see Chapter 10).

- **September 1975.** State supplementary samples were introduced. An additional sample, consisting of about 14,000 interviews each month, was introduced in July 1975 to supplement the national sample in 26 states and the District of Columbia. In all, 165 new PSUs were involved. The supplemental sample was added to meet a specific reliability standard for estimates of the annual average number of unemployed persons for each state.

---

[1]Other includes American Indian, Eskimo, Aleut, Asian, and Pacific Islander.

In August 1976, an improved estimation procedure and modified reliability requirements led to the supplement PSUs being dropped from three states.

Thus, the size of the supplemental sample was reduced to about 11,000 households in 155 PSUs.

- **October 1978.** Procedures for determining demographic characteristics were modified. At this time, changes were made in the collection methods for household relationship, race, and ethnicity. Race was now determined by the respondent rather than by the interviewer.

Other modifications included the introduction of earnings questions for the two outgoing rotations. New items focused on usual hours worked, hourly wage rate, and usual weekly earnings. Earnings items were asked of currently employed wage and salary workers.

- **January 1979.** A new two-level, first-stage ratio estimation procedure was introduced. This procedure was designed to improve the reliability of metropolitan/nonmetropolitan estimates.

Other newly introduced items were the monthly tabulation of children's demographic data, including relationship, age, sex, race, and origin.

- **September/October 1979.** The final report of the National Commission on Employment and Unemployment Statistics (NCEUS; "Levitan" Commission) (Executive Office of the President, 1976) was issued. This report shaped many of the future changes to the CPS.

- **January 1980.** To improve coverage about 450 households were added to the sample, increasing the number of total PSUs to 629.

- **May 1981.** The sample was reduced by approximately 6,000 assigned households bringing the total sample size to approximately 72,000 assigned households.

- **January 1982.** The race categories in the second-stage ratio estimation adjustment were changed from White/Non-White to Black/Non-Black. These changes were made to eliminate classification differences in race that existed between the 1980 census and the CPS. The change did not result in notable differences in published household data. Nevertheless, it did result in more variability for certain "White," "Black," and "Other" characteristics.

As is customary, the CPS uses ratio estimates from the most recent decennial census. Beginning in January 1982, these ratio estimates were based on findings from the 1980 census. The use of the 1980 census-based population estimates, in conjunction with the revised second-stage adjustment, resulted in about a 2 percent increase in the estimates for total civilian noninstitutional population 16 years and over, civilian labor force, and unemployed persons. The magnitude of the differences between 1970 and 1980 census-based ratio estimates affected the historical comparability and continuity of major labor force series; therefore, the BLS revised approximately 30,000 series back to 1970.

- **November 1982.** The question series on earnings was extended to include items on union membership and union coverage.

- **January 1983.** The occupational and industrial data were coded using the 1980 classification systems. While the effect on industry-related data was minor, the conversion was viewed as a major break in occupation-related data series. The census developed a "list of conversion factors" to translate occupation descriptions based on the 1970 census-coding classification system to their 1980 equivalents.

Most of the data historically published for the "Black and Other" population group were replaced by data which relate only to the "Black" population.

- **October 1984.** School enrollment items were added for persons 16-24 years of age.

- **April 1984.** The 1970 census-based sample was phased-out through a series of changes that were completed by July 1985. The redesigned sample used data from the 1980 census to update the sampling frame, took advantage of recent research findings to improve the efficiency and quality of the survey, and used a state-based design to improve the estimates for the states without any change in sample size.

- **September 1984.** Collection of veteran's data for females was started.

- **January 1985.** Estimation procedures were changed to use data from the 1980 census and new sample. The major changes were to the second-stage adjustment which replaced population estimates for "Black" and "Non-Black" (by sex and age groups) with population estimates for "White," "Black," and "Other" population groups. In addition, a separate, intermediate step was added as a control to the Hispanic[2] population. The combined effect of these changes on labor force estimates and aggregates for most population groups was negligible; however, the Hispanic population and associated labor force estimates were greatly affected and revisions were made back to January 1980 to the extent possible.

- **June 1985.** The CPS Computer Assisted Telephone Interviewing (CATI) facility was opened at Hagerstown, Maryland. A series of tests over the next few years were conducted to identify and resolve the operational issues associated with the use of CATI. Later tests focused on CATI-related issues, such as data quality, costs, and mode effects on labor force estimates. Samples used in these tests were not used as part of the CPS.

- **April 1987.** First CATI cases were used in CPS monthly estimates. Initially, CATI started with 300 cases a month. As operational issues were resolved and new telephone

---

[2]Hispanics may be of any race.

centers were opened—Tucson, Arizona (May 1992) and Jeffersonville, Indiana (September 1994)—the CATI workload was gradually increased to about 9,200 cases a month (January 1995).

- **June 1990.** The first of a series of experiments to test alternative labor force questionnaires was started at the Hagerstown Telephone Center. These tests used random digit dialing and were conducted in 1990 and 1991.

- **July 1992.** The CATI and Computer Assisted Personal Interviewing (CAPI) Overlap (CCO) experiments began. CATI and automated laptop versions of the revised CPS questionnaire were used in a sample of about 12,000 households selected from the National Crime Victimization Survey sample. The experiment continued through December 1993.

  The CCO ran parallel with the official CPS. The CCO's main purpose was to gauge the combined effect of the new questionnaire and computer-assisted data collection. It is estimated that the redesign had no statistically significant effect on the total unemployment rate, but it did affect statistics related to unemployment, such as the reasons for unemployment, the duration of unemployment, and the industry and occupational distribution of the unemployed with previous work experience. It also is estimated that the redesign significantly increased the employment-to-population ratio and the labor force participation rate for women, but significantly decreased the employment-to-population ratio for men. Along with the changes in employment, it was estimated that the redesign significantly influenced the measurement of characteristics related to employment, such as the proportion of employment working part-time, the proportion working part-time for economic reasons, the number of individuals classified as self-employed, and industry and occupational distribution of the employed.

- **January 1994.** A new questionnaire designed solely for use in computer assisted interviewing was introduced in the official CPS. Computerization allowed for the use of a very complex questionnaire without increasing response burden, increased consistency by reducing interviewer error, permitted editing at time of interviewing, and allowed for the use of dependent interviewing where information reported in one month (industry/occupation, retired/disabled statuses, and duration of unemployment) was confirmed or updated in subsequent months.

  Industry and occupation codes from the 1990 census were introduced. Population estimates were converted to 1990 census base for use in ratio estimation procedures.

- **April 1994.** The 16-month phase-in of the redesigned sample based on the 1990 census began. The primary purpose of this sample redesign was to maintain the efficiency of the sampling frames. Once phased-in, this resulted in a monthly sample of 56,000 eligible housing units in 792 sample areas. The details of the 1990 sample redesign are described in Chapter 3.

- **December 1994.** Starting in December 1994, a new set of response categories was phased in for the relationship to reference person. This modification was directed at individuals not formally related to the reference person to identify unmarried partners in a household. The old partner/roommate category was deleted and replaced with the following categories: unmarried partner, housemate/roommate, and roomer/boarder. This modification was phased in two rotation groups at a time and was fully in place by March 1995. This change had no effect on the family statistics produced by CPS.

- **January 1996.** The 1990 CPS design was changed because of a funding reduction. The original reliability requirements of the sample were relaxed, allowing a reduction in the national sample size from roughly 56,000 eligible housing units to 50,000 eligible housing units. The reduced CPS national sample contains 754 PSUs. The details of the sample design changes as of January 1996 are described in Appendix H.

- **January 1998.** A new two-step composite estimation method for the CPS was implemented (See Appendix I). The first step involves computation of composite estimates for the main labor force categories, classified by important demographic characteristics. The second adjusts person weights, through a series of ratio adjustments, to agree with the composite estimates, thus incorporating the effect of composite estimation into the person weights. This new technique provides increased operational simplicity for microdata users and improves the accuracy of labor force estimates by using different compositing coefficients for different labor force categories. The weighting adjustment method assures additivity while allowing this variation in compositing coefficients.

- **July 2001.** Effective with the release of July 2001 data, official labor force estimates from the CPS and Local Area Unemployment Statistics (LAUS) program reflect the expansion of the monthly CPS sample from about 50,000 to about 60,000 eligible households. This expansion of the monthly CPS sample was one part of the Census Bureau's plan to meet the requirements of the State Children's Health Insurance Program (SCHIP) legislation. The SCHIP legislation requires the Census Bureau to improve state estimates of the number of children who live in low-income families and lack health insurance. These estimates are obtained from the Annual Demographic Supplement to the CPS. In September 2000, the Census Bureau began expanding the monthly CPS sample in 31 states and the District of Columbia. States were identified for sample supplementation based on the standard error of their March estimate of low-income children without health insurance. The additional 10,000 households were added to the sample over a 3-month period. The BLS chose not to include the

additional households in the official labor force esti-mates, however, until it had sufficient time to evaluate the estimates from the 60,000 household sample. See Appen-dix J, Changes to the Current Population Survey Sample in July 2001, for details.

REFERENCES

Eckler, R.A. (1972), U.S. Census Bureau, New York: Praeger.

Executive Office of the President, Office of Management and Budget, Statistical Policy Division (1976), **Federal Statistics: Coordination, Standards, Guidelines: 1976,** Washington, DC: Government Printing Office.

Chapter 3.
Design of the Current Population Survey Sample
(See Appendix H for sample design changes as of January 1996 and Appendix J for changes in July 2001.)

INTRODUCTION

For more than five decades, the Current Population Survey (CPS) has been one of the major sources of up-to-date information on the labor force and demographic characteristics of the U.S. population. Because of the CPS's importance and high profile, the reliability of the estimates has been evaluated periodically. The design has often been under constant and close scrutiny in response to demand for new data and to improve the reliability of the estimates by applying research findings and new types of information (especially census results). All changes are implemented with concern for minimizing cost and maximizing comparability of estimates across time. A sample redesign takes place after each census. The most recent decennial revision which incorporated new information from the 1990 census was, in the main, complete as of July 1995. Thus, this chapter describes the CPS sample design as of July 1995.

In January 1996, the CPS design was again modified because of funding reductions. This time, the redesign was restricted to the most populous states where survey requirements were relaxed compared to earlier years. These changes in the sample are not reflected in the main body of this report, but the reader is referred to Appendix H which provides an up-to-date and contemporary description of the changes introduced in 1996. Periodic modifications to the CPS design are occasionally needed to respond to growth of the housing stock. Appendix C discusses how this type of design modification is implemented. Any changes to the CPS design that postdate publication of the present document will be described separately in future appendices.

Effective with the release of July 2001 data, official labor force estimates from the CPS and Local Area Unemployment Statistics (LAUS) program reflect the expansion of the monthly CPS sample. The State Children's Health Insurance Program (SCHIP) legislation requires the Census Bureau to improve state estimates of the number of children who live in low-income families and lack health insurance. These estimates are obtained from the Annual Demographic Supplement to the CPS. States were identified for sample supplementation based on the standard error of their March estimate of low-income children without health insurance. See Appendix J for details.

This chapter is directed to a general audience and presents many topics with varying degrees of detail. The following section provides a broad overview of the CPS design and is recommended for all readers. Later sections of this chapter provide a more in-depth description of the CPS design and are recommended for readers who require greater detail.

SURVEY REQUIREMENTS AND DESIGN

Survey Requirements

The following list briefly describes the major characteristics of the CPS sample as of July 1995:

1. The CPS sample is a probability sample.

2. The sample is designed primarily to produce national and state estimates of labor force characteristics of the civilian noninstitutional population 16 years of age and older (CNP16+).

3. The CPS sample consists of independent samples in each state and the District of Columbia. In other words, each state's sample is specifically tailored to the demographic and labor market conditions that prevail in that particular state. California and New York State are further divided into two substate areas that also have independent designs: the Los Angeles-Long Beach metropolitan area and the rest of California; New York City and the rest of New York State. Since the CPS design consists of independent designs for the states and substate areas, it is said to be state-based.

4. Sample sizes are determined by reliability requirements which are expressed in terms of the coefficient of variation, or CV. The CV is a relative measure of the sampling error and is calculated as sampling error divided by the expected value of the given characteristic. The specified CV for the monthly unemployment level for the nation, given a 6 percent unemployment rate, is 1.8 percent. The 1.8 percent CV is based on the requirement that a difference of 0.2 percent in the unemployment rate for two consecutive months be significant at the 0.10 level.

5. The specified CV for the monthly unemployment level for 11 states, given a 6 percent unemployment rate, is 8 percent.[1] The specified CV for the monthly unemployment level for the California and New York substate areas, given a 6 percent unemployment rate, is 9

---

[1] The 11 states are California, Florida, Illinois, Massachusetts, Michigan, New Jersey, New York, North Carolina, Ohio, Pennsylvania, and Texas.

percent. This latter specification leads to California and New York State having CVs somewhat less than 8 percent.

6. The required CV on the annual average unemployment level for the other 39 states and the District of Columbia, given a 6 percent unemployment rate, is 8 percent.

Overview of Survey Design

The CPS sample is a multistage stratified sample of approximately 56,000 housing units from 792 sample areas designed to measure demographic and labor force characteristics of the civilian noninstitutional population 16 years of age and older. The CPS samples housing units from lists of addresses obtained from the 1990 Decennial Census of Population and Housing. These lists are updated continuously for new housing built after the 1990 census. The first stage of sampling involves dividing the United States into primary sampling units (PSUs) — most of which comprise a metropolitan area, a large county, or a group of smaller counties. Every PSU falls within the boundary of a state. The PSUs are then grouped into strata on the basis of independent information, that is, information obtained from the decennial census or other sources.

The strata are constructed so that they are as homogeneous as possible with respect to labor force and other social and economic characteristics that are highly correlated with unemployment. One PSU is sampled per stratum. The probability of selection for each PSU in the stratum is proportional to its population as of the 1990 census.

In the second stage of sampling, a sample of housing units within the sample PSUs is drawn. Ultimate sampling units (USUs) are clusters of about four housing units. The bulk of the USUs sampled in the second stage consists of sets of addresses which are systematically drawn from sorted lists of addresses of housing units prepared as part of the 1990 census. Housing units from blocks with similar demographic composition and geographic proximity are grouped together in the list. In parts of the United States where addresses are not recognizable on the ground, USUs are identified using area sampling techniques. Occasionally, a third stage of sampling is necessary when actual USU size is extremely large. A final addition to the USUs is a sample of building permits, which compensates for the exclusion of construction since 1990 in the list of addresses in the 1990 census.

Each month, interviewers collect data from the sample housing units. A housing unit is interviewed for 4 consecutive months and then dropped out of the sample for the next 8 months and is brought back in the following 4 months. In all, a sample housing unit is interviewed eight times. Households are rotated in and out of the sample in a way that improves the accuracy of the month-to-month and year-to-year change estimates. The rotation scheme ensures that in any 1 month, one-eighth of the housing units are interviewed for the first time, another eighth is interviewed for the second time, and so on. That is, after the first month, 6 of the 8 rotation groups will have been in the survey for the previous month — there will always be a 75 percent month-to-month overlap. When the system has been in full operation for 1 year, 4 of the 8 rotation groups in any month will have been in the survey for the same month, 1 year ago; there will always be a 50 percent year-to-year overlap. This rotation scheme fully upholds the scientific tenets of probability sampling, so that each month's sample produces a true representation of the target population. Also, this rotation scheme is considered better than other candidate schemes. For example, undue reporting burden expected of survey respondents if they were to constitute a permanent panel is avoided. The properties of the rotation system also show that it could be used to reduce sampling error by use of a composite estimation procedure[2] and, at slight additional cost, by increasing the representation in the sample of USUs with unusually large numbers of housing units.

Each state's sample design ensures that most housing units within a state have the same overall probability of selection. Because of the state-based nature of the design, sample housing units in different states have different overall probabilities of selection. It is true that if we considered only the national level, a more efficient design would result from using the same overall probabilities for all states. Nevertheless, the current system of state-based designs ensures that both the state and national reliability requirements are met.

FIRST STAGE OF THE SAMPLE DESIGN

The first stage of the CPS sample design is the selection of counties. The purpose of selecting a subset of counties instead of having all counties in the sample is to reduce travel costs for the field representatives. Two features of the first-stage sampling are: (1) to ensure that sample counties represent other counties with similar labor force characteristics that are not selected and (2) to ensure that each field representative is allotted a manageable workload in his/her sample area.

The first stage-sample selection is carried out in three major steps:

1. Definition of the PSUs.

2. Stratification of the PSUs within each state.

3. Selection of the sample PSUs in each state.

---

[2]The complete estimation procedure results in significant reduction in the sampling error of estimates of level and change for most items. This procedure depends on the fact that data from previous months are usually highly correlated with the corresponding estimates for the current month.

Definition of the Primary Sampling Units

PSUs are delineated in such a way that they encompass the entire United States. The land areas within each PSU are made reasonably compact so they can be traversed by an interviewer without incurring unreasonable costs. The population is as heterogeneous with regard to labor force characteristics as can be made consistent with the other constraints. Strata are constructed that are homogenous in terms of labor force characteristics to minimize between-PSU variance. Between-PSU variance is a component of total variance which arises from selecting a sample of PSUs rather than selecting housing units from all PSUs. In each stratum, a PSU is selected that is representative of the other PSUs in the same stratum. When revisions are made in the sample each decade, a procedure used for reselection of PSUs maximizes the overlap in the sample PSUs with the previous CPS sample (see Appendix A).

Most PSUs are groups of contiguous counties rather than single counties. A group of counties is more likely to have diverse labor force characteristics rather than a single county. Limits are placed on the geographic size of a PSU to contain the distance a field representative must travel.

After some empirical research in the late 1940s to help establish rules, the PSUs were initially established in late 1949 and early 1950. The original definitions were subsequently modified and now conform to the rules listed below.

Rules for Defining PSUs

1. PSUs are contained within state boundaries.

2. Metropolitan areas are defined as separate PSUs using projected 1990 Metropolitan Statistical Area (MSA) definitions. (An MSA is defined to be at least one county.) If an MSA straddles state boundaries, each state-MSA intersection is a separate PSU.[3]

3. For most states, PSUs are either one county or two or more contiguous counties. For the New England states[4] and part of Hawaii, minor civil divisions (towns or townships) define the PSUs. In some states, county equivalents are used: cities, independent of any county organization, in Maryland, Missouri, Nevada, and Virginia; parishes in Louisiana; and boroughs and census divisions in Alaska.

4. The area of the PSU should not exceed 3,000 square miles except in cases where a single county exceeds the maximum area.

5. The population of the PSU is at least 7,500 except where this would require exceeding the maximum area specified in number 4.

6. In addition to meeting the limitation on total area, PSUs are formed to limit extreme length in any direction and to avoid natural barriers within the PSU.

Combining counties into PSUs. The PSU definitions are reviewed each time the CPS sample design is revised. Before 1980, almost all changes in the composition of the PSUs were made to reflect changes in definitions of MSAs. For 1980, revised PSU definitions reflect new MSA definitions and ensure that the PSU definitions were compatible with a state-based sample design. For 1990, revised PSU definitions reflect changes in MSA definitions and make the PSU definitions more consistent with those used by the other Census Bureau demographic surveys. The following are steps for combining counties, county equivalents, and independent cities into PSUs for 1990.

1. The 1980 PSUs are evaluated by incorporating into the PSU definitions those counties comprising MSAs that are new or have been redefined.

2. Any single county is classified as a separate PSU, regardless of its 1990 population, if it exceeds the maximum area limitation deemed practical for interviewer travel.

3. Other counties within the same state are examined to determine whether they might advantageously be combined with contiguous counties without violating the population and area limitations.

4. Contiguous counties with natural geographic barriers between them are placed in separate PSUs to reduce the cost of travel within PSUs.

5. The proposed combinations are reviewed. Although personal judgment can have no place in the actual selection of sample units, (known probabilities of selection can be achieved only through a random selection process) there are a large number of ways in which a given population can be structured and arranged prior to sampling. Personal judgment legitimately plays an important role in devising an optimal arrangement; that is, one designed to minimize the variances of the sample estimates subject to cost constraints.

These steps result in 2,007 CPS PSUs in the United States from which to draw a sample.

Stratification of Primary Sampling Units

The CPS sample design calls for combining PSUs into strata within each state and selecting one PSU from each stratum. For this type of sample design, sampling theory

---

[3]Final MSA definitions were not available from the Office of Management and Budget when PSUs were defined. Fringe counties having a good chance of being in final MSA definitions are separate PSUs. Most projected MSA definitions are the same as final MSA definitions (Executive Office of the President, 1993).

[4]The New England states are Rhode Island, Connecticut, Massachusetts, Maine, New Hampshire, and Vermont.

suggests forming strata with approximately equal population sizes. When the design is self-weighting (same sampling fraction in all strata) and one field representative is assigned to each sample PSU, equal stratum sizes also have the advantage of providing equal field representative workloads (at least during the early years of each decade, before population growth and migration significantly affect the PSU population sizes). The objective of the stratification, therefore, is to group PSUs with similar characteristics into strata having approximately equal 1990 populations.

Sampling theory also dictates that highly populated PSUs should be selected for sample with certainty. The rationale is that some PSUs exceed or come close to the stratum size needed for equalizing stratum sizes. These PSUs are designated as self-representing (SR); that is, each of the SR PSUs is treated as a separate stratum and is included in the sample.

The following describes the steps for stratifying PSUs for the 1990 redesign.

1. The PSUs required to be SR are identified if the PSU meets one of the following criteria:

   a. The PSU belongs to one of the 150 MSAs with the largest populations in the 1990 census or the PSU contains counties which had a good chance of joining one of these 150 MSAs under final MSA definitions.
   b. The PSU belongs to an MSA that was SR for the 1980 design and among the 150 largest following the 1980 census.

2. The remaining PSUs are grouped into nonself-representing (NSR) strata within state boundaries by adhering to the following criteria:

   a. Roughly equal-sized NSR strata are formed within a state.
   b. NSR strata are formed so as to yield reasonable field representative workloads in an NSR PSU of roughly 45 to 60 housing units. The number of NSR strata in a state is a function of 1990 population, civilian labor force, state CV, and between-PSU variance on the unemployment level. (Workloads in NSR PSUs are constrained because one field representative must canvass the entire PSU. No such constraints are placed on SR PSUs.)
   c. NSR strata are formed with PSUs homogeneous with respect to labor force and other social and economic characteristics that are highly correlated with unemployment. This helps to minimize the between-PSU variance.
   d. Stratification is performed independently of previous CPS sample designs.
      Key variables used for stratification are:

      • Number of male unemployed.

      • Number of female unemployed.

• Number of families with female head of household.

• Ratio of occupied housing units with three or more persons, of all ages, to total occupied housing units.

In addition to these, a number of other variables such as industry and wage variables obtained from the Bureau of Labor Statistics are used for some states. The number of stratification variables in a state ranges from 3 to 12.

Table 3–1 summarizes the number of SR and NSR strata in each state. (The other columns of the table are discussed in later sections of this chapter.)

The algorithm for implementing the NSR stratification criteria for the 1980 and 1990 sample designs is a modified version of the Friedman-Rubin clustering algorithm (Kostanich, 1981). The algorithm consists of three basic steps: hillclimbing, size adjustment, and an exchange pass. For each state, the algorithm identifies a stratification which meets two criteria: (1) all strata are about the same size and (2) the value of the objective function—a scaled total between-PSU variance for all the stratification variables—is relatively small. Each of the algorithm's three steps assigns slightly different priorities to criteria (1) and (2). Before the start of the first step, the program groups the PSUs within a state into randomly defined strata. The algorithm then "swaps" PSUs between strata to reduce size disparity between the strata or to decrease the value of the objective function. The hillclimbing procedure moves PSUs from stratum to stratum, subject to loose size constraints, in order to minimize the between-PSU variance for stratification variables.[5] The size adjustment tightens size constraints and adjusts stratum sizes by making moves that lead to the smallest increases in between-PSU variance. With tight size constraints, the exchange pass seeks to further reduce between-PSU variance by exchanging PSUs between strata.

The algorithm is run several times allowing stratum sizes to vary by differing degrees. A final stratification is chosen which minimizes, to the extent possible, variability in stratum workloads and total between-PSU variance for all stratification variables for the state. If a stratification results in an NSR PSU being placed in a stratum by itself, the PSU is then SR. After the strata are defined, some state sample sizes are adjusted to bring the national CV for unemployment level down to 1.8 percent assuming a 6 percent unemployment rate. (The stratification procedure for Alaska takes into account expected interview cost and between-PSU variance (Ludington, 1992).)

---

[5]Between-PSU variance is the component of total variance arising from selecting a sample of PSUs from all possible PSUs. For this stratification process, the between-PSU variance was calculated using 1990 census data for the stratification variables.

A consequence of the above stratification criteria is that states that are geographically small, mostly urban, or demographically homogeneous are entirely SR. These states are Connecticut, Delaware, Massachusetts, New Hampshire, New Jersey, Rhode Island, Vermont, and the District of Columbia.

Selection of Sample Primary Sampling Units

Each SR PSU is in the sample by definition. As shown in Table 3–1, there are 432 SR PSUs. In each of the remaining 360 NSR strata, one PSU is selected for the sample following the guidelines described next.

At each sample redesign of the CPS, it is important to minimize the cost of introducing a new set of PSUs. Substantial investment has been made in the hiring and training of field representatives in the existing sample PSUs. For each PSU dropped from the sample and replaced by another in the new sample, the expense of hiring and training a new field representative must be accepted. Furthermore, there is a temporary loss in accuracy of the results produced by new and relatively inexperienced field representatives. Concern for these factors is reflected in the procedure used for selecting PSUs.

Objectives of the selection procedure. The selection of the sample of NSR PSUs is carried out within the strata using the 1990 population. The selection procedure accomplishes the following objectives:

1. Select one sample PSU from each stratum with probability proportional to the 1990 population.

2. Retain in the new sample the maximum number of sample PSUs from the 1980 design sample.

Using the Maximum Overlap procedure described in Appendix A, one PSU is selected per stratum with probability proportional to its 1990 population. This procedure uses mathematical programming techniques to maximize the probability of selecting PSUs that are already in sample while maintaining the correct overall probabilities of selection.

Calculation of overall state sampling interval. After stratifying the PSUs within the states, the overall sampling interval in each state is computed. The overall state sampling interval is the inverse of the probability of selection of each housing unit in a state for a self-weighting design. By design, the overall state sampling interval is fixed, but the state sample size is not fixed allowing growth of the CPS sample because of housing units built after the 1990 census. (See Appendix C for details on how the desired sample size is maintained.)

The state sampling interval is designed to meet the requirements for the variance on an estimate of the unemployment level. This variance can be thought of as a sum of variances from the first stage and the second stage of

sample selection.[6] The first-stage variance is called the between-PSU variance and the second-stage variance is called the within-PSU variance. The square of the state CV, or the relative variance, on the unemployment level is expressed as

$$CV^2 = \frac{\sigma_b^2 + \sigma_w^2}{[E(x)]^2} \qquad (3.1)$$

where

σ_b^2 = between-PSU variance contribution to the variance of the state unemployment level estimator.

σ_w^2 = within-PSU variance contribution to the variance of the state unemployment level estimator.

$E(x)$ = the expected value of the unemployment level for the state.

The term, σ_w^2, can be written as the variance assuming a binomial distribution from a simple random sample multiplied by a design effect

$$\sigma_w^2 = \frac{N^2 \, p \, q \, (deff)}{n}$$

where

N   = the civilian noninstitutional population, 16 years of age and older (CNP16+), for the state.

p   = proportion of unemployed in the CNP16+ for the state, or $\frac{x}{N}$.

Substituting.
q = 1 − p.
n = the state sample size.
deff = the state within-PSU design effect. This is a factor accounting for the difference between the variance calculated from a multistage stratified sample and that from a simple random sample.

This formula can be rewritten as

$$\sigma_w^2 = SI \, (x \, q) \, (deff) \qquad (3.2)$$

where

SI   = the state sampling interval, or $\frac{N}{n}$.

Substituting (3.2) into (3.1) and rewriting in terms of the state sampling interval gives

$$SI = \frac{CV^2 x^2 - \sigma_b^2}{x \, q \, deff}$$

---

[6]The variance of an estimator, u, based on a two-stage sample has the general form
$Var(u) = Var_I E_{II}(u| \text{ set of sample PSUs}) + E_I Var_{II}(u| \text{ set of sample PSUs})$
where I and II represent the first and second stage designs, respectively. The left term represents the between-PSU variance, σ_b^2. The right term represents the within-PSU variance, σ_w^2.

Table 3–1. **Number and Estimated Population of Strata for 792-PSU Design by State**

State	Self-representing (SR)		Nonself-representing (NSR)		Overall sampling interval
	Number of strata	Estimated population[1]	Number of strata	Estimated population[1]	
Total	**432**	**141,876,295**	**360**	**45,970,646**	**2,060**
Alabama	4	1,441,118	10	1,604,901	2,298
Alaska	5	287,192	5	91,906	336
Arizona	3	2,188,880	4	544,402	2,016
Arkansas	4	612,369	13	1,151,315	1,316
California	21	20,866,697	5	1,405,500	2,700
Los Angeles	1	6,672,035	0	0	1,691
Remainder of California	20	14,194,662	5	1,405,500	3,132
Colorado	5	1,839,521	5	629,640	1,992
Connecticut	20	2,561,010	0	0	2,307
Delaware	3	509,330	0	0	505
District of Columbia	1	486,083	0	0	356
Florida	20	9,096,732	8	1,073,818	2,176
Georgia	9	2,783,587	9	2,038,631	3,077
Hawaii	3	778,364	1	49,720	769
Idaho	8	416,284	8	302,256	590
Illinois	11	6,806,874	12	1,821,548	1,810
Indiana	9	2,215,745	8	1,949,996	3,132
Iowa	5	710,910	11	1,372,871	1,582
Kansas	3	920,819	9	906,626	1,423
Kentucky	9	1,230,062	9	1,546,770	2,089
Louisiana	7	1,638,220	9	1,403,969	2,143
Maine	8	685,937	4	247,682	838
Maryland	4	3,310,546	2	352,766	3,061
Massachusetts	31	4,719,188	0	0	954
Michigan	12	5,643,817	11	1,345,074	1,396
Minnesota	4	2,137,810	7	1,119,797	2,437
Mississippi	6	540,351	14	1,334,898	1,433
Missouri	7	2,394,137	6	1,457,616	3,132
Montana	8	366,320	8	221,287	431
Nebraska	2	557,203	9	608,513	910
Nevada	4	819,424	2	98,933	961
New Hampshire	13	846,029	0	0	857
New Jersey	11	6,023,359	0	0	1,221
New Mexico	6	685,543	8	410,929	867
New York	13	12,485,029	7	1,414,548	1,709
New York City	1	5,721,495	0	0	1,159
Remainder of New York	12	6,763,534	7	1,414,548	2,093
North Carolina	14	3,089,176	23	1,973,074	1,095
North Dakota	4	232,865	9	235,364	363
Ohio	13	6,238,600	13	1,958,138	1,653
Oklahoma	2	1,244,920	11	1,092,513	1,548
Oregon	4	1,396,261	6	761,794	1,904
Pennsylvania	14	7,704,963	11	1,507,946	1,757
Rhode Island	5	784,090	0	0	687
South Carolina	7	1,434,621	7	1,161,298	2,291
South Dakota	5	211,146	11	291,546	376
Tennessee	8	2,502,671	6	1,219,915	3,016
Texas	22	8,779,997	20	3,613,170	2,658
Utah	2	888,524	3	251,746	958
Vermont	11	428,263	0	0	410
Virginia	12	3,093,947	6	1,612,667	3,084
Washington	5	2,437,454	6	1,216,557	2,999
West Virginia	10	803,797	9	581,544	896
Wisconsin	7	1,773,109	9	1,889,141	2,638
Wyoming	8	227,401	6	98,321	253

[1]Estimate of civilian noninstitutional population 16 years of age and older based on preliminary 1990 census counts.

Generally, this overall state sampling interval is used for all strata in a state yielding a self-weighting state design. (In some states, the sampling interval is adjusted in certain strata to equalize field representative workloads.)

When computing the sampling interval for the current CPS sample, a 6 percent state unemployment rate is assumed for 1995. The results are given in Table 3–1, which was provided earlier.

SECOND STAGE OF THE SAMPLE DESIGN

The second stage of the CPS sample design is the selection of sample housing units within PSUs. The objectives of within-PSU sampling are to:

1. Select a probability sample that is representative of the total civilian, noninstitutional population.

2. Give each housing unit in the population one and only one chance of selection, with virtually all housing units in a state having the same overall chance of selection.

3. For the sample size used, keep the within-PSU variance on labor force statistics (in particular, unemployment) at as low a level as possible, subject to response burden, costs, and other constraints.

4. Select enough within-PSU sample for additional samples that will be needed before the next decennial census.

5. Put particular emphasis on providing reliable estimates of monthly levels and change over time of labor force items.

USUs are the sample units selected during the second stage of the CPS sample design. As discussed earlier in this chapter, most USUs consist of a geographically compact cluster of approximately four addresses, corresponding to four housing units at the time of the census. Use of housing unit clusters lowers travel costs for field representatives. Clustering slightly increases within-PSU variance of estimates for some labor force characteristics since respondents within a compact cluster tend to have similar labor force characteristics.

Overview of Sampling Sources

To accomplish the objectives of within-PSU sampling, extensive use is made of data from the 1990 Decennial Census of Population and Housing and the Building Permit Survey. The 1990 census collected information on all living quarters existing as of April 1, 1990, including characteristics of living quarters as well as the demographic composition of persons residing in these living quarters. Data on the economic well-being and labor force status of individuals were solicited for about 1 in 6 housing units. However, since the census does not cover housing units constructed since April 1, 1990, a sample of building permits issued in

1990 and later is used to supplement the census data. These data are collected via the Building Permit Survey, which is an ongoing survey conducted by the Census Bureau. Therefore, a list sample of census addresses, supplemented by a sample of building permits, is used in most of the United States. However, where city-type street addresses from the 1990 census do not exist, or where residential construction does not need or require building permits, area samples are sometimes necessary. (See the next section for more detail on the development of the sampling frames.)

These sources provide sampling information for numerous demographic surveys conducted by the Census Bureau.[7] In consideration of respondents, sampling methodologies are coordinated among these surveys to ensure a sampled housing unit is selected for one survey only. Consistent definition of sampling frames allows for the development of separate, optimal sampling schemes for each survey. The general strategy for each survey is to sort and stratify all the elements in the sampling frame (eligible and not eligible) to satisfy individual survey requirements, select a systematic sample, and remove the selected sample from the frame. Sample is selected for the next survey from what remains. Procedures are developed to determine eligibility of sample cases at the time of interview for each survey. This coordinated sampling approach is computer intensive and was not possible in previous redesigns.[8]

Development of Sampling Frames

Results from the 1990 census, the Building Permit Survey, and the relationship between these two sources are used in developing sampling frames. Four frames are created: the unit frame, the area frame, the group quarters frame, and the permit frame. The unit, area, and group quarters frames are collectively called old construction. To describe frame development methodology, several terms must be defined.

Two types of living quarters were defined for the census. The first type is a housing unit. A housing unit is a group of rooms or a single room occupied as a separate living quarter or intended for occupancy as a separate living quarter. A separate living quarter is one in which the occupants live and eat separately from all other persons on the property and have direct access to their living quarter

---

[7]CPS sample selection was coordinated with the following demographic surveys in the 1990 redesign: the American Housing Survey - Metropolitan Sample, the American Housing Survey - National sample, the Consumer Expenditure Survey - Diary sample, the Consumer Expenditure Survey - Quarterly sample, the Current Point of Purchase Survey, the National Crime Victimization Survey, the National Health Interview Survey, the Rent and Property Tax Survey, and the Survey of Income and Program Participation.

[8]This sampling strategy is unbiased because if a random selection is removed from a frame, the part of the frame that remains is a random subset. Also, the sample elements selected and removed from each frame for a particular survey have similar characteristics as the elements remaining in the frame.

from the outside or through a common hall or lobby as found in apartment buildings. A housing unit may be occupied by a family or one person, as well as by two or more unrelated persons who share the living quarter. About 98 percent of the population counted in the 1990 census resided in housing units.

The second type of living quarter is a group quarters. A group quarters is a living quarter where residents share common facilities or receive formally authorized care. Examples include college dormitories, retirement homes, and communes. For some group quarters, such as fraternity and sorority houses and certain types of group houses, a group quarters is distinguished from a housing unit if it houses ten or more unrelated people. The group quarters population is classified as institutional or noninstitutional and as military or civilian. CPS targets only the civilian noninstitutional population residing in group quarters. Military and institutional group quarters are included in the group quarters frame and given a chance of selection in case of conversion to civilian noninstitutional housing by the time it is scheduled for interview. Less than 2 percent of the population counted in the 1990 census resided in group quarters.

Old Construction Frames

Old construction consists of three sampling frames: unit, area, and group quarters. The primary objectives in constructing the three sampling frames are maximizing the use of census information to reduce variance of estimates, ensuring adequate coverage, and minimizing cost. The sampling frames used in a particular geographic area take into account three major address features:

1. Type of living quarters — housing units or group quarters.

2. Completeness of addresses — complete or incomplete.

3. Building permit office coverage — covered or not covered.

An address is considered complete if it describes a specific location; otherwise, the address is considered incomplete. (When the 1990 census addresses cannot be used to locate sample units, area listings must be performed in those areas before sample units can be selected for interview. See Chapter 4 for more detail.) Examples of a complete address are city delivery types of mailing addresses composed of a house number, street name, and possibly a unit designation, such as "1599 Main Street" or "234 Elm Street, Apartment 601." Examples of incomplete addresses are addresses composed of postal delivery information without indicating specific locations, such as "PO Box 123" or "Box 4" on a rural route. Housing units in complete blocks covered by building permit offices are assigned to the unit frame. Group quarters in complete blocks covered by building permit offices are assigned to the group quarters frame. Other blocks are assigned to the area frame.

Unit frame. The unit frame consists of housing units in census blocks that contain a very high proportion of complete addresses and are essentially covered by building permit offices. The unit frame covers most of the population. (Although building permit offices cover nearly all blocks in the unit frame, a few exceptions may slightly compromise CPS coverage of the target population (see Chapter 16)). A USU in the unit frame consists of a compact cluster of four addresses, which are identified during sample selection. The addresses, in most cases, are those for separate housing units. However, over time some buildings may be demolished or converted to non-residential use, and others may be split up into several housing units. These addresses remain sample units, resulting in a small variability in cluster size. Also, USUs usually cover neighboring housing units, though, occasionally they are dispersed across a neighborhood, resulting in a USU with housing units from different blocks.

Area frame. The area frame consists of housing units and group quarters in census blocks that contain a high proportion of incomplete addresses, or are not covered by building permit offices. A CPS USU in the area frame also consists of about four housing unit equivalents, except in some areas of Alaska that are difficult to access where a USU is eight housing unit equivalents. The area frame is converted into groups of four housing unit equivalents called "measures" because the census addresses of individual housing units or persons within a group quarters are not used in the sampling.

An integer number of area measures is calculated at the census block level. The number is referred to as the area block measure of size (MOS) and is calculated as follows:

$$\text{area block MOS} = \frac{H}{4} + [\text{GQ block MOS}] \qquad (3.3)$$

where

H	= the number of housing units enumerated in the block for the 1990 census.
GQ block MOS	= the integer number of group quarters measures in a block (see equation 3.4).

The first term of equation (3.3) is rounded to the nearest nonzero integer. When the fractional part is 0.5 and the term is greater than 1, it is rounded to the nearest even integer.

Sometimes census blocks are combined with geographically nearby blocks before the area block MOS is calculated. This is done to ensure that newly constructed units

have a chance of selection in blocks with no housing units or group quarters at the time of the census and that are not covered by a building permit office. This also reduces the sampling variability caused by USU size differing from four housing unit equivalents for small blocks with fewer than four housing units.

Depending on whether or not a block is covered by a building permit office, area frame blocks are classified as area permit or area nonpermit. No distinction is made between area permit and area nonpermit blocks during sampling. Field procedures are developed to ensure proper coverage of housing units built after the 1990 census in the area blocks to (1) prevent these housing units from having a chance of selection in area permit blocks and (2) give these housing units a chance of selection in area nonpermit blocks. These field procedures have the added benefit of assisting in keeping USU size constant as the number of housing units in the block increases because of new construction.

Group quarters frame. The group quarters frame consists of group quarters in census blocks that contain a sufficient proportion of complete addresses and are essentially covered by building permit offices. Although nearly all blocks are covered by building permit offices, some are not, which may result in minor undercoverage. The group quarters frame covers a small proportion of the population. A CPS USU in the group quarters frame consists of four housing unit equivalents. The group quarters frame, like the area frame, is converted into housing unit equivalents because 1990 census addresses of individual group quarters or persons within a group quarters are not used in the sampling. The number of housing unit equivalents is computed by dividing the 1990 census group quarters population by the average number of persons per household (calculated from the 1990 census as 2.63).

An integer number of group quarters measures is calculated at the census block level. The number of group quarters measures is referred to as the GQ block MOS and is calculated as follows:

$$\text{GQ block MOS} = \frac{\text{NIGQPOP}}{(4)(2.63)} + \text{MIL} + \text{IGQ} \qquad (3.4)$$

where

NIGQPOP = the noninstitutional group quarters population in the block from the 1990 census.

MIL = the number of military barracks in the block from the 1990 census.

IGQ = 1 if one or more institutional group quarters are in the block or 0 if no institutional group quarters are in the block from the 1990 census.

The first term of equation (3.4) is rounded to the nearest nonzero integer. When the fractional part is 0.5 and the term is greater than 1, it is rounded to the nearest even integer.

Only the civilian noninstitutional population is interviewed for CPS. Military barracks and institutional group quarters are given a chance of selection in case group quarters convert status over the decade. A military barrack or institutional group quarters is equivalent to one measure regardless of the number of people counted there in the 1990 census.

Special situations in old construction. During development of the old construction frames, several situations are given special treatment. Military and national park blocks are treated as if covered by a building permit office to increase the likelihood of being in the unit or group quarters frames to minimize costs. Blocks in American Indian Reservations are treated as if not covered by a building permit office and are put in the area frame to improve coverage. To improve coverage of newly constructed college housing, special procedures are used so blocks with existing college housing and small neighboring blocks are in the area frame. Blocks in Ohio which are covered by building permit offices that issue permits for only certain types of structures are treated as area nonpermit blocks. Two examples of blocks excluded from sampling frames are blocks consisting entirely of docked maritime vessels where crews reside and street locations where only homeless people were enumerated in the 1990 census.

The Permit Frame

Permit frame sampling ensures coverage of housing units built since the 1990 census. The permit frame grows as building permits are issued during the decade. Data collected by the Building Permit Survey are used to update the permit frame monthly. About 92 percent of the population lives in areas covered by building permit offices. Housing units built since the 1990 census in areas of the United States not covered by building permit offices have a chance of selection in the nonpermit portion of the area frame. Group quarters built since the 1990 census are generally not covered in the permit frame, although the area frame does pick up new group quarters. (This minor undercoverage is discussed in Chapter 16.)

A permit measure which is equivalent to a CPS USU is formed within a permit date and a building permit office resulting in a cluster containing an expected four newly built housing units. The integer number of permit measures is referred to as the BPOMOS and is calculated as follows:

$$\text{BPOMOS}_t = \frac{\text{HP}_t}{4} \qquad (3.5)$$

where

HP_t = the total number of housing units for which the building permit office issues permits for a time period, t, normally a month; for example, a building permit office issued 2 permits for a total 24 housing units to be built in month t.

BPOMOS for time period t is rounded to the nearest integer except when nonzero and less than 1, then it is rounded to 1. Permit cluster size varies according to the number of housing units for which permits are actually issued. Also, the number of housing units for which permits are issued may differ from the number of housing units that actually get built.

When developing the permit frame, an attempt is made to ensure inclusion of all new housing units constructed after the 1990 census. To do this, housing units for which building permits had been issued but which had not yet been constructed by the time of the census should be included in the permit frame. However, by including permits issued prior to the 1990 census in the permit frame, there is a risk that some of these units will have been built by the time of the census and, thus, included in the old construction frame. These units will then have two chances of selection in the CPS: one in the permit frame and one in the old construction frames.

For this reason, permits issued too long before the census should not be included in the permit frame. However, excluding permits issued long before the census brings the risk of excluding units for which permits were issued but which had not yet been constructed by the time of the census. Such units will have no chance of selection in the CPS, since they are not included in either the permit or old construction frames. In developing the permit frame, an attempt is made to strike a reasonable balance between these two problems.

Summary of Sampling Frames

Before providing a summary of the various sampling frames, an exception is noted. Census blocks containing sample selected by the National Health Interview Survey (NHIS) were also included in the area frame to ensure a housing unit was in sample for only one demographic survey. That is, any sample (both housing unit and group quarters) selected for the NHIS was transferred to the area frame. Therefore, group quarters were in both the area and group quarters frames. The NHIS had an all area frame sample design because it was not conducted under Title 13; thus, it was prohibited from selecting a sample of 1990 census addresses.

Table 3–2 summarizes the features of the sampling frames and CPS USU size discussed above. Roughly 65 percent of the CPS sample is from the unit frame, 30 percent is from the area frame, and 1 percent is from the group quarters frame. In addition, about 5 percent of the sample is from the permit frame initially. The permit frame has grown, historically, about 1 percent a year. Optimal cluster size or USU composition differs for the demographic surveys. The unit frame allows each survey a choice of cluster size. For the area, group quarters, and permit frames, MOS must be defined consistently for all demographic surveys.

Table 3–2. **Summary of Sampling Frames**

Frame	Typical characteristics of frame	CPS USU
Unit frame	High percentage of complete addresses in areas covered by a building permit office	Compact cluster of four addresses
Group quarters frame .	High percentage of complete addresses in areas covered by a building permit office	Measure containing group quarters of four expected housing unit equivalents
Area frame		
Area permit	Many incomplete addresses in areas covered by a building permit office	Measure containing housing units and group quarters of four expected housing unit equivalents
Area nonpermit	Not covered by a building permit office	
Permit frame	Housing units built since 1990 census in areas covered by a building permit office	Cluster of four expected housing units

Selection of Sample Units

The CPS sample is designed to be self-weighting by state or substate area. A systematic sample is selected from each PSU at a sampling rate of 1 in k, where k is the within-PSU sampling interval which is equal to the product of the PSU probability of selection and the stratum sampling interval. The stratum sampling interval is usually the overall state sampling interval. (See the earlier section in this chapter, "Calculation of overall state sampling interval.")

The first stage of selection is conducted independently for each demographic survey involved in the 1990 redesign. Sample PSUs overlap across surveys and have different sampling intervals. To make sure housing units get selected for only one survey, the largest common geographic areas obtained when intersecting each survey's sample PSUs are identified. These intersecting areas, as well as the residual areas of those PSUs, are called basic PSU components (BPCs). A CPS stratification PSU consists of one or more BPCs. For each survey, a within-PSU sample is selected from each frame within BPCs. However, sampling by BPCs is not an additional stage of selection. After combining sample from all frames for all BPCs in a PSU, the resulting within-PSU sample is representative of the entire civilian, noninstitutional population of the PSU.

When CPS is not the first survey to select a sample in a BPC, the CPS within-PSU sampling interval is decreased to maintain the expected CPS sample size after other surveys have removed sampled USUs. When a BPC does not include enough sample to support all surveys present in the BPC for the decade, each survey proportionally reduces its expected sample size for the BPC. This makes a state no longer self-weighting, but this adjustment is rare.

CPS sample is selected separately within each sampling frame. Since sample is selected at a constant overall rate, the percentage of sample selected from each frame is proportional to population size. Although the procedure is the same for all sampling frames, old construction sample selection is performed once for the decade while permit frame sample selection is an ongoing process each month throughout the decade.

Within-PSU Sort

Units or measures are arranged within sampling frames based on characteristics of the 1990 census and geography. Sorting minimizes within-PSU variance of estimates by grouping together units or measures with similar characteristics. The 1990 census data and geography are used to sort blocks and units. (Sorting is done within BPCs since sampling is performed within BPCs.) The unit frame is sorted on block level characteristics, keeping housing units in each block together, and then by a housing unit identification to sort the housing units geographically. Sorts are different for each frame and are provided in Tables 3–3a and 3–3b.

General Sampling Procedure

The CPS sampling is a one-time operation that involves selecting enough sample for the decade. To accommodate the CPS rotation system and the phasing in of new sample designs, 19 samples are selected. A systematic sample of USUs is selected and 18 adjacent sample USUs identified. The group of 19 sample USUs is known as a hit string. Due to the sorting variables, persons residing in USUs within a hit string are likely to have similar labor force characteristics.

The within-PSU sample selection is performed independently by BPC and frame. Four dependent random numbers (one per frame) between 0 and 1 are calculated for each BPC within a PSU.[9] Random numbers are used to calculate random starts. Random starts determine the first sampled USU in a BPC for each frame.

The method used to select systematic samples of hit strings of USUs within each BPC and sampling frame follows:

1. Units or measures within the census blocks are sorted using the within-PSU sort criteria specified in Tables 3–3a and 3–3b.

2. Each successive USU not selected by another survey is assigned an index number 1 through N.

3. A random start (RS) for the BPC/frame is calculated. RS is the product of the dependent random number and the adjusted within-PSU sampling interval (SI_w).

4. Sampling sequence numbers are calculated. Given N USUs, sequence numbers are:

   RS, RS + (1(SI_w)), RS + (2(SI_w)), ..., RS + (n(SI_w))
   where n is the largest integer such that RS + (n(SI_w)) ≤ N. Sequence numbers are rounded up to the next integer. Each rounded sequence number represents the first unit or measure designating the beginning of a hit string.

5. Sequence numbers are compared to the index numbers assigned to USUs. Hit strings are assigned to sequence numbers. The USU with the index number matching the sequence number is selected as the first

---

[9]Random numbers are evenly distributed by frame within BPC and by BPC within PSU to minimize variability of sample size.

Table 3–3a. **Old Construction Within-PSU Sorts**

Sort order	Unit and area frames		Group quarters frame
	Urban PSUs	Rural PSUs	All PSUs
1	Block CBUR classification[1]	Block CBUR classification	District Office
2	Proportion of minority renter occupied housing units to all occupied housing units in the block	Proportion of households headed by females to all households in the block	Address Register Area
3	Proportion of owner occupied housing units to total population age 16 years and over	Proportion of minority population to total population in the block	County code
4	Proportion of housing units to population age 16 years and over in the block	Proportion of population age 65 and over to total population in the block	MCD/CCD code
5	Proportion of households headed by females to all households in the block	Proportion of Black renter occupied housing units to total Black population in the block	Block CBUR classification
6	County code	County code	Block number
7	Tract number	Tract number	
8	Combined block number (area frame only)	Combined block number (area frame only)	
9	Housing unit ID (unit frame only)	Housing unit ID (unit frame only)	

[1]A census block is classified as C, B, U, or R, which means: central city of 1990 MSA (C); balance of 1990 urbanized area (B); other urban (U); or rural (R).

Table 3–3b. Permit Frame Within-PSU Sort

Sort order	All PSUs
1	County code
2	Building permit office
3	Permit date

sample. The 18 USUs that follow the sequence number are selected as the next 18 samples. This method may yield hit strings with less than 19 samples (called incomplete hit strings) at the beginning or end of BPCs.[10] Allowing incomplete hit strings ensures that each USU has the same probability of selection.

6. A sample designation uniquely identifying 1 of the 19 samples is assigned to each USU in a hit string. For the 1990 design, sample designations A62 through A80 are assigned sequentially to the hit string. A62 is assigned to the first sample; A63 to the second sample; and assignment continues through A80 for the nineteenth sample. A sample designation suffix, A or B, is assigned in areas of Alaska that are difficult to access (in which USUs consist of eight housing unit equivalents).

Example of Within-PSU Sample Selection for Old Construction

The following example illustrates selection of within-PSU sample for an old construction frame within a BPC. Assume blocks have been sorted within a BPC. The BPC contains 18 unsampled USUs (N=18) and each USU is assigned an index number (1, 2, ..., 18). The dependent random number is 0.6528 and the SI_w is 5.7604. To simplify this example, four samples are selected and sample designations A1 through A4 assigned.

The random start is RS = 0.6528 x 5.7604 = 3.7604. Sequence numbers are 3.7604, 9.5208, and 15.2812, rounding up to 4, 10, and 16. These sequence numbers represent first samples and correspond to index numbers assigned to USUs. A hit string is assigned to each sequence number to obtain the remaining three samples:

4, 10, 16 (first sample),

5, 11, 17 (second sample),

6, 12, 18 (third sample), and

7, 13, 1[11] (fourth sample).

This example includes an incomplete hit string at the beginning and end of the BPC. After sample selection, corresponding sample designations are assigned. Table 3–4 illustrates results from sample selection.

_____

[10]When RS + I > SI_w, an incomplete hit string occurs at the beginning of a BPC. When (RS + I) + (n(SI_w)) > N, an incomplete hit string occurs at the end of a BPC (I = 1 to 18).

[11]Since 3.7604 + I > 5.7604 (RS + I > SI_w) where I = 3, an incomplete hit string occurs at the beginning of the BPC. The sequence number is calculated as RS + I - SI_w.

Although this example is for old construction, selecting a systematic sample for the permit frame is similar except that sampling is performed on an imaginary universe (called a skeleton universe) consisting of an estimated number of USUs within each BPC. As monthly permit information becomes available, the skeleton universe gradually fills with issued permits and eventually specific sample addresses are identified.

Table 3–4. Sampling Example Within a BPC for Any of the Three Old Construction Frames

Census block	USU number within the block	Index number	Sample designation
101 101A	1	1	A4
	2	2	
	3	3	
	4	4	A1
103 104	1	5	A2
	2	6	A3
	3	7	A4
	4	8	
	5	9	
	6	10	A1
	7	11	A2
	8	12	A3
106	1	13	A4
107 107A	1	14	
	2	15	
	3	16	A1
108 108D	1	17	A2
	2	18	A3

Assignment of Post-Sampling Codes

Two types of post-sampling codes are assigned to the sampled units. First, there are the CPS technical codes used to weight the data, estimate the variance of characteristics, and identify representative subsamples of the CPS sample units. The technical codes include final hit number, rotation group, and random group codes. Second, there are operational codes common to the demographic household surveys used to identify and track the sample units through data collection and processing. The operational codes include field PSU, segment number and segment number suffix.

Final hit number. The final hit number identifies the original within-PSU order of selection. All USUs in a hit string are assigned the same final hit number. For each PSU, this code is assigned sequentially starting with one for both the old construction and the permit frames. The final hit number is used in the application of the CPS variance estimation method discussed in Chapter 14.

Rotation group. Sample is partitioned into eight representative subsamples called rotation groups used in the CPS rotation scheme. All USUs in a hit string are assigned to the same rotation group. Assignment is performed separately

for old construction and the permit frame. Rotation groups are assigned after sorting hits by state, MSA/non-MSA status (old construction only), SR/NSR status, stratification PSU, and final hit number. Because of this sorting, the eight subsamples are balanced across stratification PSUs, states, and the nation. Rotation group is used in conjunction with sample designation to determine units in sample for particular months during the decade.

Random group. Sample is partitioned into ten representative subsamples called random groups. All USUs in the hit string are assigned to the same random group. Assignment is performed separately for old construction and the permit frame. Since random groups are assigned after sorting hits by state, stratification PSU, rotation group, and final hit number, the ten subsamples are balanced across stratification PSUs, states, and the Nation. Random groups can be used to partition the sample into test and control panels for survey research.

Field PSU. A field PSU is usually a single county within a stratification PSU, except in the New England states and part of Hawaii where a field PSU is a group of minor civil divisions. Field PSU definitions are consistent across all demographic surveys and are more useful than stratification PSUs for coordinating field representative assignments among demographic surveys.

Segment number. A segment number is assigned to each USU within a hit string. If a hit string consists of USUs from only one field PSU, then the segment number applies to the entire hit string. If a hit string consists of USUs in different field PSUs, then each portion of the hit string/field PSU combination gets a unique segment number. The segment number is a four-digit code. The first digit corresponds to the rotation group of the hit. The remaining three digits are sequence numbers. In any 1 month, a segment within a field PSU identifies one USU or expected four housing units that the field representative is scheduled to visit. A field representative's workload usually consists of a set of segments within one or more adjacent field PSUs.

Segment number suffix. Adjacent USUs with the same segment number may be in different blocks for area and group quarters sample or in different building permit office dates or ZIP Codes for permit sample, but in the same field PSU. If so, an alphabetic suffix appended to the segment number indicates that a hit string has crossed one of these boundaries. Segment number suffixes are not assigned to the unit sample.

Examples of Post-Sampling Code Assignments

Two examples are provided to illustrate assignment of codes. To simplify the examples, only two samples are selected, and sample designations A1 and A2 are assigned. The examples illustrate a stratification PSU consisting of all sampling frames (which often does not occur). Assume the index numbers (shown in Table 3–5) are selected in two BPCs.

These sample USUs are sorted and survey design codes assigned as shown in Table 3–6.

The example in Table 3–6 illustrates that assignment of rotation group and final hit number is done separately for old construction and the permit frame. Consecutive numbers are assigned across BPCs within frames. Although not shown in the example, assignment of consecutive rotation group numbers (modulo 8) carries across stratification PSUs. For example, the first old construction hit in the next stratification PSU is assigned to rotation group 1. However, assignment of final hit numbers is performed within stratification PSUs. A final hit number of 1 is assigned to the first old construction hit and the first permit hit of each stratification PSU. Operational codes are assigned as shown in Table 3–7.

After sample USUs are selected and post-sampling codes assigned, addresses are needed in order to interview sampled units. The procedure for obtaining addresses differs by sampling frame. For operational purposes, identifiers are used in the unit frame during sampling instead of actual addresses. The procedure for obtaining unit frame addresses by matching identifiers to census files is described in Chapter 4. Field procedures, usually involving a listing operation, are used to identify addresses in other frames. A description of listing procedures is also given in Chapter 4. Illustrations of the materials used in the listing phase are shown in Appendix B.

Table 3–5. **Index Numbers Selected During Sampling for Code Assignment Examples**

BPC number	Unit frame	Group quarters frame	Area frame	Permit frame
1	3-4, 27-28, 51-52	10-11	1 (incomplete), 32-33	7-8, 45-46
2	10-11, 34-35	none	6-7	14-15

Table 3–6. **Example of Postsampling Survey Design Code Assignments Within a PSU**

BPC	Frame	Index	Sample designation	Final hit number	Rotation group
1..	Unit	3	A1	1	8
1..	Unit	4	A2	1	8
1..	Unit	27	A1	2	1
1..	Unit	28	A2	2	1
1..	Unit	51	A1	3	2
1..	Unit	52	A2	3	2
1..	Group quarters	10	A1	4	3
1..	Group quarters	11	A2	4	3
1..	Area	1	A2	5	4
1..	Area	32	A1	6	5
1..	Area	33	A2	6	5
2..	Unit	10	A1	7	6
2..	Unit	11	A2	7	6
2..	Unit	34	A1	8	7
2..	Unit	35	A2	8	7
2..	Area	6	A1	9	8
2..	Area	7	A2	9	8
1..	Permit	7	A1	1	3
1..	Permit	8	A2	1	3
1..	Permit	45	A1	2	4
1..	Permit	46	A2	2	4
2..	Permit	14	A1	3	5
2..	Permit	15	A2	3	5

Table 3–7. **Example of Postsampling Operational Code Assignments Within a PSU**

BPC	County	Frame	Block	Index	Sample designation	Final hit number	Field PSU	Segment/ suffix
1....................................	1	Unit	1	3	A1	1	1	8999
1....................................	1	Unit	1	4	A2	1	1	8999
1....................................	1	Unit	2	27	A1	2	1	1999
1....................................	2	Unit	2	28	A2	2	2	1999
1....................................	2	Unit	3	51	A1	3	2	2999
1....................................	2	Unit	3	52	A2	3	2	2999
1....................................	1	Group quarters	4	10	A1	4	1	3599
1....................................	1	Group quarters	4	11	A2	4	1	3599
1....................................	1	Area	5	1	A2	5	1	4699
1....................................	1	Area	6	32	A1	6	1	5699
1....................................	1	Area	6	33	A2	6	1	5699
2....................................	3	Unit	7	10	A1	7	3	6999
2....................................	3	Unit	7	11	A2	7	3	6999
2....................................	3	Unit	8	34	A1	8	3	7999
2....................................	3	Unit	8	35	A2	8	3	7999
2....................................	3	Area	9	6	A1	9	3	8699
2....................................	3	Area	10	7	A2	9	3	8699A
1....................................	1	Permit		7	A1	1	1	3001
1....................................	1	Permit		8	A2	1	1	3001
1....................................	2	Permit		45	A1	2	2	4001
1....................................	2	Permit		46	A2	2	2	4001
2....................................	3	Permit		14	A1	3	3	5001
2....................................	3	Permit		15	A2	3	3	5001

THIRD STAGE OF THE SAMPLE DESIGN

Often, the actual USU size in the field can deviate from what is expected from the computer sampling. Occasionally, the deviation is large enough to jeopardize the successful completion of a field representative's assignment. When these situations occur, a third stage of selection is conducted to maintain a manageable field representative workload. This third stage is called field subsampling.

Field subsampling occurs when a USU consists of more than 15 sample housing units identified for interview. Usually, this USU is identified after a listing operation. (See Chapter 4 for a description of field listing.) The regional office staff selects a systematic subsample of the USU to reduce the number of sample housing units to a more manageable number, from 8 to 15 housing units. To facilitate the subsampling, an integer take-every (TE) and start-with (SW) are used. An appropriate value of the TE

reduces the USU size to the desired range. For example, if the USU consists of 16 to 30 housing units, a TE of 2 reduces USU size to 8 to 15 housing units. The SW is a randomly selected integer between 1 and the TE.

Field subsampling changes the probability of selection for the housing units in the USU. An appropriate adjustment to the probability of selection is made by applying a special weighting factor in the weighting procedure. See "Special Weighting Adjustments" in Chapter 10.

ROTATION OF THE SAMPLE

The CPS sample rotation scheme is a compromise between a permanent sample (from which a high response rate would be difficult to maintain) and a completely new sample each month (which results in more variable estimates of change). The CPS sample rotation scheme represents an attempt to strike a balance in the minimization of the following:

1. Variance of estimates of month-to-month change: three-fourths of the sample is the same in consecutive months.

2. Variance of estimates of year-to-year change: one-half of the sample is the same in the same month of consecutive years.

3. Variance of other estimates of change: outgoing sample is replaced by sample likely to have similar characteristics.

4. Response burden: eight interviews are dispersed across 16 months.

The rotation scheme follows a 4-8-4 pattern. A housing unit or group quarters is interviewed 4 consecutive months, not in sample for the next 8 months, interviewed the next 4 months, and then retired from sample. The rotation scheme is designed so outgoing housing units are replaced by housing units from the same hit string which have similar characteristics.

The following summarizes the main characteristics (in addition to the sample overlap described above) of the CPS rotation scheme:

1. In any 1 month, one-eighth of the sample housing units are interviewed for the first time; another eighth is interviewed for the second time; and so on.

2. The sample for 1 month is composed of units from two or three consecutive samples.

3. One new sample designation-rotation group is activated each month. The new rotation group replaces the rotation group retiring permanently from sample.

4. One rotation group is reactivated each month after its 8-month resting period. The returning rotation group replaces the rotation group beginning its 8-month resting period.

5. Rotation groups are introduced in order of sample designation and rotation group:

   A62(1), A62(2), ..., A62(8), A63(1), A63(2), ..., A63(8), ..., A80(1), A80(2), ..., A80(8).

The present rotation scheme has been used since 1953. The most recent research into alternate rotation patterns was prior to the 1980 redesign when state-based designs were introduced (Tegels, 1982).

The Rotation Chart

The CPS rotation chart illustrates the rotation pattern of CPS sample over time. Figure 3–1 presents the rotation chart beginning in January 1996. The following statements provide guidance in interpreting the chart:

1. Numbers in the chart refer to rotation groups. Sample designations appear in column headings. In January 1996, rotation groups 3, 4, 5, and 6 of A64; 7 and 8 of A65; and 1 and 2 of A66 are designated for interview.

2. Consecutive monthly samples have six rotation groups in common. The sample housing units in A64(4 - 6), A65(8), and A66(1-2), for example, are interviewed in January and February of 1996.

3. Monthly samples 1 year apart have four rotation groups in common. For example, the sample housing units in A65(7-8) and A66(1-2) are interviewed in January 1996 and January 1997.

4. Of the two rotation groups replaced from month-to-month, one is in sample for the first time and one returns after being excluded for 8 months. For example, in October 1996, the sample housing units in A67(3) are interviewed for the first time and the sample housing units in A65(7) are interviewed for the fifth time after last being in sample in January.

Overlap of the Sample

Table 3–8 shows the proportion of overlap between any 2 months of sample depending on the time lag between them. The proportion of sample in common has a strong effect on correlation between estimates from different months and, therefore, on variances of estimates of change.

Table 3–8. **Proportion of Sample in Common for 4-8-4 Rotation System**

Interval (in months)	Percent of sample in common between the 2 months
1...	75
2...	50
3...	25
4-8.......................................	0
9...	12.5
10..	25
11..	37.5
12..	50
13..	37.5
14..	25
15..	12.5
16 and greater............................	0

Phase-In of a New Design

When a newly redesigned sample is introduced into the ongoing CPS rotation scheme, there are a number of reasons not to discard the old CPS sample one month and replace it with a completely redesigned sample the next month. Since redesigned sample contains different sample areas, new field representatives must be hired. Modifications in survey procedures are usually made for a redesigned sample. These factors can cause discontinuity in estimates if the transition is made at one time.

Instead, a gradual transition from the old sample design to the new sample design is undertaken. Beginning in April 1994, the 1990 census-based design was phased in through a series of changes completed in July 1995 (U.S. Department of Labor, 1994).

Figure 3–1. **CPS Rotation Chart: January 1996 - April 1998**

Year/month	A64	A65	A66	A67	A68	A69
			Sample designation and rotation groups			
1996 Jan Feb Mar Apr	3456 4567 5678 678	78 8 1	12 123 1234 2345			
May June July Aug	78 8	12 123 1234 2345	3456 4567 5678 678	1		
Sept Oct Nov Dec		3456 4567 5678 678	78 8 1	12 123 1234 2345		
1997 Jan Feb Mar Apr		78 8	12 123 1234 2345	3456 4567 5678 678	1	
May June July Aug			3456 4567 5678 678	78 8 1	12 123 1234 2345	
Sept Oct Nov Dec			78 8	12 123 1234 2345	3456 4567 5678 678	1
1998 Jan Feb Mar Apr				3456 4567 5678 678	78 8 1	12 123 1234 2345

REFERENCES

Executive Office of the President, Office of Management and Budget (1993), **Metropolitan Area Changes Effective With the Office of Management and Budget's Bulletin 93-17**, June 30, 1993.

Hansen, Morris H., Hurwitz, William N., and Madow, William G. (1953), **Survey Sample Methods and Theory**, Vol. I, **Methods and Applications**. New York: John Wiley and Sons.

Hanson, Robert H. (1978), **The Current Population Survey: Design and Methodology,** Technical Paper 40, Washington, DC: Government Printing Office.

Kostanich, Donna, Judkins, David, Singh, Rajendra, and Schautz, Mindi (1981), "Modification of Friedman-Rubin's Clustering Algorithm for Use in Stratified PPS Sampling," paper presented at the 1981 Joint Statistical Meetings, American Statistical Association.

Ludington, Paul W. (1992), "Stratification of Primary Sampling Units for the Current Population Survey Using Computer Intensive Methods," paper presented at the 1992 Joint Statistical Meetings, American Statistical Association.

Statt, Ronald, Vacca, E. Ann, Wolters, Charles, and Hernandez, Rosa (1981), "Problems Associated With Using Building Permits as a Frame of Post-Census Construction: Permit Lag and ED Identification," paper presented at the 1981 Joint Statistical Meetings, American Statistical Association.

Tegels, Robert, and Cahoon, Lawrence (1982), "The Redesign of the Current Population Survey: The Investigation Into Alternate Rotation Plans," paper presented at the 1982 Joint Statistical Meetings, American Statistical Association.

U.S. Department of Labor, Bureau of Labor Statistics (1994), "Redesign of the Sample for the Current Population Survey," **Employment and Earnings**, Washington, DC: Government Printing Office, May 1994.

Chapter 4.
Preparation of the Sample

INTRODUCTION

The sample preparation operations have been developed to fulfill the following goals:

1. Implement the sampling procedures described in Chapter 3.

2. Produce virtually complete coverage of the eligible population.

3. Ensure that only a trivial number of households will appear in the sample more than once over the course of a decade, or will appear in more than one of the household surveys conducted by the Bureau of the Census.

4. Provide cost efficient collection by producing most of the sampling materials needed for both the Current Population Survey (CPS) and other household surveys in a single, integrated operation.

The CPS is one of many household surveys conducted on a regular basis by the Census Bureau. Insofar as possible, Census Bureau programs have been designed so that survey materials, survey procedures, personnel, and facilities can be used by as many surveys as possible. Sharing personnel and sampling material among a number of programs yields obvious savings. For example, CPS field representatives can be (and often are) employed on non-CPS activities because the sampling materials, listing and coverage instructions, and to a lesser extent, questionnaire content are similar for a number of different programs. In addition, the sharing of sampling materials makes it possible to keep respondents from being in more than one sample.

The postsampling codes described in Chapter 3 identify, among other information, the sample cases that are scheduled to be interviewed for the first time in each month of the decade and indicate the types of materials (maps, listing of addresses, etc.) needed by the census field representative to locate the sample addresses. This chapter describes how these materials are put together.

This chapter may occasionally indulge in detail beyond the interest of the general reader. The next section is an overview that is recommended for all readers. Subsequent sections provide a more in-depth description of the CPS sample preparation and will prove useful to those interested in greater detail.

Census Bureau headquarters are located in the Washington, DC area. Staff at headquarters coordinate CPS functions ranging from sample design, sample selection, and resolution of subject matter issues to administration of the interviewing staffs maintained under the 12 regional offices and data processing. Census Bureau staff located in Jeffersonville, IN also participate in CPS planning and administration. Their responsibilities include preparation and dissemination of interviewing materials, such as maps and segment folders. Nonetheless, the successful completion of the CPS data collection operations rests on the combined efforts of all Census Bureau offices.

Monthly sample preparation of the CPS has three major components:

1. Identifying addresses.

2. Listing living quarters.

3. Assigning sample to field representatives.

The within-PSU sample described in Chapter 3 is selected from four distinct sampling frames, not all of which consist of specific addresses. Since the field representatives need to know the exact location of the households or group quarters they are going to interview, much of sample preparation involves the conversion of selected sample (e.g., maps, lists of building permits) to a set of addresses. This conversion is described below.

Address Identification in the Unit Frame

About 65 percent of the CPS sample is selected from the unit frame. The unit frame sample is selected from a 1990 census file that contains the information necessary for within-PSU sample selection, but does not contain address information. The address information from the 1990 census is stored in a separate file. The addresses of unit segments are obtained by matching the file of 1990 census information to the file containing the associated 1990 census addresses. This is a one-time operation for the entire unit sample and is performed at headquarters. If the addresses are thought to be incomplete (missing a house number or street name), the 1990 census information is reviewed in an attempt to complete the address before sending it to the field representative for interview.

Address Identification in the Area Frame

About 30 percent of the CPS sample is selected from the area frame. Measures of expected four housing units are selected during within-PSU sampling instead of selecting

housing units because the addresses are not city-style or there is no building permit office coverage. This essentially means that no particular housing units are as yet associated with the selected measure. The only information available is a map of the block which contains the area segment, the number of measures the block contains, and which measure is associated with the area segment. Before the individual housing units associated with the area segment can be identified, additional procedures are used to ensure that field representatives can locate the housing units and that all newly built housing units have a probability of selection. A field representative will be sent to visit and canvass the block to create a complete list of the housing units located in the block. This is referred to as a listing operation which is described more thoroughly in the next section. A systematic sampling pattern is applied to this listing to identify the housing units in the area segment that are designated for each month's sample.

Systematic sampling pattern. Applying the systematic sampling pattern to a listing is referred to as a Start-with/Take-every procedure. Start-with/Take-every's are calculated at headquarters based on the first hit measure in the block and on the number of measures in the block, respectively. For more information on hits and blocks, see Chapter 3. The regional offices use the Start-with/Take-every information to identify a sample housing unit to be interviewed. In simplistic terms, the Start-with identifies the line number on the listing sheet where the sampling pattern begins; the Take-every identifies the frequency with which sample housing units or lines on the listing sheet are selected.

To apply a Start-with/Take-every for a block, the regional office would begin with the first listing sheet for the block counting down the number of lines as indicated by the Start-with. This resulting line number identifies a selected sample housing unit. Beginning with the next line listed, the regional office would count down the number of lines equal to the Take-every. The resulting line identifies the next selected sample housing unit. The regional office continues the above approach until the sampling pattern has been applied to all listed units. For example, a Start-with/Take-every of 3/4 indicates that when the regional office applies the sampling pattern, they would start on line three of the listing sheet and take every fourth line. Suppose the listing sheet had 12 listed lines, a Start-with/Take-every results in the units on lines 3, 7, and 11 being selected for interview. See Figure 4-1.

Address Identification in the Group Quarters Frame

About 1 percent of the CPS sample is selected from the group quarters frame. The decennial census files did not have information on the characteristics of the group quarters. The files contain information about the residents as of April 1, 1990, but there was no information about their living arrangements within the group quarters. Consequently, the census information did not provide a tangible sampling unit for the CPS. Measures were selected during within-PSU sampling since there was no way to associate the selected sample cases with persons to interview at a group quarters. A two-step process is used to identify the group quarters segment. First, the group quarters addresses are obtained by matching to the file of 1990 census addresses, similar to the process for the unit frame. This is a one-time operation done at headquarters. Before the persons living at the group quarters associated with the group quarters segment can be identified, an interviewer visits the group quarters and creates a complete list of eligible sample units (consisting of persons, rooms, or beds). This is referred to as a listing operation. Then a systematic sampling pattern, as described above, is applied to the listing to identify the individuals at the group quarters in the group quarters segment.

Address Identification in the Permit Frame

The proportion of the CPS sample selected from the permit frame keeps increasing over the decade as new housing units are constructed. The CPS sample redesigns are introduced about 4 or 5 years after each decennial census, and at this time the permit sample makes up about 5 percent of the CPS sample; this has historically increased about 1 percent a year. Hypothetical measures are selected during within-PSU sampling (skeleton sampling) in anticipation of the construction of new housing units. Identifying the addresses for these permit measures involves a listing operation at the building permit office, clustering of addresses to form measures, and associating these addresses with the hypothetical measures (or USUs) in sample.

The Census Bureau conducts the Building Permit Survey which collects information on a monthly basis from each building permit office in the Nation about the number of housing units authorized to be built. The Building Permit Survey results are converted to measures of expected four housing units. These measures are continuously accumulated and linked with the frame of hypothetical measures used to select the CPS sample. This matching identifies which building permit office contains the measure that is in sample. A field representative then visits the building permit office to obtain a list of addresses of units that were authorized to be built; this is the Permit Address List (PAL) operation. This list of addresses is keyed and transmitted to headquarters where clusters are formed that correspond one-to-one with the measures. Using this link, the clusters of four addresses in each permit segment are identified.

Forming clusters. To help ensure some geographic clustering of addresses within permit measures and to make PAL listing more efficient, information collected by the

Figure 4–1. **Example of Systematic Sampling**

FORM 11-5 (2-16-93) — U.S. DEPARTMENT OF COMMERCE BUREAU OF THE CENSUS **AREA SEGMENT LISTING SHEET**	1. RO: 23	2. PSU: 42063	3. Segment Number: 4509	4. Survey: CPS
	5. Post Office and ZIP Code: Indiana PA 15701			6. NHIS Week/Stratum

7. Listing

Line No. (a)	Street Name (b)	House No. OR Description/Location of unit (c)	Sample Desig. (d)	Serial No. (e)	Remarks — Reason & date for changes (f)	Year Built B=Before 4/1/90 A=After 4/1/90 ND=Not to be determined (g)
1	Hwy 12	14 Name / Extra Unit:				B A ND
2		18 Name / Extra Unit:				B A ND
3		20 Name / Extra Unit:	A64			B A ND
4	Road A	3 Name / Extra Unit:				B A ND
5		57 Name / Extra Unit:				B A ND
6		2 story brick, white shutters, shed on left. Name / Extra Unit:				B A ND
7	Elkton Dr.	70 Name / Extra Unit:	A64			B A ND
8		72 Name / Extra Unit:				B A ND
9		76 Name / Extra Unit:				B A ND
10		78 Name / Extra Unit:				B A ND
11		82 Name / Extra Unit:	A64			B A ND
12		86 Name / Extra Unit:				B A ND
13		Name / Extra Unit:				B A ND
14		Name / Extra Unit:				B A ND
15		Name / Extra Unit:				B A ND

8. Listed and Updated — a. FR code/initials: CC/ABC — b. Date Month/Year: 1/94

9. Footnotes

Copy distribution: **WHITE** – Segment Folder **YELLOW** – Regional Office Sheet 1 of 1 sheets

Survey of Construction (SOC)[1] is used to identify many of the addresses in the permit frame. For building permit offices included in SOC, the Census Bureau collects information on the characteristics of units to be built for each permit issued by the building permit office. This information is used to form measures in SOC building permit offices. This is not the case for non-SOC building permit offices.

1. SOC PALs are listings from building permit offices that are in the SOC. If a building permit office is in the SOC, then the actual permits issued by the building permit office and the number of units authorized by each permit (though not the addresses) are known in advance of the match to the skeleton universe. Therefore, the measures for sampling are formed directly from the actual permits. The sample permits can then be identified. These sample permits are the only ones for which addresses are collected.

   Because measures for SOC permits were constrained to be within permits, once the listed addresses are complete, the formation of clusters follows easily. The measures formed at the time of sampling are, in effect, the clusters. The sample measures within permits are fixed at the time of sampling; that is, there cannot be any later rearrangement of these units into more geographically compact clusters without voiding the original sampling results.

   Even without geographic clustering, there is some degree of compactness inherent in the manner in which measures are formed for sampling. The units within the SOC permits are assigned to measures in the same order in which they were listed for SOC; thus units within apartment buildings will normally be in the same clusters, and permits listed on adjacent lines on the SOC listing often represent neighboring structures.

2. Non-SOC PALs are listings from building permit offices not in the SOC. At the time of sampling, the only data known for non-SOC building permit offices is a cumulative count of the units authorized on all permits for an office for a given month (or year). Therefore, all addresses for a building permit office/date are collected, together with the number of apartments in multiunit buildings. The addresses are clustered using all units on the PAL.

   The purpose of clustering is to group units together geographically, thus enabling a reduction in field travel costs. For multiunit addresses, as many whole clusters as possible are created from the units within each address. The remaining units on the PAL are clustered within ZIP Code and permit day of issue.

---

[1]The Survey of Construction (SOC) is conducted by the U.S. Census Bureau in conjunction with the Department of Housing and Urban Development. It provides current regional statistics on starts and completions of new single-family and multifamily units and sales of new one-family homes.

LISTING ACTIVITIES

When address information from the census is not available or the address information from the census no longer corresponds to the current address situation, then a listing of all eligible units must be created. Creating this list of basic addresses is referred to as listing. Listing can occur in all four frames: units within multiunit structures, living quarters in blocks, units or residents within group quarters, and addresses for building permits issued.

The living quarters to be listed are usually housing units. In group quarters such as transient hotels, rooming houses, dormitories, trailer camps, etc., where the occupants have special living arrangements, the living quarters listed may be housing units, rooms, beds, etc. In this discussion of listing, all of these living quarters are included in the term "unit" when it is used in context of listing or interviewing. Completed listings are sampled by regional office staff. Performing the listing and sampling in two separate steps allows each step to be verified and allows a more complete control of sampling to avoid bias in designating the units to be interviewed.

The units to be interviewed are selected as objectively as possible to reduce bias. However, the sample rotation system used in the CPS tends to remove biases that may have been introduced in the selection process. For example, if a unit USU is to be defined within a large structure where units will be assigned to several USUs, the field representative could conceivably manipulate the order of listing of the housing units to bias the selection for a given USU. However, as the CPS rotation system eventually replaces each USU by others made up of units adjacent on the listing at the address, the unknown error from the selection bias is replaced by a (measurable) variance.

In order to ensure accurate and complete coverage of the area and group quarters segments, updating of the initial listings is done periodically throughout the decade. The updating ensures that changes such as units missed in the initial listing, demolished units, residential/commercial conversions, and new construction are accounted for.

Listing in the Unit Frame

Listing in the unit frame is usually not necessary. The only time it is done is when the field representative discovers that the address information from the 1990 census is no longer accurate for a multiunit structure and the field representative cannot adequately reconcile the differences.

For multiunit addresses (addresses where the expected number or unit designations is two or more), the field representative receives preprinted (computer-generated) Unit/Permit Listing Sheets (Form 11-3) showing unit designations as recorded in the 1990 census. For multiunits containing between two and nine units, the listing sheet displays the unit designations of all units in the structure

even if some of the units are not in sample. For multiunit structures with ten or more units, the listing sheet contains only the unit designations of sample cases. Other important information on the Unit/Permit Listing Sheet includes: the address, the expected number of units at the address, the sample designations, and serial numbers for the selected sample units, and the 1990 census IDs corresponding to those units. For an example of a Unit/Permit Listing Sheets see Appendix B.

The first time a multiunit address enters the sample, the field representative does one or more of the following:

• Verifies that the 1990 census information on the Unit/Permit Listing Sheet (Form 11-3) is accurate.

• Corrects the listing sheet when it does not agree with what the field representative finds at the address.

• Relists the address on a blank Unit/Permit Listing Sheet (only for some large multiunit listings).

After the field representative has an accurate listing sheet, he/she conducts an interview for each unit that has a current (preprinted) sample designation. If an address is relisted, the field representative provides information to the regional office on the relisting. The regional office staff will resample the listing sheet and provide the line numbers for the specific lines on the listing sheet that identify the units that should be interviewed.

A regular system of updating the listing in unit segments is not followed; however, the field representative may correct an in-sample listing during any visit if a change is noticed. The change may result in additional units being added or removed from sample.

For single-unit addresses, a preprinted listing sheet is not provided to the field representative since only one unit is expected based on the 1990 census information. If a field representative discovers other units at the address at the time of interview, these additional units are interviewed.

Listing in the Area Frame

All blocks that contain area frame sample cases must be listed. Several months before the first area segment in a block is to be interviewed, a field representative visits the block to establish a list of living quarters. Units within the block are recorded on an Area Segment Listing Sheet (Form 11-5). The heading entries on the form are all filled by a regional office clerk, and the field representative records a systematic list of all occupied and vacant housing units and group quarters within the block boundaries.

If the area segment is within a jurisdiction where building permits are issued, housing units constructed since April 1, 1990, are eliminated from the area segment through the year built procedure to avoid giving an address more than one chance of selection. This is required because housing units constructed since April 1, 1990, Census Day, are represented by segments in the permit frame. When there

has been significant growth in housing units since April 1, 1990, it is necessary to determine the year built for every structure in the segment at the time of listing. To determine ''year built,'' the field representative inquires at each listed unit and circles the appropriate code on the listing sheet. For an example of an Area Segment Listing Sheet see Appendix B. The inquiry at the time of listing is omitted in areas with low new construction activity; in such cases, new construction units are identified later in completing the coverage questions during the interview. This process avoids an extra contact at units built before 1990.

If an area segment is not in a building permit issuing jurisdiction, then housing units constructed after the 1990 census do not have a chance of being selected for interview in the permit frame. The field representative does not determine ''year built'' for units in such blocks.

After the listing of living quarters in the area segment has been completed, the listing forms are returned to the regional office. A clerk in the regional office then applies the sampling pattern to identify the units to be interviewed. (See the Systematic sampling pattern section above.)

To reflect changes in the number of units in each sample block, it is desirable to periodically review and correct (i.e., update) each segment's listing so that the current USU, and later USUs to be interviewed, will represent the current status. The following rule is used: The USU being interviewed for the first time for CPS must be identified from a listing that has been updated within the last 24 months.

The area listing sheet is updated by retracing the path of travel of the original lister, verifying the existence of each unit on the listing sheet, accounting for units no longer in existence, and appending any unlisted housing units or other living quarters that are found at the end of the list. By extending the original sampling pattern, the appended units are given their chance of being in sample.

Listing in the Group Quarters Frame

Group quarters addresses in the CPS sample are listed if they are in the group quarters, area, or unit frames. Group quarters found in the permit frame are not listed. Before the first interviews at a group quarters address can be conducted, a field representative visits the group quarters to establish a list of eligible units at the group quarters. The same group quarters procedures for creating a list of eligible units applies for group quarters found in the area or unit frame. The only difference is the time frame in which the listing is done. If the group quarters is in the group quarters or area frame, then the listing is done several months before interview. If the group quarters is discovered at the address of unit frame sample, then the listing is done at the time of interview.

Group Quarters Listing Sheets (Form 11-1) are used to record the group quarters name, group quarters type, address, the name and telephone number of a contact person, and to list the eligible units within the group

quarters. Institutional and verified military group quarters are not listed by a field representative. For an example of a Group Quarters Listing Sheet, see Appendix B.

Each eligible unit of a group quarters is listed on a separate line of the group quarters listing sheet, regardless of the number of units. If there is more than one group quarters within the block, each group quarters must be listed on a separate listing sheet. More detailed information on the listing of group quarters can be found in the Listing and Coverage Manual for Field Representatives. The rule for the frequency of updating group quarters listings is the same as for area segments.

Listing in the Permit Frame

There are two phases of listing in the permit frame. The first is the PAL operation which establishes a list of addresses authorized to be built by a building permit office. This is done shortly after the permit has been issued by the building permit office and associated with a sample hypothetical measure. The second listing is required when the field representative visits the unit to conduct an interview and discovers that the original address information collected through the PAL operation is not accurate and cannot be reconciled.

PAL operation. For each building permit office containing a sample measure, a PAL Form (Form 11-193A) is computer-generated. A field representative visits the building permit office and completes the PALs by listing the necessary permit and address information. If an address given on a permit is missing a house number or street name (or number), then the address is considered incomplete. In this case, the field representative visits the new construction site and draws a sketch map showing the location of the structure and, if possible, completes the address.

Permit listing. Listing in the permit frame is necessary for all multiunit addresses. The PAL operation does not obtain unit designations at multiunit addresses. Therefore, listing is necessary to complete the addresses. Additionally, listing is required when the addresses obtained through the PAL operation do not correspond to what the field representative actually finds when visiting the address for the first interview. When this occurs, the unit frame listing procedures are followed.

THIRD STAGE OF THE SAMPLE DESIGN (SUBSAMPLING)

Chapter 3 describes a third stage of the sample design. This third stage is referred to as subsampling. The need for subsampling is dependent on the results of the listing operations and the results of the clerical sampling. Subsampling is required when the number of housing units in a segment for a given sample is greater than 15 or when 1 of

the 4 units in the USU yields more than 4 total units. For unit segments (and permit segments) this can happen when more units than expected are found at the address at the time of the first interview. For more information on subsampling, see Chapter 3.

INTERVIEWER ASSIGNMENTS

The final stage of sample preparation includes those operations that are needed in order to break the sample down into manageable interviewer workloads and to get the resulting assignments to the field representatives for interview. At this point, all the sample cases for the month have been identified and all the necessary information about these sample cases is available in a central database at headquarters. The listings have been completed, sampling patterns have been applied, and the addresses are available. The central database also includes additional information, such as telephone numbers for those cases which have been interviewed in previous months. The rotation of sample used in the CPS is such that seven-eighths of the sample cases each month have been in sample in previous months (see Chapter 3). This central database is part of an integrated system described briefly below. This integrated system also affects the management of the sample and the way the interview results are transmitted to central headquarters, as described in Chapter 8.

Overview of the Integrated System

In recent years, technological advances have changed the face of data preparation and collection activities at the Census Bureau. The Census Bureau has been developing computer-based methods for survey data collection, communications, management, and analysis. Within the Census Bureau, this integrated data collection system is called the Computer Assisted Interviewing System. The computer assisted interviewing system has two principal components: computer assisted personal interviewing (CAPI) and centralized computer assisted telephone interviewing (CATI). See Chapter 7 on the data collection aspects of this system. The integrated system is designed to manage decentralized data collection using laptop computers, a centralized telephone collection, and a central database for data management and accounting.

The integrated system is made up of three main parts:

1. **Headquarters operations in the central database.** The headquarters operations include transmission of cases to field representatives, transmission of CATI cases to the telephone centers, and database maintenance.

2. **Regional office operations in the central database.** The regional office operations include sample maintenance, keying of addresses as necessary, preparation of assignments, determination of CATI assignments, reinterview selection, and review and reassignment of cases.

3. **Field representative case management operations on the laptop computer.** The field representative operations include receipt of assignments, completion of interview assignments, and transmittal of completed work to the central database for processing (see Chapter 8).

The central database resides at headquarters where the file of sample cases is maintained. The database stores field representative data (name, phone number, address, etc.), information for making assignments (Field PSU, segment, address, etc.), and all the data for cases in sample.

Regional Office Operations

When the information in the database is complete, the regional offices can begin the assignment preparation phase. This includes making assignments to field representatives and sending cases to a centralized telephone facility. Regional offices access the central database to break down the assignment areas geographically and key in information that is used to aid in the monthly field representative assignment operations. The regional office supervisor considers such characteristics as the size of the Field PSU, the workload in that Field PSU, and the number of field representatives working in that Field PSU when deciding the best geographic method for dividing the workload in Field PSUs among field representatives.

As part of the process of making field representative assignments, the CATI assignments are also made. Each regional office is informed of the recommended number of cases that should be sent to CATI. The regional office attempts to assign the recommended number of cases for centralized telephone interviewing. The selection of cases for CATI involves several steps. Cases may be assigned to CATI if several criteria are met. These criteria pertain to the Field PSU, the household, and the time in sample. In general terms, the criteria are as follows:

1. The sample case must be in a Field PSU and random group that is CATI eligible. (The random group restriction is imposed to allow for statistical testing. See Chapter 3 for further description of random group.)

2. The household must have a telephone and be willing to accept a telephone interview.

3. First and fifth month cases are generally not eligible for a telephone or CATI interview.

The regional offices have the ability to temporarily assign cases to CATI in order to cover their workloads in certain situations. Each region has provided headquarters with the number of CATI cases that they could provide to CATI on a temporary basis each month. This is done primarily to fill in for field representatives that are ill or on vacation. When interviewing for the month is completed, these cases will automatically be reassigned for CAPI interviewing.

Many of the cases sent to CATI are successfully completed as telephone interviews. Those that cannot be completed from the telephone centers are returned to the field prior to the end of the interview period. These cases are called ''CATI recycles.'' See Chapter 8 for further detail.

The final step is the certification of assignments. After all changes to the interview and CATI assignments have been made, the regional offices certify the assignments. This must be done after all assignments have been reviewed for geographic efficiency and have the proper balance among field representatives.

After assignments are made and certified, the regional offices transmit the assignments to the central database. This transmission places the assignments on the telecommunications server for the field representatives and centralized telephone facilities. Prior to the interview period, field representatives receive their assignments by initiating a transmission to the telecommunications server at headquarters. Assignments include the instrument (questionnaire and/or supplements) and the cases they must interview that month. These files are copied to the laptop during the transmission from the server. The instrument includes the control card or household demographic information and labor force questions. All data sent and received from the field representatives pass through the central communications system maintained at headquarters. See Chapter 8 for more information on the transmission of interview results.

Finally, the regional offices prepare the remaining paper materials needed by the field representatives to complete their assignments. The materials include:

1. Field Representative Assignment Listing (CAPI-35).

2. Segment folders for cases to be interviewed (including maps, listing sheets, and other pertinent information that will aid the interviewer in locating specific cases).

3. Blank listing sheets, payroll forms, respondent letters, and other supplies requested by the field representative.

4. Segment folders and listing sheets for blocks and group quarters to be listed for future CPS samples.

Once the field representative has received the above materials and has successfully completed a transmission to retrieve his/her assignments, the sample preparation operations are complete and the field representative is ready to conduct the interviews.

Chapter 5.
Questionnaire Concepts and Definitions for the Current Population Survey

INTRODUCTION

An extremely important component of the Current Population Survey (CPS) is the questionnaire, also called the survey instrument. Although the concepts and definitions of both labor force items and associated demographic information collected in the CPS have, with a few exceptions, remained relatively constant over the past several decades, the survey instrument was radically redesigned in January 1994. The purpose of the redesign was to reflect changes that have occurred in the economy and to take advantage of the possibilities inherent in the automated data collection methods introduced in 1994, that is, computer assisted personal interviewing (CAPI) and computer assisted telephone interviewing (CATI). This chapter briefly describes and discusses the current survey instrument: its concepts, definitions, and data collection procedures and protocols.

STRUCTURE OF THE SURVEY INSTRUMENT

The CPS interview is divided into three basic parts: (1) household and demographic information, (2) labor force information, and (3) supplement information in months that include supplements. Household and demographic information historically was called "control card information" because, in the paper-and-pencil environment, this information was collected on a separate cardboard form. With electronic data collection, this distinction is no longer apparent to the respondent or the interviewer. The order in which interviewers attempt to collect information is as follows: (1) housing unit data, (2) demographic data, (3) labor force data, (4) more demographic data, (5) supplement data, and finally (6) more housing unit data.

Only the concepts and definitions of the household, demographic, and labor force data are discussed below. (For more information about supplements to the CPS see Chapter 10.)

CONCEPTS AND DEFINITIONS

Household and Demographic Information

Upon contacting a household, interviewers proceed with the interview unless there is a clear indication that the case is a definite noninterview. (Chapter 7 discusses the interview process and explains refusals and other types of noninterviews.) When interviewing a household for the first time, interviewers collect information about the housing unit and all individuals who usually live at the address.

Housing unit information. Upon first contact with a housing unit, interviewers collect information on the housing unit's physical address, its mailing address, the year it was constructed, the type of structure (single or multiple family), whether it is renter- or owner-occupied, whether other units or persons are part of the sample unit, whether the housing unit has a telephone and, if so, the telephone number.

Household roster. After collecting or updating the housing unit data, the interviewer either creates or updates a list of all individuals living in the unit and determines whether or not they are members of the household. This list is referred to as the household roster.

Household respondent. One person may provide all of the CPS data for the entire sample unit, provided that the person is a household member 15 years of age or older who is knowledgeable about the household. The person who responds for the household is called the household respondent. Information collected from the household respondent for other members of the household is referred to as proxy response.

Reference person. To create the household roster, the interviewer asks the household respondent to give "the names of all persons living or staying" in the housing unit, and to "start with the name of the person or one of the persons who owns or rents" the unit. The person whose name the interviewer enters on line one (presumably one of the individuals who owns or rents the unit) becomes the reference person. Note that the household respondent and the reference person are not necessarily the same. For example, if you are the household respondent and you give your name "first" when asked to report the household roster, then you also are the reference person. If, on the other hand, you are the household respondent and you give your spouse's name "first" when asked to report the household roster, then your spouse is the reference person. (Sometimes the reference person is referred to as the "householder.")

Household. A household is defined as all individuals (related family members and all unrelated individuals) whose usual place of residence at the time of the interview

is the sample unit. Individuals who are temporarily absent and who have no other usual address are still classified as household members even though they are not present in the household during the survey week. College students comprise the bulk of such absent household members, but persons away on business or vacation are also included. (Not included are persons in institutions or the military.) Once household/nonhousehold membership has been established for all persons on the roster, the interviewer proceeds to collect all other demographic data for household members only.

Relationship to reference person. The interviewer will show a flash card with relationship categories (e.g., spouse, child, grandchild, parent, brother/sister) to the household respondent and ask him/her to report each household member's relationship to the reference person (the person listed on line one). Relationship data also are used to define families, subfamilies, and individuals whose usual place of residence is elsewhere. A family is defined as a group of two or more individuals residing together who are related by birth, marriage, or adoption; all such individuals are considered members of one family. Families are further classified either as married-couple families or as families maintained by women or men without spouses. Subfamilies are defined as families that live in housing units where none of the members of the family are related to the reference person. It is also possible that a household contains unrelated individuals; that is, people who are not living with any relatives. An unrelated individual may be part of a household containing one or more families or other unrelated individuals, may live alone, or may reside in group quarters, such as a rooming house.

Additional demographic information. In addition to asking for relationship data, the interviewer asks for other demographic data for each household member including: birth date, marital status, who a particular person's parent is among the members of the household (entered as the line number of parent) and who a person's spouse is (if it cannot be determined from the relationship data), Armed Forces status, level of education, race, ethnicity, nativity, and social security number (for those 15 years of age or older in selected months). Total household income also is collected. The social security number for each household member 15 years of age or older is collected so that it is possible to match the CPS data to other government data in order to evaluate the accuracy, consistency, and comparability of various statistics derived from CPS data. The following terms are used to define an individual's marital status at the time of the interview: married spouse present, married spouse absent, widowed, divorced, separated, or never married. The term "married spouse present" applies to a husband and wife who both live at the same address, even though one may be temporarily absent due to business, vacation, a visit away from home, a hospital stay, etc. The term "married spouse absent" applies to individuals

who live apart for reasons such as marital problems, as well as husbands and wives who are living apart because one or the other is employed elsewhere, on duty with the Armed Forces, or any other reason. The information collected during the interview is used to create three marital status categories: single never married, married spouse present, and other marital status. The latter category includes those who were classified as widowed; divorced; separated; or married, spouse absent.

Starting in January 1992, educational attainment for each person in the household age 15 or older was obtained through a question asking about the highest grade or degree completed. In January 1996, additional questions were added for several educational attainment categories to ascertain the total number of years of school or credit years completed. (Prior to January 1992, the questions referred to the highest grade or year of school each person had attended and, then, if they had completed that grade or year of school.)

The household respondent identifies his/her own race as well as that of other household members by selecting a category from a flash card presented by the interviewer. The categories on the flash card are: White, Black, American Indian, Aleut, Eskimo, Asian or Pacific Islander, or Other. The "nativity" items ask about a person's country of birth, as well as that of the individual's parents, and whether they are American citizens and if so, whether by birth or naturalization. Persons born outside of the 50 states also are asked date of immigration to the United States. The nativity items were added in April of 1993.

The additional demographic data collected are used to classify individuals into racial groups, education categories, and groups based on ethnic origin. Currently, the racial categories for CPS data are White, Black, and Other, where the Other group includes American Indians, Alaskan Natives, and Asians and Pacific Islanders. The data about individuals' ethnic origin primarily are collected for the purpose of identifying Hispanics,[1] where Hispanics are those who identify their origin or descent as Mexican, Puerto Rican, Cuban, Central or South American, or some other Hispanic descent. It should be noted that race and ethnicity are distinct categories. Thus, individuals of Hispanic origin may be of any race.

Labor Force Information

Labor force information is obtained after the household and demographic information has been collected. One of the primary purposes of the labor force information is to classify individuals as employed, unemployed, or not in the labor force. Other information collected includes hours worked, occupation, and industry and related aspects of the working population. It should be noted that the major labor force categories are defined hierarchically and, thus,

---

[1]Hispanics may be of any race.

are mutually exclusive. Employed supersedes unemployed which supersedes not in the labor force. For example, individuals who are classified as employed, even if they worked less than full time, are not asked the questions about having looked for work, and hence cannot be classified as unemployed. Similarly, an individual who is classified as unemployed is not asked the questions used to determine one's primary nonlabor market activity. For instance, retired persons who are currently working are classified as employed even though they have retired from previous jobs. Consequently, they are not asked the questions about their previous employment nor can they be classified as retired. The current concepts and definitions underlying the collection and estimate of the labor force data are presented below.

Reference week. The CPS labor force questions ask about labor market activities for 1 week each month. This week is referred to as the "reference week." The reference week is defined as the 7-day period, Sunday through Saturday, that includes the 12th of the month.

Civilian noninstitutional population. In the CPS, labor force data are restricted to persons 16 years of age and older, who currently reside in 1 of the 50 states or the District of Columbia, who do not reside in institutions (e.g., penal and mental facilities, homes for the aged), and who are not on active duty in the Armed Forces.

Employed persons. Employed persons are those who, during the reference week (a) did any work at all (for at least 1 hour) as paid employees; worked in their own businesses, professions, or on their own farms; or worked 15 hours or more as unpaid workers in an enterprise operated by a family member or (b) were not working, but who had a job or business from which they were temporarily absent because of vacation, illness, bad weather, childcare problems, maternity or paternity leave, labor-management dispute, job training, or other family or personal reasons whether or not they were paid for the time off or were seeking other jobs. Each employed person is counted only once, even if he or she holds more than one job. (See the discussion of multiple jobholders below.)

Employed citizens of foreign countries who are temporarily in the United States but not living on the premises of an embassy are included. Excluded are persons whose only activity consisted of work around their own house (painting, repairing, cleaning, or other home-related housework) or volunteer work for religious, charitable, or other organizations.

The initial survey question, asked only once for each household, inquires whether anyone in the household has a business or a farm. Subsequent questions are asked for each household member to determine whether any of them did any work for pay (or for profit if there is a household business) during the reference week. If no work for pay or profit was performed and a family business exists, respondents are asked whether they did any unpaid work in the family business or farm.

Multiple jobholders. These are employed persons who, during the reference week, had either two or more jobs as wage and salary workers; were self-employed and also held one or more wage and salary jobs; or worked as unpaid family workers and also held one or more wage and salary jobs. A person employed only in private households (cleaner, gardener, babysitter, etc.) who worked for two or more employers during the reference week is not counted as a multiple jobholder since working for several employers is considered an inherent characteristic of private household work. Also excluded are self-employed persons with multiple unincorporated businesses and persons with multiple jobs as unpaid family workers.

Since January 1994, CPS respondents have been asked questions each month to identify multiple jobholders. First, all employed persons are asked "Last week, did you have more than one job (or business, if one exists), including part-time, evening, or weekend work?" Those who answer "yes" are then asked, "Altogether, how many jobs (or businesses) did you have?" Prior to 1994, this information had been available only through periodic CPS supplements.

Hours of work. Beginning with the CPS redesign in January 1994, both actual and usual hours of work have been collected. Prior to the redesign, only actual hours were requested for all employed individuals.

Published data on hours of work relate to the actual number of hours spent "at work" during the reference week. For example, persons who normally work 40 hours a week but were off on the Memorial Day holiday, would be reported as working 32 hours, even though they were paid for the holiday. For persons working in more than one job, the published figures relate to the number of hours worked at all jobs during the week.

Data on persons "at work" exclude employed persons who were absent from their jobs during the entire reference week for reasons such as vacation, illness, or industrial dispute. Data also are available on usual hours worked by all employed persons, including those who were absent from their jobs during the reference week.

At work part time for economic reasons. Sometimes referred to as involuntary part time, this category refers to individuals who gave an economic reason for working 1 to 34 hours during the reference week. Economic reasons include slack work or unfavorable business conditions, inability to find full-time work, and seasonal declines in demand. Those who usually work part time also must indicate that they want and are available to work full time to be classified as being part time for economic reasons.

At work part time for noneconomic reasons. This group includes those persons who usually work part time and were at work 1 to 34 hours during the reference week for a

noneconomic reason. Noneconomic reasons include ill-ness or other medical limitation, childcare problems or other family or personal obligations, school or training, retirement or social security limits on earnings, and being in a job where full-time work is less than 35 hours. The group also includes those who gave an economic reason for usually working 1 to 34 hours but said they do not want to work full time or were unavailable for such work.

Usual full- or part-time status. In order to differentiate a person's normal schedule from his/her activity during the reference week, persons also are classified according to their usual full- or part-time statuses. In this context, full-time workers are those who usually work 35 hours or more (at all jobs combined). This group includes some individuals who worked less than 35 hours in the reference week — for either economic or noneconomic reasons — as well as those who are temporarily absent from work. Similarly, part-time workers are those who usually work less than 35 hours per week (at all jobs), regardless of the number of hours worked in the reference week. This may include some individuals who actually worked more than 34 hours in the reference week, as well as those who were temporarily absent from work. The full-time labor force includes all employed persons who usually work full time and unemployed persons who are either looking for full-time work or are on layoff from full-time jobs. The part-time labor force consists of employed persons who usually work part time and unemployed persons who are seeking or are on layoff from part-time jobs.

Prior to 1994, persons who worked full time during the reference week were not asked about their usual hours. Rather, it was assumed that they usually worked full time, and hence they were classified as full-time workers.

Occupation, industry, and class-of-worker. For the employed, this information applies to the job held in the reference week. A person with two or more jobs is classified accord-ing to the job at which he or she worked the greatest number of hours. The unemployed are classified according to their last jobs. The occupational and industrial classifi-cation of CPS data is based on the coding systems used in the 1990 census. A list of these codes can be found in **Alphabetical Index of Industries and Occupations** (Bureau of the Census, January 1992). The class-of-worker clas-sification assigns workers to one of the following catego-ries: wage and salary workers, self-employed workers, and unpaid family workers. Wage and salary workers are those who receive wages, salary, commissions, tips, or pay in kind from a private employer or from a government unit.

The class-of-worker question also includes separate response categories for "private for profit company" and "nonprofit organization" to further classify private wage and salary workers (this distinction has been in place since January 1994). Self-employed persons are those who work for profit or fees in their own businesses, professions, trades, or farms. Only the unincorporated self-employed

are included in the self-employed category since those whose businesses are incorporated technically are wage and salary workers because they are paid employees of a corporation. Unpaid family workers are persons working without pay for 15 hours a week or more on a farm or in a business operated by a member of the household to whom they are related by birth or marriage.

Occupation, industry, and class-of-worker on second job. The occupation, industry, and class-of-worker informa-tion for individuals' second jobs is collected in order to obtain a more accurate measure of multiple jobholders, to obtain more detailed information about their employment characteristics, and to provide information necessary for comparing estimates of number of employees in the CPS and in BLS's establishment survey (the Current Employ-ment Statistics; for an explanation of this survey see BLS Handbook of Methods, April 1997). For the majority of multiple jobholders, occupation, industry, and class-of-worker data for their second jobs are collected only from a quarter of the sample–those in their fourth or eighth monthly interviews. However, for those classified as "self-employed unincorporated" on their main jobs, class-of-worker of the second job is collected each month. This is done because, according to the official definition, individuals who are self-employed unincorporated on both of their jobs are not considered multiple jobholders.

The questions used to determine whether an individ-ual is employed or not, along with the questions an employed person typically will receive, are presented in Figure 5–1. A copy of the entire questionnaire can be obtained from the Internet. The address for this file is: http://www.bls.census.gov/cps/bqestair.htm.

Earnings. Information on what people earn at their main jobs is collected only for those who are receiving their fourth or eighth monthly interviews. This means that earn-ings questions are asked of only one-fourth of the survey respondents. Respondents are asked to report their usual earnings before taxes and other deductions and to include any overtime pay, commissions, or tips usually received. The term "usual" is as perceived by the respondent. If the respondent asks for a definition of usual, however, inter-viewers are instructed to define the term as more than half the weeks worked during the past 4 or 5 months. Respon-dents may report earnings in the time period they prefer—for example, hourly, weekly, biweekly, monthly, or annually. (Allowing respondents to report in a periodicity with which they were most comfortable was a feature added in the 1994 redesign.) Based on additional information collected during the interview, earnings reported on a basis other than weekly are converted to a weekly amount in later processing. Data are collected for wage and salary workers (excluding the self-employed who respond that their busi-nesses were incorporated). These earnings data are used to construct estimates of the distribution of usual weekly earnings and median earnings. Individuals who do not

report their earnings on an hourly basis are asked if they are, in fact, paid at an hourly rate and if so, what the hourly rate is. The earnings of those who reported hourly and those who are paid at an hourly rate is used to analyze the characteristics of hourly workers, for example, those who are paid the minimum wage.

Unemployed persons. All persons who were not employed during the reference week but were available for work (excluding temporary illness) and had made specific efforts to find employment some time during the 4-week period ending with the reference week are classified as unemployed. Individuals who were waiting to be recalled to a job from which they had been laid off need not have been looking for work to be classified as unemployed.

A relatively minor change was incorporated into the definition of unemployment with the implementation of the January 1994 redesign. Under the former definition, persons who volunteered that they were waiting to start a job within 30 days (a very small group numerically) were classified as unemployed, whether or not they were actively looking for work. Under the new definition, by contrast, people waiting to start a new job must have actively looked for a job within the last 4 weeks in order to be counted as unemployed. Otherwise, they are classified as not in the labor force.

As the definition indicates, there are two ways people may be classified as unemployed. They are either looking for work (job seekers) or they have been temporarily separated from a job (persons on layoff). Job seekers must have engaged in active job search during the above mentioned 4-week period in order to be classified as unemployed. (Active methods are defined as job search methods that have the potential to result in a job offer without any further action on the part of the job seeker.) Examples of active job search methods include going to an employer directly or to a public or private employment agency, seeking assistance from friends or relatives, placing or answering ads, or using some other active method. Examples of the "other active" category include being on a union or professional register, obtaining assistance from a community organization, or waiting at a designated labor pickup point. Passive methods, which do not qualify as job search, include reading (as opposed to answering or placing) "help wanted" ads and taking a job training course. The response categories for active and passive methods are clearly delineated in separately labeled columns on the interviewers' computer screens. Job search methods are identified by the following questions: "Have you been doing anything to find work during the last 4 weeks?" and "What are all of the things you have done to find work during the last 4 weeks?" To ensure that respondents report all of the methods of job search used, interviewers ask "Anything else?" after the initial or a subsequent job search method is reported.

Persons "on layoff" are defined as those who have been separated from a job to which they are waiting to be recalled (i.e., their layoff status is temporary). In order to measure layoffs accurately, the questionnaire determines whether people reported to be on layoff did in fact have an expectation of recall; that is, whether they had been given a specific date to return to work or, at least, had been given an indication that they would be recalled within the next 6 months. As previously mentioned, persons on layoff need not be actively seeking work to be classified as unemployed.

Reason for unemployment. Unemployed individuals are categorized according to their status at the time they became unemployed. The categories are: (1) *Job losers:* a group comprised of (a) persons on temporary layoff from a job to which they expect to be recalled and (b) permanent job losers, whose employment ended involuntarily and who began looking for work; (2) *Job leavers:* persons who quit or otherwise terminated their employment voluntarily and began looking for work; (3) *Persons who completed temporary jobs:* persons who began looking for work after their jobs ended; (4) *Reentrants:* persons who previously worked but were out of the labor force prior to beginning their job search; (5) *New entrants:* persons who never worked before and who are entering the labor force for the first time. Each of these five categories of unemployed can be expressed as a proportion of the entire civilian labor force or as a proportion of the total unemployed. Prior to 1994, new entrants were defined as job seekers who had never worked at a full-time job lasting 2 weeks or longer; reentrants were defined as job seekers who had held a full-time job for at least 2 weeks and had then spent some time out of the labor force prior to their most recent period of job search. These definitions have been modified to encompass any type of job, not just a full-time job of at least 2 weeks duration. Thus, new entrants are now defined as job seekers who have never worked at all, and reentrants are job seekers who have worked before but not immediately prior to their current job search.

Duration of unemployment. The duration of unemployment is expressed in weeks. For individuals who are classified as unemployed because they are looking for work, the duration of unemployment is the length of time (through the current reference week) that they have been looking for work. For persons on layoff, the duration of unemployment is the number of full weeks (through the reference week) they have been on layoff.

The questions used to classify an individual as unemployed can be found in Figure 5–1.

Not in the labor force. Included in this group are all persons in the civilian noninstitutional population who are neither employed nor unemployed. Information is collected on their desire for and availability to take a job at the time of the CPS interview, job search activity in the prior year, and reason for not looking in the 4-week period prior to the survey week. This group includes discouraged workers,

defined as persons not in the labor force who want and are available for a job and who have looked for work sometime in the past 12 months (or since the end of their last job if they held one within the past 12 months), but are not currently looking, because they believe there are no jobs available or there are none for which they would qualify. (Specifically, the main reason identified by discouraged workers for not recently looking for work is one of the following: Believes no work available in line of work or area; could not find any work; lacks necessary schooling, training, skills, or experience; employers think too young or too old; or other types of discrimination.)

Data on a larger group of persons outside the labor force, one that includes discouraged workers as well as persons who desire work but give other reasons for not searching (such as childcare problems, family responsibilities, school, or transportation problems) are also published regularly. This group is made up of persons who want a job, are available for work, and have looked for work within the past year. This group is generally described as having some marginal attachment to the labor force.

Prior to January 1994, questions about the desire for work among those who were not in the labor force were asked only of a quarter of the sample. Since 1994, these questions have been asked of the full CPS sample. Consequently since 1994, estimates of the number of discouraged workers as well as those with a marginal attachment to the labor force are published monthly rather than just quarterly.

Additional questions relating to individuals' job histories and whether they intend to seek work continue to be asked only of persons not in the labor force who are in the sample for either their fourth or eighth month. Data based on these questions are tabulated only on a quarterly basis.

Estimates of the number of employed and unemployed are used to construct a variety of measures. These measures include:

- **Labor force:** The labor force consists of all persons 16 years of age or older classified as employed or unemployed in accordance with the criteria described above.

- **Unemployment rate:** The unemployment rate represents the number of unemployed as a percentage of the labor force.

- **Labor force participation rate:** The labor force participation rate is the proportion of the age-eligible population that is in the labor force.

- **Employment-population ratio:** The employment-population ratio represents the proportion of the age-eligible population that is employed.

Figure 5–1. **Questions for Employed and Unemployed**

1. Does anyone in this household have a business or a farm?

2. LAST WEEK, did you do ANY work for (either) pay (or profit)?

   Parenthetical filled in if there is a business or farm in the household. If 1 is "yes" and 2 is "no," ask 3. If 1 is "no" and 2 is "no," ask 4.

3. LAST WEEK, did you do any unpaid work in the family business or farm?

   If 2 and 3 are both "no," ask 4.

4. LAST WEEK, (in addition to the business,) did you have a job, either full or part time? Include any job from which you were temporarily absent.

   Parenthetical filled in if there is a business or farm in the household.

   If 4 is "no," ask 5.

5. LAST WEEK, were you on layoff from a job?

   If 5 is "yes," ask 6. If 5 is "no," ask 8.

6. Has your employer given you a date to return to work?

   If "no," ask 7.

7. Have you been given any indication that you will be recalled to work within the next 6 months?

   If "no," ask 8.

8. Have you been doing anything to find work during the last 4 weeks?

   If "yes," ask 9.

9. What are all of the things you have done to find work during the last 4 weeks?

Individuals are classified as employed if they say "yes" to questions 2, 3 (and work 15 hours or more in the reference week or receive profits from the business/farm), or 4.

Individuals who are available to work are classified as unemployed if they say "yes" to 5 and either 6 or 7, or if they say "yes" to 8 and provide a job search method that could have brought them into contact with a potential employer in 9.

REFERENCES

U.S. Department of Commerce, U.S. Census Bureau (1992),**Alphabetical Index of Industries and Occupations,** January 1992.

U.S. Department of Labor, Bureau of Labor Statistics (1997), **BLS Handbook of Methods,** Bulletin 2490, April 1997.

Chapter 6.
Design of the Current Population Survey Instrument

INTRODUCTION

Chapter 5 describes the concepts and definitions underpinning the Current Population Survey (CPS) data collection instrument. The current survey instrument is the result of an 8-year research and development effort to redesign the data collection process and to implement previously recommended changes in the underlying labor force concepts. The changes described here were introduced in January 1994. Of particular interest in the redesign was the wording of the labor force portion of the questionnaire and the data collection methods. For virtually every labor force concept the current questionnaire wording is different from what was used previously. Data collection has also been redesigned so that the instrument is fully automated and is administered either on a laptop computer or from a centralized telephone facility.

MOTIVATION FOR THE REDESIGN OF THE QUESTIONNAIRE COLLECTING OF LABOR FORCE DATA

The CPS produces some of the most important data used to develop economic and social policy in the United States. Although the U.S. economy and society have undergone major shifts in recent decades, the survey questionnaire had remained unchanged since 1967. The growth in the number of service-sector jobs and the decline in the number of factory jobs have been two key developments. Other changes include the more prominent role of women in the workforce and the growing popularity of alternative work schedules. The 1994 revisions were designed to accommodate these changes. At the same time, the redesign took advantage of major advances in survey research methods and data collection technology. Furthermore, recommendations for changes in the CPS had been proposed in the late 1970s and 1980s, primarily by the Presidentially-appointed National Commission on Employment and Unemployment Statistics (commonly referred to as the Levitan Commission). No changes were implemented at that time, however, due to the lack of funding for a large overlap sample necessary to assess the effect of the redesign. In the mid-1980s, funding for an overlap became available. Spurred by all of these developments, it was decided to redesign the CPS questionnaire.

OBJECTIVES OF THE REDESIGN

There were five main objectives in redesigning the CPS questionnaire: (1) to better operationalize existing definitions and reduce reliance on volunteered responses; (2) to reduce the potential for response error in the questionnaire-respondent-interviewer interaction and, hence, improve measurement of CPS concepts; (3) to implement minor definitional changes within the labor force classifications; (4) to expand the labor force data available and improve longitudinal measures; and (5) to exploit the capabilities of computer assisted interviewing for improving data quality and reducing respondent burden (see Copeland and Rothgeb (1990) for a fuller discussion).

Enhanced Accuracy

In redesigning the CPS questionnaire, the BLS and Census Bureau attempted to develop questions that would lessen the potential for response error. Among the approaches used were: (1) shorter, clearer question wording; (2) splitting complex questions into two or more separate questions; (3) building concept definitions into question wording; (4) reducing reliance on volunteered information; (5) explicit and implicit strategies for the respondent to provide numeric data on hours, earnings, etc.; and (6) the use of revised precoded response categories for open-ended questions (Copeland and Rothgeb, 1990.)

Definitional Changes

The labor force definitions used in the CPS have undergone only minor modifications since the survey's inception in 1940, and with only one exception, the definitional changes and refinements made in 1994 were uniformly small. The one major definitional change dealt with the concept of discouraged workers; that is, persons outside the labor force who are not looking for work because they believed that there are no jobs available for them. As was noted in Chapter 5, discouraged workers are similar to the unemployed in that they are not working and want a job. Since they are not conducting an active job search, however, they do not satisfy a key element necessary to be classified as unemployed. The former measurement of discouraged workers was criticized by the Levitan Commission as too arbitrary and subjective. It was deemed arbitrary that assumptions about a person's availability for work were made from responses to a question on why the respondent was not currently looking for work. It was considered too subjective that the measurement was based on a person's stated desire for a job regardless of whether the individual had ever looked for work. A new, more

precise measurement of discouraged workers was introduced that specifically required that a person had searched for a job during the prior 12 months and was available for work. The new questions also enable estimation of the number of persons outside the labor force who, although they cannot be precisely defined as discouraged, satisfy many of the same criteria as discouraged workers and thus show some marginal attachment to the labor force.

Other minor changes were made to fine-tune the definitions of unemployment, categories of unemployed persons, and persons who were employed part time for economic reasons.

New Labor Force Information Introduced

With the revised questionnaire, several types of labor force data became available regularly for the first time. For example, information is now available each month on employed persons who have more than one job. Also, by collecting information on the number of hours multiple jobholders work on their main job and secondary jobs separately, estimates of the number of workers who combined two or more part-time jobs into a full-time workweek, and the number of full- and part-time jobs in the economy can be made. The inclusion of the multiple jobholding question also improves the accuracy of answers to the questions on hours worked and facilitates comparisons of employment estimates from the CPS with those from the Current Employment Statistics program, the survey of nonfarm business establishments (for a discussion of CES survey see **BLS Handbook of Methods,** Bureau of Labor Statistics, April 1997). In addition, beginning in 1994, monthly data on the number of hours usually worked per week and data on the number of discouraged workers are collected from the entire CPS sample rather than from the one-quarter sample of respondents in their fourth or eighth monthly interviews.

Computer Technology

A key feature of the redesigned CPS is that the new questionnaire was designed for a computer assisted interview. Prior to the redesign, CPS data were primarily collected using a paper-and-pencil form. In an automated environment, most interviewers now use laptop computers on which the questionnaire has been programmed. This mode of data collection is known as computer assisted personal interviewing (CAPI). Interviewers ask the survey questions as they appear on the screen of the laptop and then type the responses directly into the computer. A portion of sample households—currently about 18 percent—is interviewed via computer assisted telephone interviewing (CATI) from three centralized telephone centers located in Hagerstown, MD; Tucson, AZ; and Jeffersonville, IN.

Automated data collection methods allow greater flexibility in questionnaire designing than paper-and-pencil data collection methods. Complicated skips, respondent-specific question wording, and carry-over of data from one interview to the next are all possible with an automated environment. For example, automated data collection permits capabilities such as (1) the use of dependent interviewing, that is the carrying over of information from the previous month — for industry, occupation, and duration of unemployment data and (2) the use of respondent-specific question wording based on the person's name, age, and sex, answers to prior questions, household characteristics, etc. Furthermore, by automatically bringing up the next question on the interviewer's screen, computerization reduces the probability of an interviewer asking the wrong set of questions. The computerized questionnaire also permits the inclusion of several built-in editing features, including automatic checks for internal consistency, checking for unlikely responses, and verification of answers. With these built-in editing features, errors can be caught and corrected during the interview itself.

Evaluation and Selection of Revised Questions

Planning for the revised CPS questionnaire began in 1986, when BLS and the Census Bureau convened a task force to identify areas for improvement. Studies employing methods from the cognitive sciences were conducted to test possible solutions to the problems identified. These studies included interviewer focus groups, respondent focus groups, respondent debriefings, a test of interviewers' knowledge of concepts, in-depth cognitive laboratory interviews, response categorization research, and a study of respondents' comprehension of alternative versions of labor force questions (Campanelli, Martin, and Rothgeb, 1991; Edwards, Levine, and Cohany, 1989; Fracasso, 1989; Gaertner, Cantor, and Gay,1989; Martin, 1987; Palmisano, 1989).

In addition to the qualitative research mentioned above, the revised questionnaire, developed jointly by Census Bureau and BLS staff, used information collected in a large two-phase test of question wording. During Phase I, two alternative questionnaires were tested with the then official questionnaire as the control. During Phase II, one alternative questionnaire was tested with the control. The questionnaires were tested using computer assisted telephone interviewing and a random digit dialing sample (CATI/RDD). During these tests, interviews were conducted from the centralized telephone interviewing facilities of the Census Bureau.

Both quantitative and qualitative information was used in the two phases to select questions, identify problems, and suggestion solutions. Analyses were based on information from item response distributions, respondent and interviewer debriefing data, and behavior coding of interviewer/respondent interactions.

Item Response Analysis

The primary use of item response analysis was to determine whether different questionnaires produce different response patterns, which may, in turn, have affected

the labor force estimates. Unedited data were used for this analysis. Statistical tests were conducted to ascertain whether response patterns for different questionnaire versions differed significantly. The statistical tests were adjusted to take into consideration the use of a nonrandom clustered sample, repeated measures over time, and multiple persons in a household.

Response distributions were analyzed for all items on the questionnaires. The response distribution analysis indicated the degree to which new measurement processes produced different patterns of responses. Data gathered using the other methods outlined above also aided in the interpretation of the response differences observed. (Response distributions were calculated on the basis of persons who responded to the item, excluding those for whom a ''don't know'' or ''refused'' was obtained.)

Respondent Debriefings

At the end of the interview, respondent debriefing questions were administered to a sample of respondents to measure respondent comprehension and response formulation. Question-specific probes were used to ascertain whether certain words, phrases, or concepts were understood by respondents in the manner intended. (Esposito et al., 1992.). From these data, indicators of how respondents interpret and answer the questions and some measures of response accuracy were obtained.

The debriefing questions were tailored to the respondent and depended on the path the interview had taken. Two forms of respondent debriefing questions were administered— probing questions and vignette classification. For example, those who did not indicate that they had done any work in the main survey were asked the direct probe ''LAST WEEK did you do any work at all, even for as little as 1 hour?'' An example of the vignettes respondents received is ''Last week, Amy spent 20 hours at home doing the accounting for her husband's business. She did not receive a paycheck.'' Individuals were asked to classify the person in the vignette as working or not working based on the wording of the question they received in the main survey (e.g., ''Would you report her as working last week not counting work around the house?'' if the respondent received the unrevised questionnaire or ''Would you report her as working for pay or profit last week?'' if the respondent received the current, revised questionnaire (Martin and Polivka, 1995).

Behavior Coding

Behavior coding entails monitoring or audiotaping interviews, and recording significant interviewer and respondent behaviors (e.g., minor/major changes in question wording, probing behavior, inadequate answers, requests for clarification). During early stages of testing, behavior coding data were useful in identifying problems with proposed questions. For example, if interviewers frequently

reword a question, this may indicate that the question was too difficult to ask as worded; respondents' requests for clarification may indicate that they were experiencing comprehension difficulties; and interruptions by respondents may indicate that a question was too lengthy (Esposito et al., 1992).

During later stages of testing, the objective of behavior coding was to determine whether the revised questionnaire improved the quality of interviewer/respondent interactions as measured by accurate reading of the questions and adequate responses by respondents. Additionally, results from behavior coding helped identify areas of the questionnaire that would benefit from enhancements to interviewer training.

Interviewer Debriefings

The primary objective of interviewer debriefing was to identify areas of the revised questionnaire or interviewer procedures that were problematic for interviewers or respondents. The information collected during the debriefings was useful in identifying which questions needed revision. The information was also useful in modifying initial interviewer training and the interviewer manual. A secondary objective of interviewer debriefing was to obtain information about the questionnaire, interviewer behavior, or respondent behavior that may help explain differences observed in the labor force estimates from the different measurement processes.

Two different techniques were used to debrief interviewers. The first was the use of focus groups at the centralized telephone interviewing facilities and in geographically dispersed regional offices. The focus groups were conducted after interviewers had at least 3 to 4 months experience using the revised CPS instrument. Approximately 8 to 10 interviewers were selected for each focus group. Interviewers were selected to represent different levels of experience and ability.

The second technique was the use of a self-administered standardized interviewer debriefing questionnaire. Once problematic areas of the revised questionnaire were identified through the focus groups, a standardized debriefing questionnaire was developed and administered to all interviewers.

HIGHLIGHTS OF THE QUESTIONNAIRE REVISION

A copy of the questionnaire can be obtained from the Internet. The address for this file is:

http://www.bls.census.gov/cps/bqestair.htm

General

Definition of reference week. In the interviewer debriefings that were conducted in 13 different geographic areas during 1988, interviewers reported that the current question 19 (Q19, major activity question) ''What were you

doing most of LAST WEEK, working or something else?" was unwieldy and sometimes misunderstood by respondents. In addition to not always understanding the intent of the question, respondents were unsure what was meant by the time period "last week" (BLS, 1988). A respondent debriefing conducted in 1988 found that only 17 percent of respondents had definitions of "last week" that matched the CPS definition of Sunday through Saturday of the reference week. The majority (54 percent) of respondents defined "last week" as Monday through Friday (Campanelli et al., 1991).

In the revised questionnaire, an introductory statement was added with the reference period clearly stated. The introductory statement reads as follows: "I am going to ask a few questions about work-related activities LAST WEEK. By last week I mean the week beginning on Sunday, August 9 and ending Saturday, August 15." This statement makes the reference period more explicit to respondents. Additionally, the former Q19 has been deleted from the questionnaire. In the past, Q19 had served as a preamble to the labor force questions, but in the revised questionnaire the survey content is defined in the introductory statement, which also defines the reference week.

Direct question on presence of business. The definition of employed persons includes those who work without pay for at least 15 hours per week in a family business. In the former questionnaire, there was no direct question on the presence of a business in the household. Such a question is included in the revised questionnaire. This question is asked only once for the entire household prior to the labor force questions. The question reads as follows: "Does anyone in this household have a business or a farm?" With this question it can be determined whether a business exists and who in the household owns the business. The primary purpose of this question is to screen for households that may have unpaid family workers, not to obtain an estimate of household businesses. (See Rothgeb et al. (1992), Copeland and Rothgeb (1990), and Martin (1987) for a fuller discussion of the need for a direct question on presence of a business.)

For households that have a family business, direct questions are asked about unpaid work in the family business of all persons who were not reported as working last week. BLS produces monthly estimates of unpaid family workers who work 15 or more hours per week.

Employment Related Revisions

Revised "At Work" question. Having a direct question on the presence of a family business not only improved the estimates of unpaid family workers, but also permitted a revision of the "at work" question. In the former questionnaire, the "at work" question read: "LAST WEEK, did you do any work at all, not counting work around the house?" In the revised questionnaire, the wording reads, "LAST WEEK

did you do ANY work for (either) pay (or profit)?" (The parentheticals in the question are read only when there is a business or farm in the household.) The revised wording "work for pay (or profit)" better captures the concept of work that BLS is attempting to measure. (See Martin (1987) or Martin and Polivka (1995) for a fuller discussion of problems with the concept of "work.")

Direct question on multiple jobholding. In the former questionnaire, the actual hours question read: "How many hours did you work last week at all jobs?" During the interviewer debriefings conducted in 1988, it was reported that respondents do not always hear the last phrase "at all jobs." Some respondents who work at two jobs may have only reported hours for one job (BLS, 1988). In the revised questionnaire, a question is included at the beginning of the hours series to determine whether or not the person is a multiple jobholder. A followup question also asks for the number of jobs the multiple jobholder has. Multiple jobholders are asked about their hours on their main job and other job(s) separately to avoid the problem of multiple jobholders not hearing the phrase "at all jobs." These new questions also allow monthly estimates of multiple jobholders to be produced.

Hours series. The old question on "hours worked" read: "How many hours did you work last week at all jobs?" If a person reported 35-48 hours worked, additional followup probes were asked to determine whether the person worked any extra hours or took any time off. Interviewers were instructed to correct the original report of actual hours, if necessary, based on responses to the probes. The hours data are important because they are used to determine the sizes of the full-time and part-time labor forces. It is unknown whether respondents reported exact actual hours, usual hours, or some approximation of actual hours.

In the revised questionnaire, a revised hours series was adopted. An anchor-recall estimation strategy was used to obtain a better measure of actual hours and to address the issue of work schedules more completely. For multiple jobholders, it also provides separate data on hours worked at a main job and other jobs. The revised questionnaire first asks about the number of hours a person usually works at the job. Then, separate questions are asked to determine whether a person worked extra hours, or fewer hours, and finally a question is asked on the number of actual hours worked last week. It also should be noted that the new hours series allows monthly estimates of usual hours worked to be produced for all employed persons. In the former questionnaire, usual hours were obtained only in the outgoing rotation for employed private wage and salary workers and were available only on a quarterly basis.

Industry and occupation – dependent interviewing. Prior to the revision, CPS industry and occupation (I&O) data were not always consistent from month-to-month for the same person in the same job. These inconsistencies

arose, in part, because the household respondent frequently varies from 1 month to the next. Furthermore, it is sometimes difficult for a respondent to describe an occupation consistently from month-to-month. Moreover, distinctions at the three-digit occupation and industry level, that is, at the most detailed classification level, can be very subtle.

To obtain more consistent data and make full use of the automated interviewing environment, dependent interviewing for the I&O questions was implemented in the revised questionnaire for month-in-sample 2-4 households and month-in-sample 6-8 households. Dependent interviewing uses information collected during the previous month's interview in the current month's interview. (Different variations of dependent interviewing were evaluated during testing. See Rothgeb et al. (1991) for more detail.)

In the revised CPS, respondents are provided with the name of their employer as of the previous month and asked if they still work for that employer. If they answer "no," respondents are asked the independent questions on industry and occupation.

If they answer "yes," respondents are asked "Have the usual activities and duties of your job changed since last month?" If individuals say "yes," their duties have changed, these individuals are then asked the independent questions on occupation, activities or duties, and class-of-worker. If their duties have not changed, individuals are asked to verify the previous month's description through the question "Last month, you were reported as (previous month's occupation or kind of work performed) and your usual activities were (previous month's duties). Is this an accurate description of your current job?"

If they answer "yes," the previous month's occupation and class-of-worker are brought forward and no coding is required. If they answer "no," persons are asked the independent questions on occupation activities and duties and class-of-worker. This redesign permits a direct inquiry about job change before the previous month's information is provided to the respondent.

Earnings. The earnings series in the revised questionnaire is considerably different from that in the former questionnaire. In the former questionnaire, persons were asked whether they were paid by the hour, and if so, what the hourly wage was. All wage and salary workers were then asked for their usual weekly earnings. In the former version, earnings could be reported as weekly figures only, even though that may not have been the easiest way for the respondent to recall and report earnings. Data from early tests indicated that a small proportion (14 percent) (n=853) of nonhourly wage workers were paid at a weekly rate and less than 25 percent (n=1623) of nonhourly wage workers found it easiest to report earnings as a weekly amount.

In the revised questionnaire, the earnings series is designed to first request the periodicity for which the respondent finds it easiest to report earnings and then request an earnings amount in the specified periodicity, as displayed below. The wording of questions requesting an earnings amount is tailored to the periodicity identified earlier by the respondent. (Because data on weekly earnings are published quarterly by BLS, earnings data provided by respondents in periodicities other than weekly are converted to a weekly earnings estimate later during processing operations.)

Revised Earnings Series (Selected items)

1. For your (MAIN) job, what is the easiest way for you to report your total earnings BEFORE taxes or other deductions: hourly, weekly, annually, or on some other basis?

2. Do you usually receive overtime pay, tips, or commissions (at your MAIN job)?

3. (Including overtime pay, tips and commissions,) What are your usual (weekly, monthly, annual, etc.) earnings on this job, before taxes or other deductions?

As can be seen from the revised questions presented above, other revisions to the earnings series include a specific question to determine whether a person usually receives overtime pay, tips, or commissions. If so, a preamble precedes the earnings questions that reminds respondents to include overtime pay, tips, and commissions when reporting earnings. If a respondent reports that it is easiest to report earnings on an hourly basis, then a separate question is asked regarding the amount of overtime pay, tips and commissions usually received, if applicable.

An additional question is asked of persons who do not report that it is easiest to report their earnings hourly. The question determines whether they are paid at an hourly rate and is displayed below. This information, which allows studies of the effect of the minimum wage, is used to identify hourly wage workers.

"Even though you told me it is easier to report your earnings annually, are you PAID AT AN HOURLY RATE on this job?"

Unemployment Related Revisions

Persons on layoff - direct question. Previous research (Rothgeb, 1982; Palmisano, 1989) demonstrated that the former question on layoff status—"Did you have a job or business from which you were temporarily absent or on layoff LAST WEEK?"—was long, awkwardly worded, and frequently misunderstood by respondents. Some respondents heard only part of the question, while others thought that they were being asked whether they had a business.

In an effort to reduce response error, the revised questionnaire includes two separate direct questions about layoff and temporary absences. The layoff question is: "LAST WEEK, were you on layoff from a job?" Questions asked later screen out those persons who do not meet the criteria for layoff status.

Persons on layoff—expectation of recall. The official definition of layoff includes the criterion of an expectation of being recalled to the job. In the former questionnaire, persons reported to be on layoff were never directly asked whether they expected to be recalled. In an effort to better capture the existing definition, persons reported to be on layoff in the revised questionnaire are asked ''Has your employer given you a date to return to work?'' Persons who respond that their employers have not given them a date to return are asked ''Have you been given any indication that you will be recalled to work within the next 6 months?'' If the response is positive, their availability is determined by the question, ''Could you have returned to work LAST WEEK if you had been recalled?'' Persons who do not meet the criteria for layoff are asked the job search questions so they still have an opportunity to be classified as unemployed.

Job search methods. The concept of unemployment requires, among other criteria, an active job search during the past 4 weeks. In the former questionnaire, the following question was asked to determine whether a person conducted an active job search. ''What has ... been doing in the last 4 weeks to find work?'' checked with—

- public employment agency

- private employment agency

- employer directly

- friends and relatives

- placed or answered ads

- nothing

- other

Interviewers were instructed to code all passive job search methods into the ''nothing'' category. This included such activities as looking at newspaper ads, attending job training courses, and practicing typing. Only active job search methods for which no appropriate response category exists were to be coded as ''other.''

In the revised questionnaire, several additional response categories were added and the response options re-ordered and reformatted to more clearly represent the distinction between active job search methods and passive methods. The revisions to the job search methods question grew out of concern that interviewers were confused by the precoded response categories. This was evident even before the analysis of the CATI/RDD test. Martin (1987) conducted an examination of verbatim entries for the ''other'' category and found that many of the ''other'' responses should have been included in the ''nothing'' category instead. The analysis also revealed responses coded as ''other'' that were too vague to determine whether or not an active job search method had been undertaken.

Fracasso (1989) also concluded that the current set of response categories was not adequate for accurate classification of active and passive job search methods.

During development of the revised questionnaire, two additional passive categories were included: (1) ''looked at ads'' and (2) ''attended job training programs/courses,'' and two additional active categories were included: (1) ''contacted school/university employment center'' and (2) ''checked union/ professional registers.'' Later research also demonstrated that interviewers had difficulty coding relatively common responses such as ''sent out resumes'' and ''went on interviews''; thus the response categories were further expanded to reflect these common job search methods.

Duration of job search and layoff. The duration of unemployment is an important labor market indicator published monthly by BLS. In the former questionnaire, this information was collected by the question: ''How many weeks have you been looking for work?'' This wording forced people to report in a periodicity that may not have been meaningful to them, especially for the longer-term unemployed. Also, asking for the number of weeks (rather than months) may have led respondents to underestimate the duration. In the revised questionnaire, the question reads: ''As of the end of LAST WEEK, how long had you been looking for work?'' Respondents can select the periodicity themselves and interviewers are able to record the duration in weeks, months, or years.

To avoid clustering of answers around whole months, the revised questionnaire also asks persons who report duration in whole months (between 1 and 4 months) a followup question to obtain an estimated duration in weeks: ''We would like to have that in weeks, if possible. Exactly how many weeks had you been looking for work?'' The purpose of this is to lead people to report the exact number of weeks instead of multiplying their monthly estimates by four as was done in an earlier test and may have been done in the former questionnaire.

As mentioned earlier, the CATI/CAPI technology makes it possible to automatically update duration of job search and layoff for persons who are unemployed in consecutive months. For persons reported to be looking for work for 2 consecutive months or longer, the previous month's duration is updated without re-asking the duration questions. For persons on layoff for at least 2 consecutive months, the duration of layoff is also automatically updated. This revision was made to reduce respondent burden and enhance the longitudinal capability of the CPS. This revision also will produce more consistent month-to-month estimates of duration. Previous research indicates that only about 25 percent of those unemployed in consecutive months who received the former questionnaire (where duration was collected independently each month) increased their reported durations by 4 weeks plus or minus a week. (Polivka and Rothgeb, 1993; Polivka and Miller, 1995). A very small bias is introduced when a person has a brief (less than 3 or 4

weeks) period of employment in between surveys. However, testing revealed that only 3.2 percent of those who had been looking for work in consecutive months said that they had worked in the interlude between the surveys. Furthermore, of those who had worked, none indicated that they had worked for 2 weeks or more.

Revisions Included in "Not in the Labor Force" Related Questions

Response options of retired, disabled, and unable to work at key labor force items. In the former questionnaire, when individuals reported they were retired in response to any of the labor force items, the interviewer was required to continue asking whether they worked last week, were absent from a job, were looking for work, and, in the outgoing rotation, when they last worked and their job histories. Interviewers commented that elderly respondents frequently complained that they had to respond to questions that seemed to have no relevance to their own situations.

In an attempt to reduce respondent burden, a response category of "retired" was added to each of the key labor force status questions in the revised questionnaire. If individuals 50 years of age or older volunteer that they are retired, they are immediately asked a question inquiring whether they want a job. If they indicate that they want to work, they are then asked questions about looking for work and the interview proceeds as usual. If they do not want to work, the interview is concluded and they are classified as not in the labor force - retired. (If they are in the outgoing rotation, an additional question is asked to determine whether they worked within the last 12 months. If so, the industry and occupation questions are asked about the last job held.)

A similar change has been made in the revised questionnaire to reduce the burden for individuals reported to be "unable to work" or "disabled." (Individuals who may be "unable to work" for a temporary period of time may not consider themselves as "disabled" so both response options are provided.) If a person is reported to be "disabled" or "unable to work" at any of the key labor force classification items, a followup question is asked to determine whether he/she can do any gainful work during the next 6 months. Different versions of the followup probe are used depending on whether the person is disabled or unable to work.

Dependent interviewing for persons reported to be retired, disabled, or unable to work. The revised questionnaire also is designed to use dependent interviewing for persons reported to be retired, disabled, or unable to work. An automated questionnaire increases the ease with which information from the previous month's interview can be used during the current month's interview.

Once it is reported that the person did not work during the current month's reference week, the previous month's status of retired (if a person is 50 years of age or older),

disabled, or unable to work is verified, and the regular series of labor force questions is not asked. This revision reduces respondent and interviewer burden.

Discouraged workers. The implementation of the Levitan Commission's recommendations on discouraged workers resulted in one of the major definitional changes in the 1994 redesign. The Levitan Commission criticized the former definition because it was based on a subjective desire for work and questionable inferences about an individual's availability to take a job. As a result of the redesign, two requirements were added: For persons to qualify as discouraged, they must have engaged in some job search within the past year (or since they last worked if they worked within the past year), and they must currently be available to take a job. (Formerly, availability was inferred from responses to other questions; now there is a direct question.)

Data on a larger group of persons outside the labor force (one that includes discouraged workers as well as persons who desire work, but give other reasons for not searching, such as child care problems, family responsibilities, school, or transportation problems) also are published regularly. This group is made up of persons who want a job, are available for work, and have looked for work within the past year. This group is generally described as having some marginal attachment to the labor force. Also beginning in 1994, questions on this subject are asked of the full CPS sample rather than limited to a quarter of the sample, permitting estimates of the number of discouraged workers to be published monthly rather than quarterly.

From data available during the tests of the revised questionnaire, it is clear that improvements in data quality for labor force classification have been obtained as a result of the redesign of the CPS questionnaire, and in general, measurement error has been reduced. Data from respondent debriefings, interviewer debriefings, and response analysis demonstrated that the revised questions are more clearly understood by respondents and the potential for labor force misclassification is reduced. Results from these tests formed the basis for the design of the final revised version of the questionnaire. This revised version was tested in a separate year and a half parallel survey prior to implementation as the official survey in January 1994. In addition, from January 1994 through May 1994, the unrevised procedures were used with the parallel survey sample. These parallel surveys were conducted to assess the effect of the redesign on national labor force estimates. Estimates derived from the initial year and a half of the parallel survey indicated that the redesign might increase the unemployment rate by 0.5 percentage points. However, subsequent analysis using the entire parallel survey indicates that the redesign did not have a statistically significant effect on the unemployment rate. (Analysis of the effect of the redesign on the unemployment rate and other labor force estimates can be found in Cohany, Polivka, and

Rothgeb (1994). Analysis of the redesign on the unemployment rate along with a wide variety of other labor force estimates using data from the entire parallel survey can be found in Polivka and Miller (1995).)

CONTINUOUS TESTING AND IMPROVEMENTS OF THE CURRENT POPULATION SURVEY AND ITS SUPPLEMENTS

Experience gained during the redesign of the CPS has demonstrated the importance of testing questions and monitoring data quality. The experience, along with contemporaneous advances in research on questionnaire design, also has helped inform the development of methods for testing new or improved questions for the basic CPS and its periodic supplements (Martin, 1987; Oksenberg, Bischoping, K., Cannell and Kalton, 1991; Campanelli, Martin, and Rothgeb, 1991; Esposito et al., 1992; and Forsyth and Lessler, 1991). Methods to continuously test questions and assess data quality are discussed in Chapter 15. It is important to note, however, that despite the benefits of adding new questions and improving existing ones, changes in the CPS should be approached cautiously and the effects measured and evaluated. When possible, methods to bridge differences caused by changes or techniques to avoid the disruption of historical series should be included in the testing of new or revised questions.

REFERENCES

Bischoping, K. (1989), An Evaluation of Interviewer Debriefings in Survey Pretests, In C. Cannell et al. (eds.), **New Techniques for Pretesting Survey Questions,** Chapter 2.

Campanelli, P.C., Martin, E.A., and Rothgeb, J.M. (1991), The Use of Interviewer Debriefing Studies as a Way to Study Response Error in Survey Data, **The Statistician,** vol. 40, pp. 253-264.

Cohany, S.R., Polivka, A.E., and Rothgeb, J.M., (1994) Revisions in the Current Population Survey Effective January 1994, **Employment and Earnings,** February 1994 vol. 41 no. 2 pp. 1337.

Copeland, K. and Rothgeb, J.M. (1990), Testing Alternative Questionnaires for the Current Population Survey, **Proceedings of the Section on Survey Research Methods,** American Statistical Association, pp. 63-71.

Edwards. S.W., Levine, R., and Cohany, S.R., (1989), Procedures for Validating Reports of Hours Worked and for Classifying Discrepancies Between Reports and Validation Totals, **Proceedings of the Section on Survey Research Methods,** American Statistical Association.

Esposito, J.L., Rothgeb,J.M., Polivka, A.E., Hess, J., and Campanelli, P.C. (1992), Methodologies for Evaluating Survey Questions: Some Lessons From the Redesign of the Current Population Survey, Paper presented at the International Conference on Social Science Methodology, Trento, Italy, June, 1992.

Forsyth, B.H. and Lessler, J.T. (1991), Cognitive Laboratory Methods: A Taxonomy, in P.P Biemer, R.M. Groves, L.E. Lyberg, N.A. Mathiowetz, and S. Sudman (eds.), **Measurement Errors in Surveys,** New York: Wiley, pp. 393-418.

Fracasso, M.P. (1989), Categorization of Responses to the Open-Ended Labor Force Questions in the Current Population Survey, **Proceedings of the Section on Survey Research Methods,** American Statistical Association, pp. 481-485.

Gaertner, G., Cantor, D., and Gay, N. (1989), Tests of Alternative Questions for Measuring Industry and Occupation in the CPS, **Proceedings of the Section on Survey Research Methods,** American Statistical Association.

Martin, E.A., (1987), Some Conceptual Problems in the Current Population Survey, **Proceedings of the Section on Survey Research Methods,** American Statistical Association, pp. 420-424.

Martin, E.A. and Polivka, A.E. (1995), Diagnostics for Redesigning Survey Questionnaires: Measuring Work in the Current Population Survey, **Public Opinion Quarterly,** Winter 1995 vol. 59, no.4, pp. 547-67.

Oksenberg, L., Cannell, C., and Kalton, G. (1991), New Strategies for Pretesting Survey Questions, **Journal of Official Statistics**, vol. 7, no. 3, pp. 349-365.

Palmisano, M. (1989), Respondents Understanding of Key Labor Force Concepts Used in the CPS, **Proceedings of the Section on Survey Research Methods,** American Statistical Association.

Polivka, A.E. and Miller, S.M. (1998),The CPS After the Redesign Refocusing the Economic Lens, **Labor Statistics Measurement Issues,** edited by J. Haltiwanger et al., pp. 249-86.

Polivka, A.E. and Rothgeb, J.M. (1993), Overhauling the Current Population Survey: Redesigning the Questionnaire, **Monthly Labor Review,** September 1993 vol. 116, no. 9, pp. 10-28.

Rothgeb, J.M. (1982), Summary Report of July Followup of the Unemployed, Internal memorandum, December 20, 1982.

Rothgeb, J.M., Polivka, A.E., Creighton, K.P., and Cohany, S.R. (1992), Development of the Proposed Revised Current Population Survey Questionnaire, **Proceedings of the Section on Survey Research Methods,** American Statistical Association.

U.S. Department of Labor, Bureau of Labor Statistics (1988), Response Errors on Labor Force Questions Based on Consultations With Current Population Survey Interviewers in the United States. Paper prepared for the OECD Working Party on Employment and Unemployment Statistics.

U.S. Department of Labor, Bureau of Labor Statistics (1997), **BLS Handbook of Methods,** Bulletin 2490, April 1997.

Chapter 7.
Conducting the Interviews

INTRODUCTION

Each month during interview week, field representatives (FRs) and computer assisted telephone interviewers (CATI) attempt to contact and interview a responsible person living in each sample unit to complete a Current Population Survey (CPS) interview. Typically, the week containing the 19th of the month is the interview week. The week containing the 12th is the reference week (i.e., the week about which the labor force questions are asked). In December, the week containing the 12th is used as interview week, provided the reference week (in this case the week containing the 5th) falls entirely within the month of December. As outlined in Chapter 3, households are in sample for 8 months. Each month, one-eighth of the households are in sample for the first time (month-in-sample 1 (MIS-1)), one-eighth for the second time, etc. Because of this, different types of interviews (due to differing MIS) are conducted by each FR within his/her assignment. An introductory letter is sent to each household in sample prior to its first and fifth month interviews. The letter describes the CPS, announces the forthcoming visit, and provides respondents with information regarding their rights under the Privacy Act, the voluntary nature of the survey, and the guarantees of confidentiality for the information they provide. Figure 7–1 shows the introductory letter sent to sample units in the area administered by the Chicago Regional Office. A personal visit interview is required for all first month-in-sample households. This is due to the fact that the CPS sample is strictly a sample of addresses. The Census Bureau has no way of knowing who the occupants of the sample household are, much less whether the household is occupied or eligible for interview. (Note: In some MIS-1 households, telephone interviews are conducted. This occurs when, during the initial personal contact, the respondent requests a telephone interview.)

NONINTERVIEWS AND HOUSEHOLD ELIGIBILITY

The FR's first task is to establish the eligibility of the sample address for CPS. There are many reasons an address may not be eligible for interview. The address may have been converted to a permanent business, condemned or demolished, or falls outside the boundaries of the segment for which it was selected. Regardless of the reason, such sample addresses are classified as Type C noninterviews. The Type C units have no chance of becoming eligible for the CPS interview in future months, because the condition is considered permanent. These addresses are stricken from the roster of sample addresses and are never visited again with regard to CPS. All households classified as Type C undergo a full supervisory review of the circumstances surrounding the case before such a determination is made final.

Other sample addresses are also ineligible for CPS interview. These are units which are intended for occupancy but are not occupied by any eligible individuals. Reasons for such ineligibility include a vacant housing unit (either for sale or rent), occupied entirely by individuals who are not eligible for a CPS labor force interview (individuals with a usual residence elsewhere (URE), in the Armed Forces, or children under the age of 15), or other reasons why a housing unit is temporarily not occupied. Such units are classified as Type B noninterviews. These noninterviews are considered unavoidable. Type B noninterview units have a chance of becoming eligible for interview in future months, because the condition is considered temporary (e.g., a vacant unit could become occupied). Therefore, Type B units are reassigned to FRs in subsequent months. These sample addresses remain in sample for the entire 8 months that households are eligible for interview. Each succeeding month, an FR visits the unit to determine whether the unit has changed status, and either continues the Type B classification, revises the noninterview classification, or conducts an interview as applicable. Some of these Type B households are found to be eligible for the Housing Vacancy Survey (HVS). See Chapter 11 for a description of the HVS.

Additionally, one final set of households is not interviewed for CPS; these are called Type A households. These are households that the FR has determined as eligible for a CPS interview, but for which no useable data were collected. To be eligible the unit has to be occupied by at least one person eligible for an interview (an individual who is a civilian, at least 15 years old, and does not have a usual residence elsewhere). Even though such households are eligible for interview, they are not interviewed because the household members refuse, are absent during the interviewing period, or are unavailable for other reasons. All Type A cases are subject to full supervisory review before such a determination is made final. Every effort is made to keep such noninterviews to a minimum. All Type A cases remain in sample and are assigned for interview in all succeeding months. Even in cases of confirmed refusals

Figure 7–1. **Introductory Letter**

CPS-263(L) CHICAGO
(11-97)

UNITED STATES DEPARTMENT OF COMMERCE
Bureau of the Census
Washington, DC 20233-0001

OFFICE OF THE DIRECTOR

FROM THE DIRECTOR
BUREAU OF THE CENSUS

You may have read in the newspaper – – or heard on the radio or television – – the official Government figures on total employment and unemployment issued each month. The Census Bureau obtains these figures, as well as information about persons not in the labor force, from the Current Population Survey (CPS). This information, which we collect for the Bureau of Labor Statistics, provides vital up-to-date estimates of the number of persons working, the number who are unemployed, and many other related facts. Occasionally, we ask additional questions on education, health, family income, housing, and other important subjects.

A Census Bureau representative, who will show an official identification card, will call on you during the week in which the 19th of the month falls. The representative will ask questions concerning the ages, employment status, and occupations of the members of your household, as well as other related information. By law, Census Bureau employees hold all information you give in strict confidence. Your answers will be used only for statistical purposes in a manner in which no information about you as an individual can be identified.

We have selected your address and about 48,000 others throughout the United States for this survey. Because this is a sample survey, your answers represent not only yourself and your household, but also hundreds of other households like yours. For this reason, your participation in this voluntary survey is extremely important to ensure the completeness and accuracy of the final results. Although there are no penalties for failure to answer any question, each unanswered question lessens the accuracy of the final data. Your cooperation will be a distinct service to our country.

On the other side of this letter are answers to questions which participants ask most frequently about this survey.

Thank you for your cooperation.

Sincerely,

Martha Farnsworth Riche

Further information
may be obtained from:

Regional Director
Bureau of the Census
2255 Enterprise Dr., Ste. 5501
Westchester, IL   60154-5800

Telephone: 708-562-1350

(cases that still refuse to be interviewed despite supervisory attempts to convert the case), the FR must verify that the same household still resides at that address before submitting a Type A noninterview.

Figure 7–2 shows how the three types of noninterviews are classified and the various reasons for reaching such results. Even if a unit is designated as a noninterview, FRs are responsible for collecting information about the unit. Figure 7–3 lists the main housing unit items that are collected for noninterviews and summarizes each item briefly.

Figure 7–2. **Noninterviews: Types A, B, and C**

Note: See the CPS Interviewing Manual for more details regarding the answer categories under each type of noninterview. See figure 7–3 for a list of the main housing unit items asked for noninterview and a brief description of what each item asks.

TYPE A
1   No one home
2   Temporarily absent
3   Refusal
4   Other occupied

TYPE B
1   Vacant regular
2   Temporarily occupied by persons with usual residence elsewhere
3   Vacant -- storage of household furniture
4   Unfit or to be demolished
5   Under construction, not ready
6   Converted to temporary business or storage
7   Unoccupied tent site or trailer site
8   Permit granted, construction not started
9   Other Type B -- specify

TYPE C
1   Demolished
2   House or trailer moved
3   Outside segment
4   Converted to permanent business or storage
5   Merged
6   Condemned
8   Unused line of listing sheet
9   Other -- specify

INITIAL INTERVIEW

If the unit is not classified as a noninterview, the FR initiates the CPS interview. The FR attempts to interview a knowledgeable adult household member (known as the household respondent). The FRs are trained to ask the questions exactly as worded as they appear on the computer screen. The interview begins with the verification of the unit's address and confirmation of its eligibility for a CPS interview. Part 1 of Figure 7–4 shows the household items asked at the beginning of the interview. Once this is established, the interview moves into the demographic portion of the instrument. The primary task of this portion of the interview is to establish the household's roster (the listing of all household residents at the time of the interview). At this point in the interview, the main concern is to establish an individual's usual place of residence. (These rules are summarized in Figure 7–5.) For all individuals residing in the household without a usual residence elsewhere, a number of personal and family demographic characteristics are collected. Part 1 of Figure 7–6 shows the demographic items asked in MIS-1 households.

These characteristics are the relationship to the reference person (the person who owns or rents the home), parent or spouse pointers (if applicable), age, sex, marital status, educational attainment, veteran's status, current Armed Forces status, race, and origin. As discussed in Figure 7–7, these characteristics are collected in an interactive format that includes a number of consistency edits embedded in the interview itself. The goal is to collect as consistent a set of demographic characteristics as possible. The final steps in this portion of the interview are to verify the accuracy of the roster. To this end, a series of questions is asked to ensure that all household members have been accounted for. Before moving on to the labor force portion of the interview, the FR is prompted to review the roster and all data collected up to this point. A chance is provided for the FR to correct any incorrect or inconsistent information at this time. The instrument then begins the labor force portion of the interview.

In a household's initial interview, a few additional characteristics are collected after completion of the labor force portion of the interview. This information includes a question on family income and data on all household members' countries of birth (along with the country of birth of his/her father and mother) and, for the foreign born, data on year of entry into the United States and citizenship status. See Part 2 of Figure 7–6 for a list of these items.

After completing the household roster, the FR collects the labor force data described in Chapter 6. The labor force data are collected from all civilian adult individuals (age 15 and older) who do not have a usual residence elsewhere. To the greatest extent possible, the FR attempts to collect this information from each eligible individual him/herself. In the interest of timeliness and efficiency, however, a household respondent (any knowledgeable adult household member) generally is used to collect the data. Just over one-half of the CPS labor force data are collected by self-response. The bulk of the remainder is collected by proxy from the household respondent. Additionally, in certain limited situations, collection of the data from a nonhousehold member is allowed. All such cases receive direct supervisory review before the data are accepted into the CPS processing system.

Figure 7–3. **Noninterviews: Main Housing Unit Items Asked for Types A, B, and C**

Note: This list of items is not all inclusive. The list covers only the main data items and does not include related items used to arrive at the final response (e.g., probes and verification screens). See CPS Interviewing Manual for illustrations of the actual instrument screens for all CPS items.

Housing Unit Items for Type A Cases

	Item Name	Item Asks
1	TYPEA	Which specific kind of Type A is the case.
2	ABMAIL	What is the property's mailing address.
3	PROPER	If there is any other building in the property (occupied or vacant).
4	ACCES-scr	If access to the household is direct or through another unit; this item is answered by the interviewer based on observation.
5	LIVQRT	What type of housing unit is it (house/apt., mobile home or trailer, etc.); this item is answered by the interviewer based on observation.
6	INOTES-1	If the interviewer wants to make any notes about the case that might help with the next interview.

Housing Unit Items for Type B Cases

	Item Name	Item Asks
1	TYPEB	Which specific kind of Type B is the case.
2	ABMAIL	What is the property's mailing address.
3	BUILD	If there are any other units (occupied or vacant) in the unit.
4	FLOOR	If there are any occupied or vacant living quarters besides this one on this floor.
5	PROPER	If there is any other building in the property (occupied or vacant).
6	ACCES-scr	If access to the household is direct or through another unit; this item is answered by the interviewer based on observation.
7	LIVQRT	What is the type of housing unit (house/apt., mobile home or trailer, etc.); this item is answered by the interviewer based on observation.
8	SEASON	If the unit is intended for occupancy year round, by migratory workers, or seasonally.
9	BCINFO	What are the name, title, and phone number of contact who provided Type B or C information; or if the information was obtained by interviewer observation.
10	INOTES-1	If the interviewer wants to make any notes about the case that might help with the next interview.

Housing Unit Items for Type C Cases

	Item Name	Item Asks
1	TYPEC	Which specific kind of Type C is the case.
2	PROPER	If there is any other building in the property (occupied or vacant).
3	ACCES-scr	If access to the household is direct or through another unit; this item is answered by the interviewer based on observation.
4	LIVQRT	What type of housing unit is it (house/apt., mobile home or trailer, etc.); this item is answered by the interviewer based on observation.
5	BCINFO	What are the name, title, and phone number of contact who provided Type B or C information; or if the information was obtained by interviewer observation.
6	INOTES-1	If the interviewer wants to make any notes about the case that might help with the next interview.

SUBSEQUENT MONTHS' INTERVIEWS

For households in sample for the second, third, and fourth months, the FR has the option of conducting the interview over the telephone. Use of this interviewing mode must be approved by the respondent. Such approval is obtained at the end of the first month's interview upon completion of the labor force and supplemental, if any, questions. This is the preferred method for collecting the data; it is much more time and cost efficient. We obtain approximately 85 percent of interviews in these 3 months-in-samples (MIS) via the telephone. See Part 2 of Figure 7–4 for the questions asked to determine household eligibility and obtain consent for the telephone interview. A personal visit interview attempt is required for the fifth-month interview. After this initial attempt, a telephone interview may be conducted provided the original household still occupies the sample unit. This is after a sample unit's 8-month dormant period and is used to reestablish rapport with the household. Fifth-month households are more likely than any other MIS household to be a replacement household. A replacement household is one in which all the previous month's residents have moved out and been replaced by an entirely different group of residents. This can and does occur in any MIS except for MIS 1 households. As with their MIS 2, 3, and 4 counterparts, households in their sixth, seventh, and eighth MIS are eligible for telephoning interviewing. Once again we collect about 85 percent of these cases via the telephone.

The first thing the FR does in subsequent interviews is update the household roster. The instrument presents a screen (or a series of screens for MIS 5 interviews) that

Figure 7–4. **Interviews: Main Housing Unit Items Asked in MIS 1 and Replacement Households**

Note: This list of items is not all inclusive. The list covers only the main data items and does not include related items used to arrive at the final response (e.g., probes and verification screens). See CPS Interviewing Manual for illustrations of the actual instrument screens for all CPS items.

Part 1. **Items Asked at the Beginning of the Interview**

	Item Name	Item Asks
1	INTRO-b	If interviewer wants to classify case as a noninterview.
2	NONTYP	What type of noninterview the case is (A, B, or C); asked depending on answer to INTRO-b.
3	VERADD	What is the street address (as verification).
4	MAILAD	What is the mailing address (as verification).
5	STRBLT	If the structure was originally built before or after 4/1/90.
6	BUILD	If there are any other units (occupied or vacant) in the building.
7	FLOOR	If there are any occupied or vacant living quarters besides the sample unit on the same floor.
8	PROPER	If there is any other building in the property (occupied or vacant).
9	TENUR-scrn	If unit is owned, rented, or occupied without paid rent.
10	ACCES-scr	If access to household is direct or through another unit; this item is answered by the interviewer (not read to the respondent).
11	MERGUA	If the sample unit has merged with another unit.
12	LIVQRT	What type of housing unit is it (house/apt., mobile home or trailer, etc.); this item is answered by the interviewer (not read to the respondent).
13	LIVEAT	If all persons in the household live or eat together.
14	HHLIV	If any other household on the property lives or eats with the interviewed household.

Part 2. **Items Asked at the End of the Interview**

	Item Name	Item Asks
15	TELHH-scrn	If there is a telephone in the unit.
16	TELAV-scrn	If there is a telephone elsewhere on which people in this household can be contacted; asked depending on answer to TELHH-scrn.
17	TELWHR-scr	If there is a telephone elsewhere, where is the phone located; asked depending on answer to TELAV-scrn.
18	TELIN-scrn	If a telephone interview is acceptable.
19	TELPHN	What is the phone number and whether it is a home or office phone.
20	BSTTM-scrn	When is the best time to contact the respondent.
21	NOSUN-scrn	If a Sunday interview is acceptable.
22	THANKYOU	If there is any reason why the interviewer will not be able to interview the household next month.
23	INOTES-1	If the interviewer wants to make any notes about the case that might help with the next interview; also asks for a list of names/ages of ALL additional persons if there are more than 16 household members.

verifies the accuracy of the roster. Since households in MIS 5 are returning to sample after an 8-month hiatus, additional probing questions are asked to correctly establish the household's current roster and update some characteristics. See Figure 7–8 for a list of major items asked in MIS 5 interviews. If there are any changes, the instrument goes through the steps necessary to add or delete an individual(s). Once all the additions/deletions are completed, the instrument then prompts the FR/interviewer to correct or update any relationship items (e.g., relationship to reference person, marital status, and parent and spouse pointers) that may be subject to change. After completion of the appropriate corrections, the instrument will then take the interview to any items, such as educational attainment, that require periodic updating. The labor force interview in MIS 2, 3, 5, 6, and 7 collects the same information as the MIS 1 interview. MIS 4 and 8 interviews are different in several respects. Additional information collected in these interviews includes a battery of questions for employed wage and salary workers on their usual weekly earnings at their only or main job. For all individuals who are multiple

jobholders, information is collected on the industry and occupation of their second job. For individuals who are not in the labor force, we obtain additional information on their previous labor force attachment.

Dependent interviewing is another enhancement contributed to the subsequent months' interviewing by the computerization of the labor force interview. Information collected in the previous month's interview is imported into the current interview to ease response burden and improve the quality of the labor force data. This change is most noticeable in the collection of main job industry and occupation data. By importing the previous month's job description into the current month's interview, we can ascertain whether an individual has the same job as he/she had the preceding month. Not only does this enhance analysis of month-to-month job mobility, it also frees the FR/interviewer from re-entering the detailed industry and occupation descriptions. This speeds the flow through the labor force interview. Other information collected using dependent interviewing is the duration of unemployment (either job search or layoff duration), and the not-in-labor-force subgroups of

Figure 7–5. **Summary Table for Determining Who Is To Be Included As a Member of the Household**

	Include as member of household	
A. PERSONS STAYING IN SAMPLE UNIT AT TIME OF INTERVIEW		
Person is member of family, lodger, servant, visitor, etc.		
1. Ordinarily stays here all the time (sleeps here)	Yes	
2. Here temporarily - no living quarters held for person elsewhere	Yes	
3. Here temporarily - living quarters held for person elsewhere		No
Person is in Armed Forces		
1. Stationed in this locality, usually sleeps here	Yes	
2. Temporarily here on leave - stationed elsewhere		No
Person is a student - Here temporarily attending school - living quarters held for person elsewhere		
1. Not married **or** not living with immediate family.		No
2. Married and living with immediate family	Yes	
3. Student nurse living at school	Yes	
B. ABSENT PERSON WHO USUALLY LIVES HERE IN SAMPLE UNIT		
Person is inmate of institutional special place - Absent because inmate in a specified institution regardless of whether or not living quarters held for person here		No
Person is temporarily absent on vacation, in general hospital, etc. (including veterans' facilities that are general hospitals) - Living quarters held here for person .	Yes	
Person is absent in connection with job		
1. Living quarters held here for person - temporarily absent while ''on the road'' in connection with job (e.g., traveling salesperson, railroad conductor, bus driver) . . .	Yes	
2. Living quarters held here and elsewhere for person but comes here infrequently (e.g., construction engineer)		No
3. Living quarters held here at home for unmarried college student working away from home during summer school vacation	Yes	
Person is in Armed Forces - was member of this household at time of induction but currently stationed elsewhere		No
Person is a student in school - away temporarily attending school - living quarters held for person here		
1. Not married or not living with immediate family.	Yes	
2. Married and living with immediate family		No
3. Attending school overseas.		No
4. Student nurse living at school		No
C. EXCEPTIONS AND DOUBTFUL CASES		
Person with two concurrent residences - determine length of time person has maintained two concurrent residences		
1. Has slept greater part of that time in another locality.		No
2. Has slept greater part of that time in sample unit	Yes	
Citizen of foreign country temporarily in the United States		
1. Living on premises of an Embassy, Ministry, Legation, Chancellery, or Consulate .		No
2. Not living on premises of an Embassy, Ministry, etc.		
a. Living here and no usual place of residence elsewhere in the United States	Yes	
b. Visiting or traveling in the United States		No

retired and disabled. Dependent interviewing is not used in the MIS 5 interviews nor is it used for any of the data collected solely in MIS 4 and 8.

Another recent development in the computerization of the CPS is the use of centralized facilities for computer assisted telephone interviewing (CATI). CPS has been experimenting with the use of CATI interviewing since 1983 and has gradually been introducing it into the CPS operations. The first use of CATI in production was the Tri-Cities Test which started in April 1987. Since that time, more and more cases have been sent to our CATI facilities for interviewing. Currently, about 9,200 cases are sent to the facilities each month. The facilities generally interview about 88-90 percent of the cases assigned to them. The net result is that about 15 percent of all CPS interviews are completed at a CATI facility.

Three facilities are in use: Hagerstown, MD; Tuscon, AZ; and Jeffersonville, IN. During the time of the initial phasein of CATI data, and continuing to this day, there is a controlled selection criteria (see Chapter 4) that allows the analysis of any effects of the CATI collection methodology. See Chapter 16 for a discussion of CATI effects on the labor force data. One of the main reasons for using CATI is to ease the recruiting and hiring effort in hard to enumerate areas. It is much easier to hire an individual to work in the CATI facilities than it is to hire individuals to work as FRs in most major metropolitan areas. This is especially true in most large cities. Most of the cases sent to CATI are from the major metropolitan areas. CATI is not used in most rural areas because the sample sizes in these areas are small enough to keep a single FR busy the entire interview week without causing any undo hardship in completing the assignment. A concerted effort is made to hire some Spanish speaking interviewers in the Tuscon Telephone Center. This allows us to send Spanish speaking cases to this facility for interviewing. For obvious reasons stated above, no MIS 1 or 5 cases are sent to the facilities.

As stated above, the facilities complete all but 10-12 percent of the cases sent to them. These uncompleted cases are recycled back to the field for followup and final determination. For this reason, the CATI facilities generally cease interviewing the labor force portions of the interview on Wednesday of interview week. This allows the field staff 3 to 4 days to check on the case and complete any required interviewing or classify the case as a noninterview. The field staff is highly successful in completing these cases as interviews, generally interviewing about 80-85 percent of the cases. The cases that are sent to the CATI facilities are selected by the supervisors in each of the regional offices. The FRs provide information on a household's probable acceptance of a CATI interview, but ultimately it is the supervisor's decision. This decision is based on the FR's input and the need to balance workloads and meet specific goals on the number

Figure 7–6. **Interviews: Main Demographic Items Asked in MIS 1 and Replacement Households**

Note: This list of items in not all inclusive. The list covers only the main data items and does not include related items used to arrive at the final response (e.g., probes and verification screens). See CPS Interviewing Manual for illustrations of the actual instrument screens for all CPS items.

Part 1. **Items Asked at Beginning of Interview**

	Item Name	Item Asks
1	HHRESP	What is the line number of the household respondent.
2	RPNAME	What is the name of the reference person (i.e., person who owns/rents home, whose name should appear on line number 1 of the household roster).
3	NEXTNM	What is the name of the next person in the household (lines number 2 through a maximum of 16).
4	VERURE	If the sample unit is the person's usual place of residence.
5	HHMEM-scrn	If the person has his/her usual place of residence elsewhere; asked only when the sample unit is not the person's usual place of residence.
6	SEX-scrn	What is the person's sex; this item is answered by the interviewer (not read to the respondent).
7	MCHILD	If the household roster (displayed on the screen) is missing any babies or small children.
8	MAWAY	If the household roster (displayed on the screen) is missing usual residents temporarily away from the unit (e.g., traveling, at school, in a hospital).
9	MLODGE	If the household roster (displayed on the screen) is missing any lodgers, boarders, or live-in employees.
10	MELSE	If the household roster (displayed on the screen) is missing anyone else staying in the unit.
11	RRP-nscr	How is the person related to the reference person; the interviewer shows the respondent a flashcard from which he/she chooses the appropriate relationship category.
12	VR-NONREL	If the person is related to anyone else in the household; asked only when the person is not related to the reference person.
13	SBFAMILY	Who on the household roster (displayed on the screen) is the person related to; asked depending on answer to VR-NONREL.
14	PAREN-scrn	What is the parent's line number.
15	BMON-scrn	What is the month of birth.
	BDAY-scrn	What is the day of birth.
	BYEAR-scrn	What is the year of birth.
16	AGEVR	How many years old is the person (as verification).
17	MARIT-scrn	What is the person's marital status; asked only of persons 15+ years old.
18	SPOUS-scrn	What is the spouse's line number; asked only of persons 15+ years old.
19	AFEVE-scrn	If the person ever served on active duty in the U.S. Armed Forces; asked only of persons 17+ years old.
20	AFWHE-scrn	When did the person serve; asked only of persons 17+ years old who have served in the U.S. Armed Forces.
21	AFNOW-scrn	If the person is now in the U.S. Armed Forces; asked only of persons 17+ years old who have served in the U.S. Armed Forces. Interviewers will continue to ask this item each month as long as the answer is ''yes.''
22	EDUCA-scrn	What is the highest level of school completed or highest degree received; asked only of persons 15+ years old. This item is asked for the first time in MIS 1, and then verified in MIS 5 and in specific months (i.e., February, July, and October).
23	RACE-scrn	What is the person's race; the interviewer shows the respondent a flashcard from which he/she chooses the appropriate race category.
24	ORIGI-scrn	What is the person's origin; the interviewer shows the respondent a flashcard from which he/she chooses the appropriate origin category.
25	SSN-scrn	What is the person's social security number; asked only of persons 15+ years old. This item is asked only from December through March, regardless of month in sample.
26	CHANGE	If there has been any change in the household roster (displayed with full demographics) since last month, particularly in the marital status.

Part 2. **Items Asked at the End of the Interview**

	Item Name	Item Asks
27	NAT1	What is the person's country of birth.
28	MNAT1	What is his/her mother's country of birth.
29	FNAT1	What is his/her father's country of birth.
30	CITZN-scr	If the person is a citizen of the U.S.; asked only when neither the person nor both of his/her parents were born in the U.S. or U.S. territory.
31	CITYA-scr	If the person was born a citizen of the U.S.; asked when the answer to CITZN-scr is yes.
32	CITYB-scr	If the person became a citizen of the U.S. through naturalization; asked when the answer to CITYA-scr is no.
33	INUSY-scr	When did the person come to live in the U.S.; asked of U.S. citizens born outside of the 50 states (e.g., Puerto Ricans, U.S. Virgin Islanders, etc.) and of non-U.S. citizens.
34	FAMIN-scrn	What is the household's total combined income during the past 12 months; the interviewer shows the respondent a flashcard from which he/she chooses the appropriate income category.

Figure 7–7. **Demographic Edits in the CPS Instrument**

Note: The following list of edits is not all inclusive; only the major edits are described. The demographic edits in the CPS instrument take place while the interviewer is creating or updating the roster. After the roster is in place, the interviewer may still make changes to the roster (e.g., add/delete persons, change variables) at the Change screen. However, the instrument does not include any more demographic edits past the Change screen, because inconsistencies—between the data collected at that point and the data collected earlier in the interview—are difficult to resolve without risking endless loops.

Education Edits

1. The instrument will force interviewers to probe if the education level is inconsistent with the person's age; interviewers will probe for the correct response if the education entry fails any of the following range checks:
    • If 19 years old, the person should have an education level below the level of a master's degree (EDUCA-scrn < 44).
    • If 16-18 years old, the person should have an education level below the level of a bachelor's degree (EDUCA-scrn 43).
    • If younger than 15 years old, the person should have an education below college level (EDUCA-scrn < 40).

2. The instrument will force the interviewer to probe before it allows him/her to lower an education level reported in a previous month in sample.

Veterans' Edit
1. The instrument will display only the answer categories that apply (i.e., periods of service in the Armed Forces), based on the person's age. For example, the instrument will not display certain answer categories for a 40 year old veteran (e.g., World War I, World War II, Korean war), but it will display them for a 99 year old veteran.

Nativity Edits
1. The instrument will force the interviewer to probe if the person's year of entry into the U.S. is earlier than his/her year of birth.

Spouse Line Number Edits
1. If the household roster does not include a spouse for the reference person, the instrument will set the reference person's SPOUSE line number equal to zero. It will also omit the first answer category (i.e., married spouse present) when it asks for the marital status of the reference person).
2. The instrument will not ask SPOUSE line number for both spouses in a married couple. Once it obtains the SPOUSE line number for the first spouse on the roster, it will fill the second spouse's SPOUSE line number with the line number of the first spouse. Likewise, the instrument will not ask marital status for both spouses. Once it obtains the marital status for the first spouse on the roster, it will set the second spouse's marital status equal to that of his/her spouse.
3. Before assigning SPOUSE line numbers, the instrument will verify that there are opposing sex entries for each spouse. If both spouses are of the same sex, the interviewer will be prompted to fix whichever one is incorrect.
4. For each household member with a spouse, the instrument will ensure that his/her SPOUSE line number is not equal to his/her own line number, nor to his/her own PARENT line number (if any). In both cases, the instrument will not allow the interviewer to make the wrong entry and will display a message telling the interviewer to ''TRY AGAIN.''

Parent Line Number Edits
1. The instrument will never ask for the reference person's PARENT line number. It will set the reference person's PARENT line number equal to the line number of whomever on the roster was reported as the reference person's parent (i.e., an entry of 24 at RRP-nscr), or equal to zero if no one on the roster fits that criteria.
2. Likewise, for each individual reported as the reference person's child (an entry of 22 at RRP-nscr), the instrument will set his/her PARENT line number equal to the reference person's line number, without asking for each individual's PARENT line number.
3. The instrument will not allow more than two parents for the reference person.
4. If the individual is the reference person's brother or sister (i.e., an entry of 25 at RRP-nscr), the instrument will set his/her PARENT line number equal to the reference person's PARENT line number. However, the instrument will not do so without first verifying that the parent that both siblings have in common is indeed the one whose line number appears in the reference person's PARENT line number (since not all siblings have both parents in common).
5. For each household member, the instrument will ensure that his/her PARENT line number is not equal to his/her own line number. In such a case, the instrument will not allow the interviewer to make the wrong entry and will display a message telling the interviewer to ''TRY AGAIN.''

Figure 7–8. **Interviews: Main Items (Housing Unit and Demographic) Asked in MIS 5 Cases**

Note: This list of items is not all inclusive. The list covers only the main data items and does not include related items used to arrive at the final response (e.g., probes and verification screens). See CPS Interviewing Manual for illustrations of the actual instrument screens for all CPS items.

Housing Unit Items

	Item Name	Item Asks
1	HHNUM-vr	If household is a replacement household.
2	VERADD	What is the street address (as verification).
3	CHNGPH	If current phone number needs updating.
4	MAILAD	What is the mailing address (as verification).
5	TENUR-scrn	If unit is owned, rented, or occupied without paid rent.
6	TELHH-scrn	If there is a telephone in the unit.
7	TELIN-scrn	If a telephone interview is acceptable.
8	TELPHN	What is the phone number and whether it is a home or office phone.
9	BSTTM-scrn	When is the best time to contact the respondent.
10	NOSUN-scrn	If a Sunday interview is acceptable.
11	THANKYOU	If there is any reason why the interviewer will not be able to interview the household next month.
12	INOTES-1	If the interviewer wants to make any notes about the case that might help with the next interview; also asks for a list of names/ages of ALL additional persons if there are more than 16 household members.

Demographic Items

	Item Name	Item Asks
13	RESP1	If respondent is different from the previous interview.
14	STLLIV	If all persons listed are still living in the unit.
15	NEWLIV	If anyone else is staying in the unit now.
16	MCHILD	If the household roster (displayed on the screen) is missing any babies or small children.
17	MAWAY	If the household roster (displayed on the screen) is missing usual residents temporarily away from the unit (e.g., traveling, at school, in hospital).
18	MLODGE	If the household roster (displayed on the screen) is missing any lodgers, boarders, or live-in employees.
19	MELSE	If the household roster (displayed on the screen) is missing anyone else staying in the unit.
20	EDUCA-scrn	What is the highest level of school completed or highest degree received; asked for the first time in MIS 1, and then verified in MIS 5 and in specific months (i.e., February, July, and October).
21	CHANGE	If, since last month, there has been any change in the household roster (displayed with full demographics), particularly in the marital status.

of cases sent to the facilities. The selection of cases to send to CATI also must take into consideration statistical aspects of the selection. With this constraint in mind, selection of cases is controlled by a series of random group numbers assigned at the time of sample selection that either allow or disallow the assignment of a case to CATI. This control procedure allows the Census Bureau to continue to analyze statistical differences that may exist between the data collected in CATI mode versus data collected in CAPI mode.

Figures 7–9 and 7–10 show the results of a typical month's (December 1996) CPS interviewing. Figure 7–9 lists the outcomes of all the households in the CPS sample. The expectations for normal monthly interviewing are a Type A rate around 6.0 percent with an overall noninterview rate in the 20-22 percent range. In December 1996, the Type A rate was 5.81. For the April 1996 - March 1997 period, the CPS Type A rate was 6.36 percent. The months of January, February, and March 1996 were not used due to the effects of the Federal government shut downs. The overall noninterview rate for December 1996 was 20.34 percent, compared to the 12-month average of 20.75 percent.

Figure 7–10 shows the rates of personal and telephone interviewing in December 1996. It is highly consistent with the usual monthly results for personal and telephone interviews.

Figure 7–9. **Interviewing Results (December 1996)**

Description	
Total Assigned Workload	60,234
Interviewed Cases	47,981
Response Rate	94.19%
Interviewed CAPI	41,464
CAPI Partial Interviews	108
Assigned CATI	7,112
Interviewed CATI	6,517
CATI Partial Interviews	58
CATI Recycles to Region	595
Noninterviews	12,253
NI Rate	20.34%
Type A Noninterviews	2,960
Type A Rate	5.81%
No One Home	638
Temporarily Absent	292
Refused	1,928
Other Occupied	99
Accessed Instr. - No Progress	3
Type B Noninterviews	8,921
Type B Rate	14.90%
Armed Forces Occupied or < age 15	93
Temp. Occupied With Persons With URE	1,254
Vacant Regular (REG)	6,005
Vacant HHLD Furniture Storage	374
Unfit, to be Demolished	424
Under Construction, Not Ready	259
Converted to Temp. Business or Storage	186
Unoccupied Tent or Trailer Site	274
Permit Granted, Construction Not Started	37
Other Type B	15
Type C Noninterviews	372
Type C Rate	0.62%
Demolished	39
House or Trailer moved	36
Outside Segment	13
Converted to Perm. Business or Storage	26
Merged	33
Condemned	1
Built after April 1, 1980	76
Unused Serial No#/Listing Sheet Line	73
Other Type C	75

Figure 7–10. **Telephone Interview Rates (December 1996)**

		Telephone		Personal		Indeterminate	
	Total	Number	Percent	Number	Percent	Number	Percent
Total	47,981	34,285	71.5	13,548	28.2	148	0.3
MIS 1&5	11,710	2,677	22.9	8,968	76.6	65	0.6
MIS 2-4, 6-8	36,271	31,608	87.1	4,580	12.6	83	0.2

Chapter 8.
Transmitting the Interview Results

INTRODUCTION

With the advent of the completely electronic interviewing environment, transmission of the interview results took on a heightened importance to the field representative (FR). Until recently, the FR dropped the completed interviews in the mail and waited for the U.S. Postal Service to complete the task. Now, the data must be prepared for transmission and sent out via computer. This chapter provides a summary of these procedures and how the data are prepared for production processing.

The system for transmission of data is centralized at headquarters. All data transfers must pass through headquarters even if that is not the final destination of the information. The system was designed this way for ease of management and to ensure uniformity of procedures within a given survey and between different surveys. The transmission system was designed to satisfy the following requirements:

- Provide minimal user intervention.

- Upload and/or download in one transmission.

- Transmit all surveys in one transmission.

- Transmit software upgrades with data.

- Maintain integrity of software and assignment.

- Prevent unauthorized access.

- Handle mail messages.

Another important aspect about the system is the headquarters central database system cannot initiate transmissions. Either the FR or the regional offices (ROs) must initiate any transmissions. Computers in the field are not connected to the headquarters computers at all times. Instead, the field computers contact the headquarters computers and all data exchanges take place at the time of such call-ins. The central database system contains a group of servers that store messages and case information required by the FRs or the ROs. When an interviewer calls in, the transfer of data from the FR's computer to headquarters computers is completed first and then any outgoing data are transferred to the FR's computer.

A major concern with the use of a electronic method of transmitting interview data is the need for complete security of these data. Both the Census Bureau and the Bureau of Labor Statistics (BLS) are required to honor the pledge of confidentiality given to all Current Population Survey (CPS) respondents. The system was designed to safeguard this pledge. All transmissions between the headquarters central database and the FR's computers are compacted and encrypted for this reason. All transmissions between the headquarters, the ROs, the Centralized Telephone Facilities, and Jeffersonville are over secure telecommunications lines that are leased by the Census Bureau and are accessible only by Census Bureau employees through their computers.

TRANSMISSION OF INTERVIEW DATA

Telecommunications exchanges between a FR and the central database system usually take place once per day during the interview period. Additional transmissions may be made at any time as needed. Each transmission is a batch process in which all relevant files are automatically transmitted and received.

Each FR is expected to make a telecommunications transmission at the end of every work day during the interview period. This is usually accomplished by a preset transmission; that is, each evening the FR sets the FR's computer up to transmit the completed work and then hooks the computer up to a modem within the FR's residence. During the night, at a preset time, the computer automatically dials into the headquarters central database and transmits the completed cases. At the same time, the central database returns any messages and other data to complete the FR's assignment. It is also possible for a FR to make an immediate transmission at any time of the day. The results of such a transmission are identical to a preset transmission, both in the types and directions of various data transfers, but allow the FR instant access to the central database as necessary. This type of procedure is used primarily around the time of closeout when a FR might have one or two straggler cases that need to be received by headquarters before field staff can close out the month's workload and production processing can begin. The RO staff may also perform a daily data transmission sending in cases that required supervisory review or were completed at the RO.

Centralized Telephone Facility Transmission

Most of the cases sent to a Centralized Telephone Facility are successfully completed as computer assisted telephone interviews (CATI). Those that cannot be completed from the telephone center are transferred to a FR

prior to the end of the interview period. These cases are called "CATI recycles." Each telephone facility makes daily transmissions of completed cases and recycles to the headquarters database. All the completed cases are batched for further processing. Each recycled case is transmitted directly to the computer of the FR who is assigned the case. Case notes that include the reason for recycle are also transmitted to the FR to assist in follow-up. CATI recycles are placed on the server when a daily transmission is performed for the region. Daily transmissions are performed automatically for each region every hour during the CPS interview week. This provides cases to the FRs that are reassigned or recycled.

The RO staff also monitor the progress of the CATI recycled cases Recycle Report. All cases that are sent to a CATI facility are also assigned to a FR by the RO staff. The RO staff keep a close eye on recycled cases to ensure that they are completed on time and to monitor the reasons for recycling so that future recycling can be minimized. The reason for recycling is also closely monitored to ensure that recycled cases are properly handled by the CATI facility and correctly identified as CATI eligible by the FR.

Transmission of Interviewed Data From the Centralized Database

Each day during the production cycle (see Figure 8-1 for an overview of the daily processing cycle) the field staff send to the production processing system at headquarters four files containing the results of the previous day's interviewing. A separate file is received from each of the CATI facilities, and all data received from the FRs are batched together and sent as a single file. At this time, cases requiring industry and occupation coding (I&O) are identified, and a file of such cases is created. This file is then used by Jeffersonville coders to assign the appropriate I&O codes. This cycle repeats itself until all data are received by headquarters, usually Tuesday or Wednesday of the week after interviewing begins. By the middle of the interview week, the CATI facilities close down, usually Wednesday, and only one file is received daily by headquarters production processing system. This continues until field closeout day when multiple files may be sent to expedite the preprocessing.

Data Transmissions for I&O Coding

The I&O data are not actually transmitted to Jeffersonville. Rather, the coding staff directly access the data on headquarters computers through the use of remote monitors in Jeffersonville. When a batch of data has been completely coded, that file is returned to headquarters, and the appropriate data are loaded into the headquarters database. See Chapter 9 for a complete overview of the I&O coding and processing system.

Once these transmission operations have been completed, final production processing begins. Chapter 9 provides a description of the processing operation.

Figure 8–1. **Overview of CPS Monthly Operations**

Week 1 (week containing the 5th)			Week 2 (week containing the 12th)							Week 3 (week containing the 19th)							Week 4 (week containing the 26th)							Week 5	
Thu	Fri	Sat	Sun	Mon	Tue	Wed	Thu	Fri	Sat	Sun	Mon	Tue	Wed	Thu	Fri	Sat	Sun	Mon	Tue	Wed	Thu	Fri	Sat	Sun thru Thu	Fri

Field reps pick up computerized questionnaire via modem

Field reps and CATI interviewers complete practice interviews and home studies

Field reps pick up assignments via modem

CATI interviewing
CATI facilities transmit completed work to headquarters nightly | CATI interviewing ends*

CAPI interviewing
● FRs transmit completed work to headquarters nightly
● ROs monitor interviewing assignments

Processing begins
● Files received overnight are checked in by ROs and headquarters daily
● Headquarters performs initial processing and sends eligible cases to Jeffersonville for industry and occupation coding

Jeffersonville sends back completed I&O cases

All assignments completed except for telephone holds

All interviews completed

ROs complete final field close-out

Initial process close-out

J'ville close-out

Headquarters performs final computer processing
● Edits
● Recodes
● Weighting

Census delivers data to BLS

BLS analyzes data

Results released to public at 8:30 am by BLS

* CATI interviewing is extended in March and certain other months.

Chapter 9.
Data Preparation

INTRODUCTION

For the Current Population Survey (CPS), the goal of post data collection processing is to transform a raw data file, as collected by interviewers, into a microdata file that can be used to produce estimates. There are several processes needed in order for this to be accomplished. The raw data files must be read and processed. Textual industry and occupation responses are coded. Even though some editing takes place in the instrument at the time of interview (see Chapter 7), further editing is required once all the data are received. Editing and imputations are performed to improve the consistency and completeness of the micro-data. New data items are created based upon responses to multiple questions. All of this is in preparation for weighting and estimation (see Chapter 10).

DAILY PROCESSING

For a typical month, computer assisted telephone inter-viewing (CATI) starts on Sunday of the week containing the 19th of the month and continues through Wednesday of the same week. The answer files representing these interviews are received on a daily basis from Monday through Thurs-day of this interview week. Separate files are received for each of the three CATI facilities: Hagerstown, Tucson, and Jeffersonville. Computer assisted personal interviewing (CAPI) also begins on the same Sunday and continues through Monday of the following week. The CAPI answer files are received for each day of interviewing, until all the interviewers and regional offices have transmitted the workload for the month. This is generally completed by Wednesday of the following week. These answer files are read, and various computer checks are performed to ensure the data can be accepted into the CPS processing system. These checks include, but are not limited to, ensuring the successful transmission and receipt of the files, item range checks, and rejection of invalid cases. Files containing records needing three-digit industry and occupation (I&O) codes are electronically sent to Jefferson-ville for assignment of these codes. Once Jeffersonville has completed the I&O coding, the files are electronically transferred back to headquarters, where the codes are placed on the CPS production file. When all of the expected data for the month are accounted for and all of Jefferson-ville's I&O coding files have been returned and placed on the appropriate records on the data file, editing and impu-tation are performed.

INDUSTRY AND OCCUPATION (I&O) CODING

The industry and occupation coding operation for a typical month requires ten coders for a period of just over 1 week to code data from 27,000 individuals.[1] At other times, these same coders are available for similar activities on other surveys, where their skills can be maintained. The volume of codes has decreased significantly with the introduction of dependent interviewing for I&O codes (see Chapter 6). Only the new monthly CPS cases, as well as those persons whose industry or occupation has changed since the previous month of interviewing, are sent to Jeffersonville to be coded. For those persons whose industry and occupation have not changed, the three-digit codes are brought forward from the previous month of interviewing and require no further coding.

A computer assisted industry and occupation coding system is used by the Jeffersonville I&O coders. Files of all eligible I&O cases are sent to this system each day. Each coder works at a computer terminal where the computer screen displays the industry and occupation descriptions that were captured by the field representatives at the time of the interview. The coder then enters three-digit numeric industry and occupation codes used in the 1990 census that represent the industry and occupation descriptions.

A substantial effort is directed at supervision and control of the quality of this operation. The supervisor is able to turn the dependent verification setting "on" or "off" at any time during the coding operation. The "on" mode means that a particular coder's work is verified by a second coder. In addition, a 10 percent sample of each month's cases is selected to go through a quality assurance system to evaluate the work of each coder. The selected cases are verified by another coder after the current monthly process-ing has been completed.

After this operation is complete, the batch of records is electronically returned to headquarters to be used in the monthly production processing.

EDITS AND IMPUTATIONS

The CPS suffers from two sources of nonresponse. The largest results from noninterview households. We compen-sate for this data loss in the weighting where essentially the

---

[1]Because of the CPS sample increase in July 2001 from the State Children's Health Insurance Program, the number of cases for I&O coding has increased to about 30,000.

weights of noninterviewed households are distributed among interviewed households (see Chapter 10). The second source of data loss is from item nonresponse. Item nonresponse occurs when a respondent either does not know the answer to a question or refuses to provide the answer. Item nonresponse in the CPS is modest (see Chapter 16).

We compensate for item nonresponse in the CPS by using 1 of 3 imputation methods. Before the edits are applied, the daily data files are merged and the combined file is sorted by state and PSU within state. This sort ensures that allocated values are from geographically related records; that is, missing values for records in Maryland will not receive values from records in California. This is an important distinction since many labor force and industry and occupation characteristics are geographically clustered.

The edits effectively blank all entries in inappropriate questions and ensure that all appropriate questions have valid entries. For the most part, illogical entries or out-of-range entries have been eliminated since the use of electronic instruments; however, the edits still address these possibilities due to data transmission problems and occasional instrument malfunctions. The main purpose of the edits, however, is to assign values to questions where the response was "Don't know" or "Refused." This is accomplished by using 1 of 3 imputation techniques described below.

Before discussing these imputation techniques, it is important to note that the edits are run in a deliberate and logical sequence. That is, demographic variables are edited first because several of those variables are used to allocate missing values in the other modules. The labor force module is edited next since labor force status and related items are used to impute missing values for industry and occupation codes and so forth.

The three imputation methods used by the CPS edits are described below.

1. Relational imputation infers the missing value from other characteristics on the person's record or within the household. For instance, if race is missing, it is assigned based on the race of another household member or failing that, taken from the previous record on the file. Similarly, if relationship is missing, it is assigned by looking at age and sex of the person in conjunction with the known relationship of other household members. Missing occupation codes are sometimes assigned by viewing the industry codes and vice versa. This technique is used exclusively in the demographic and industry and occupation edits. If missing values cannot be assigned using this technique, they are assigned using 1 of the 2 following methods.

2. Longitudinal edits are used primarily in the labor force edits. If a question is blank and the record is in the overlap sample, the edit looks at last month's data to determine whether there was a nonallocated entry for that item. If so, last month's entry is assigned; otherwise, the item is assigned a value using the appropriate hot deck, as described next.

3. This imputation method is commonly referred to as "hot deck" allocation. This method assigns a missing value from a record with similar characteristics. Hot decks are always defined by age, race, and sex . Other characteristics used in hot decks vary depending on the nature of the question being referenced. For instance, most labor force questions use only age, race, sex, and occasionally another labor force item such as full- or part-time status. This means the number of cells in labor force item hot decks are relatively small, perhaps less than 100. On the other hand, the weekly earnings hot deck is defined by age race, sex, usual hours, occupation, and educational attainment. This hot deck has several thousand cells.

All CPS items that require imputation for missing values have an associated hot deck . The initial values for the hot decks are the ending values from the preceding month. As a record passes through the edits, it will either donate a value to each hot deck in its path or receive a value from the hot deck. For instance, in a hypothetical case, the hot deck for question X is defined by the characteristics Black/non-Black, male/female, and age 16-25/25+. Further assume a record has the value of White, male, and age 64. When this record reaches question X, the edits determine whether it has a valid entry. If so, that record's value for question X replaces the value in the hot deck reserved for non-Black, male, and age 25+. Comparably, if the record was missing a value for item X, it would be assigned the value in the hot deck designated for non-Black, male, and age 25+.

As stated above the various edits are logically sequenced, in accordance with the needs of subsequent edits. The edits and codes, in order of sequence, are:

1. **Household edits and codes.** This processing step performs edits and creates recodes for items pertaining to the household. It classifies households as interviews or noninterviews and edits items appropriately. Hot deck allocations defined by geography are used in this edit.

2. **Demographic edits and codes.** This processing step ensures consistency between all demographic variables for all individuals within a household. It ensures all interview households have one and only one reference person and that entries stating marital status, spouse, and parents are all consistent. It also creates families based upon these characteristics. It uses longitudinal editing, hot deck allocation defined by related demographic characteristics, and relational imputation.

    Demographic related recodes are created for both individual and family characteristics.

3. **Labor force edits and codes.** This processing step first establishes an edited Major Labor Force Recode (MLR), which classifies adults as either employed, unemployed, or not in the labor force.

   Based upon MLR, the labor force items related to each series of classification are edited. This edit uses longitudinal editing and hot deck allocation matrices. The hot decks are defined by age, race, and sex and, on occasion, by a related labor force characteristic.

4. **Industry and occupation (I&O) edits and codes.** This processing step assigns three-digit industry and occupation codes to those I&O eligible persons for which the I&O coders were unable to assign a code. It also ensures consistency, wherever feasible, between industry, occupation, and class of worker. I&O related recodes are also created. This edit uses relational allocation and hot deck allocation. The hot decks are defined by employment status, age, sex, race, and educational attainment.

5. **Earnings edits and codes.** This processing step performs edits on the earnings series of items for earnings eligible individuals. A usual weekly earnings recode is created to allow earnings amounts to be in a comparable form for all eligible individuals. There is no longitudinal editing because this series of questions is asked only of MIS 4 and 8 households. Hot deck allocation is used here. The hot deck for weekly earnings is defined by age, race, sex, major occupation recode, educational attainment, and usual hours worked. Additional earnings recodes are created.

6. **School enrollment edits and codes.** School enrollment items are edited for individuals 16-24 years old. Hot deck allocation based on age, race, and sex is used.

Chapter 10.
Estimation Procedures for Labor Force Data

INTRODUCTION

The Current Population Survey (CPS) is a multistage probability sample of housing units in the United States. It produces monthly labor force and related estimates for the total U.S. civilian noninstitutional population and for various age, sex, race, and ethnic groups. In addition, estimates for a number of other population subdomains of the nation (e.g., families, veterans, persons with earnings, households) are produced on either a monthly or quarterly basis. Each month a sample of eight panels (called rotation groups) is interviewed, with demographic data collected for all occupants of the sample housing units. Labor force data are collected for persons 15 years and older. Each rotation group is itself a representative sample of the U.S. population. The labor force estimates are derived through a number of steps in the estimation procedure.

The weighting procedures of the supplements are discussed in Chapter 11. The supplements tend to have higher nonresponse rates. In addition, many of the supplements apply to specific demographic subpopulations and thus differ in coverage from the basic CPS universe.

In order to produce national and state estimates from survey data, a weight for each person in the sample is developed through the following steps:

- Preparation of simple unbiased estimates from baseweights and special weights derived from CPS sampling probabilities.

- Adjustment for nonresponse.

- First-stage ratio adjustment to reduce variances due to the sampling of PSUs.

- Second-stage ratio adjustment to reduce variances by controlling CPS estimates of population to independent estimates of the current population.

- Composite estimation which uses survey data from previous months to reduce the variances.

- Seasonal adjustment for key labor force statistics.

In addition to estimates of basic labor force characteristics, several other types of estimates are also produced either on a monthly or a quarterly basis. These include:

- Household-level estimates and estimates of married couples living in the same household using household and family weights.

- Estimates of earnings, union affiliation, and industry and occupation of second jobs collected from respondents in the quarter sample using outgoing rotation group weights.

- Estimates of labor force status by age for veterans and nonveterans using veterans' weights.

- Estimates of monthly gross flows using longitudinal weights.

The additional estimation procedures provide highly accurate estimates for particular subdomains of the civilian noninstitutional population. The processes described in this chapter have remained essentially unchanged since January 1978. Seasonal adjustment for selected labor force categories has been a part of the estimation procedure since June 1975. Some of these processes have been slightly modified after each decennial census, when the CPS sample is restratified and a new set of sample PSUs is identified. Modifications have been made in the noninterview adjustment and the source and scope of the independent estimates of current population used in the second-stage ratio adjustment. In January 1998, a new compositing procedure was introduced. (See Appendix I.)

UNBIASED ESTIMATION PROCEDURE

A probability sample is defined as a sample that has a known nonzero probability of selection for each sample unit. With probability samples, unbiased estimators can be obtained. These are estimates that on average, over repeated samples, yield the population values.

An unbiased estimator of the population total for any characteristic investigated in the survey may be obtained by multiplying the value of that characteristic for each sample unit (person or household) by the reciprocal of the probability with which that unit was selected and summing the products over all units in the sample (Hansen, 1953). By starting with unbiased estimates from a probability sample, various kinds of estimation and adjustment procedures (such as for noninterview) can be applied with reasonable assurance that the overall accuracy of the estimates will be improved.

In the CPS sample for any given month, not all units respond, and this nonresponse is a potential source of bias. This nonresponse averages about 6 to 7 percent. Other factors, such as occasional errors occurring in the sample selection procedure and households or persons missed by interviewers, can also introduce bias. These

missing households or persons can be considered as having zero probability of selection. These two exceptions not withstanding, the probability of selecting each unit in the CPS is known, and every attempt is made to keep departures from true probability sampling to a minimum.

If all units in a sample have the same probability of selection, the sample is called self-weighting, and unbiased estimators can be computed by multiplying sample totals by the reciprocal of this probability. Most of the state samples in the CPS come close to being self-weighting.

Basic Weighting

The sample designated for the 792-area design was selected with probabilities equal to the inverse of the required state sampling intervals shown in Table 3–1. These sampling intervals are called the basic weights (or baseweights). Almost all sample persons within the same state have the same probability of selection. Exceptions include sample persons in New York and California, where households in New York City and Los Angeles-Long Beach Metropolitan area are selected with higher probability than those in the remainder of these two states. As the first step in the estimation procedure, raw counts from the sample housing units are multiplied by the baseweights. Every person in the same housing unit receives the same baseweight.

Effect of Sample Reductions on Basic Weights

As time goes on, the number of households and the population as a whole increases, and continued use of the original sampling interval leads to a larger sample size (with an increase in costs). The sampling interval is regularly adjusted (the probability of selection is reduced) in order to maintain a fairly constant sample size and cost (see Appendix C).

Special Weighting Adjustments

As discussed in Chapter 3, some ultimate sampling units (USUs) are subsampled in the field, because their observed size is much larger than the expected four housing units. During the estimation procedure, housing units in these USUs must receive special weighting factors to account for the change in their probability of selection. For example, an area sample USU expected to have 4 housing units (HUs) but found at the time of interview to contain 36 HUs, could be subsampled at the rate of 1 in 3 to reduce the interviewer's workload. Each of the 12 designated housing units in this case would be given a special weighting factor of 3. In order to limit the effect of this adjustment on the variance of sample estimates, these special weighting factors are limited to a maximum value of 4. At this stage of CPS

estimation process, the special weighting factors are multiplied by the baseweights. The resulting weights are then used to produce "unbiased" estimates. Although this estimate is commonly called "unbiased," it does still include some negligible bias because the size of the special weighting factor is limited to 4. The purpose of this limitation is to achieve a compromise between an increase in the bias and the variance.

ADJUSTMENT FOR NONRESPONSE

Nonresponse arises when households or other units of observation that have been selected for inclusion in a survey fail to yield all or some of the data that were to be collected. This failure to obtain complete results from all the units selected can arise from several different sources, depending upon the survey situation. There are two major types of nonresponse: item nonresponse and complete (or unit) nonresponse. Unit nonresponse refers to the failure to collect any survey data from an occupied sample unit. For example, data may not be obtained from an eligible household in the survey because of respondent's absence, impassable roads, refusal to participate in the interview, or unavailability of the respondents for other reasons. This type of nonresponse in the CPS is called a Type A noninterview. Historically, between 4 and 5 percent of the eligible units in a given month were Type A noninterviews. Recently, the Type A rate has risen to between 6 and 7 percent (see Chapter 16). Item nonresponse occurs when a cooperating unit fails or refuses to provide some specific items of information. Procedures for dealing with this type of nonresponse are discussed in Chapter 9.

In the CPS estimation process, the weights for all interviewed households are adjusted to account for occupied sample households for which no information was obtained because of unit nonresponse (Type A noninterviews). This noninterview adjustment is made separately for similar sample areas that are usually, but not necessarily, contained within the same state. Increasing the weights of interviewed sample units to account for eligible sample units which are not interviewed assumes that the interviewed units are similar to the noninterviewed units with regard to their demographic and socioeconomic characteristics. This may or may not be true. Nonresponse bias results when the nonresponding units differ in relevant respects from those which respond to the survey or to the particular items.

Noninterview Clusters and Noninterview Adjustment Cells

To reduce the size of the bias, the noninterview adjustment is performed within sample PSUs that are similar in MSA (metropolitan statistical area) status and MSA size. These PSUs are grouped together to form noninterview clusters. In general, PSUs belonging to MSAs of the same

(or similar) size in the same state belong to the same noninterview cluster. PSUs classified as MSA are assigned to MSA clusters. Likewise, non-MSA PSUs are assigned to non-MSA clusters. Within each cluster, there is a further breakdown into two noninterview adjustment cells (also called residence cells). Each MSA cluster is split into "central city" and "not central city" cells. Each non-MSA cluster is divided into "urban" and "rural" cells, making a total of 254 adjustment cells (or 127 noninterview clusters).

Computing Noninterview Adjustment Factors

Weighted counts of interviewed and noninterviewed households are tabulated separately for each noninterview adjustment cell. The basic weight multiplied by any special weighting factor is used as the weight for this purpose. The noninterview factor F_{ij} is computed as:

$$F_{ij} = \frac{Z_{ij} + N_{ij}}{Z_{ij}}$$

where

$Z_{ij} =$ is the weighted count of interviewed households in cell j of cluster i, and

$N_{ij} =$ is the weighted count of Type A noninterviewed households in cell j of cluster i.

These factors are applied to data for each interviewed person except in cells where either of the following situations occur:

- The computed factor is greater than or equal to 2.0.

- There are fewer than 50 unweighted interviewed households in the cell.

If any of these situations occur, the weighted counts are combined for the residence cells within the noninterview cluster. A common adjustment factor is computed and applied to weights for interviewed persons within the cluster. Generally, fewer than 10 noninterview clusters require this type of collapsing in a given month.

Weights After the Noninterview Adjustment

At the completion of the noninterview adjustment procedure, the weight for each interviewed person is:

(baseweight) x (special weighting factor) x
(noninterview adjustment factor)

At this point, records for all individuals in the same household have the same weight, since the adjustments discussed so far depend only on household characteristics.

RATIO ESTIMATION

Distributions of demographic characteristics derived from the CPS sample in any month will be somewhat different from the true distributions even for such basic characteristics as age, race, sex, and Hispanic[1] ethnicity. These particular population characteristics are closely correlated with labor force status and other characteristics estimated from the sample. Therefore, the variance of sample estimates based on these characteristics can be reduced when, by the use of appropriate weighting adjustments, the sample population distribution is brought as closely into agreement as possible with the known distribution of the entire population with respect to these characteristics. This is accomplished by means of ratio adjustments. There are two ratio adjustments in the CPS estimation process: the first-stage ratio adjustment and the second-stage ratio adjustment.

In the first-stage ratio adjustment, weights are adjusted so that the distribution of Black and non-Black census population from the sample PSUs in a state corresponds to the Black/non-Black population distribution from the census for all PSUs in the state. In the second-stage ratio adjustment, weights are adjusted so that aggregated CPS sample estimates match independent estimates of population in various age/sex/race and age/sex/ethnicity cells at the national level. Adjustments are also made so that the estimated state populations from CPS match independent state population estimates.

FIRST-STAGE RATIO ADJUSTMENT

Purpose of the First-Stage Ratio Adjustment

The purpose of the first-stage ratio adjustment is to reduce the contribution to the variance of sample state-level estimates arising from the sampling of PSUs. That is, the variance that would still be associated with the state-level estimates even if we included in the survey all households in every sample PSU. This is called the between-PSU variance. For some states, the between-PSU variance makes up a relatively large proportion of the total variance, while the relative contribution of the between-PSU variance at the national level is generally quite small. As can be seen in Table 14–3 the first-stage ratio adjustment causes a rise in the national relative variance factor, but the second-stage ratio adjustment decreases the relative variance factor below that of the noninterview adjustment. Further research into the effect of the combined first- and second-stage ratio adjustment is needed to determine whether the first-stage ratio adjustment is in fact meeting its purpose.

---

[1]Hispanics may be of any race.

There are several factors to be considered in determining what information to use in applying the first-stage adjustment. The information must be available for each PSU, correlated with as many of the statistics of importance published from the CPS as possible, and reasonably stable over time so that the gain from the ratio adjustment procedure does not deteriorate. The basic labor force categories (unemployed, nonagricultural employed, etc.) could be considered. However, this information could badly fail the stability criterion. The distribution of population by race (Black/non-Black) satisfies all three criteria.

By using the Black/non-Black categories, the first-stage ratio adjustment compensates for the fact that the racial composition of an NSR (nonself-representing) sample PSU could differ substantially from the racial composition of the stratum it is representing. This adjustment is not necessary for SR (self-representing) PSUs since they represent only themselves.

Computing First-Stage Ratio Adjustment Factors

The first-stage adjustment factors are based on 1990 census data and are applied only to sample data for the NSR PSUs. Factors are computed for the two race categories (Black, non-Black) for each state containing NSR PSUs. The following formula is used to compute the first-stage adjustment factors for each state:

$$FS_{sj} = \frac{\sum\limits_{i=1}^{n} C_{sij}}{\sum\limits_{k=1}^{m} \left[\dfrac{1}{\pi_{sk}}\right] (C_{skj})}$$

where

FS_{sj} = the first-stage factor for state s and race cell j (j=Black, non-Black).

C_{sij} = the 1990 16+ census population for NSR PSU i (sample or nonsample) in state s, race cell j.

C_{skj} = the 1990 16+ census population for NSR sample PSU k in state s, race cell j.

π_{sk} = 1990 probability of selection for sample PSU k in state s.

n = total number of NSR PSUs (sample and nonsample) in state s.

m = number of sample NSR PSUs in state s.

The estimate in the denominator of each of the ratios is obtained by multiplying the 1990 census population in the appropriate race cell for each NSR sample PSU by the inverse of the probability of selection for that PSU and summing over all NSR sample PSUs in the state.

The Black and non-Black cells are collapsed within a state when a cell meets one of the following criteria:

- The factor (FS_{sj}) is greater than 1.3.

- The factor is less than 1/1.3=.769230.

- There are fewer than 4 NSR sample PSUs in the state.

- There are fewer than ten expected interviews in a race cell in the state.

Race cells are collapsed within 23 states, resulting in first-stage factors of 1.0 for these states. Eight states are self-representing and have first-stage factors of 1.0 by definition. (See Table 10–1).

Weights After First-Stage Ratio Adjustment

At the completion of the first-stage ratio adjustment, the weight for each responding person is the product of:

(baseweight) x (special weighting adjustment factor) x (noninterview adjustment factor) x (first-stage ratio adjustment factor).

The weight after the first-stage adjustment is called the first-stage weight. Records for all individuals in the same household should have the same first-stage weight, except for mixed-race households having some members who are Black and others who are non-Black.

Table 10–1: **State First-Stage Factors for CPS 1990 Design (September - December 1995)**

State	Black	Non-Black
Alabama	0.92986976	1.02360321
Alaska	1.00000000	1.00000000**
Arizona	1.00000000	1.00000000**
Arkansas	1.03854268	0.99625041
California	0.92824011	1.01550280
Colorado	1.00000000	1.00000000**
Connecticut	1.00000000	1.00000000*
Delaware	1.00000000	1.00000000*
District of Columbia	1.00000000	1.00000000*
Florida	1.07779736	1.00025192
Georgia	1.06981965	0.98237807
Hawaii	1.00000000	1.00000000**
Idaho	1.00000000	1.00000000**
Illinois	1.01947743	1.00254363
Indiana	1.16715920	0.99747055
Iowa	1.00000000	1.00000000**
Kansas	1.00000000	1.00000000**
Kentucky	1.09352656	0.99897341
Louisiana	1.04956759	0.98344772
Maine	1.00000000	1.00000000**
Maryland	1.00000000	1.00000000**
Massachusetts	1.00000000	1.00000000*
Michigan	0.92441097	0.99798724
Minnesota	1.00000000	1.00000000**
Mississippi	0.98243024	1.00997154
Missouri	1.00000000	1.00000000**
Montana	1.00000000	1.00000000**
Nebraska	1.00000000	1.00000000**
Nevada	1.00000000	1.00000000**
New Hampshire	1.00000000	1.00000000*
New Jersey	1.00000000	1.00000000*
New Mexico	1.00000000	1.00000000**
New York	0.83647167	1.00163779
North Carolina	1.07378643	0.98057928
North Dakota	1.00000000	1.00000000**
Ohio	0.97362223	1.00000085
Oklahoma	1.01775196	1.00400963
Oregon	1.00000000	1.00000000**
Pennsylvania	1.17560284	0.99367587
Rhode Island	1.00000000	1.00000000*
South Carolina	0.93971915	1.05454366
South Dakota	1.00000000	1.00000000**
Tennessee	1.08638935	0.98680199
Texas	1.23277658	0.97475648
Utah	1.00000000	1.00000000**
Vermont	1.00000000	1.00000000*
Virginia	1.00000000	1.00000000**
Washington	1.00000000	1.00000000**
West Virginia	1.22587622	0.99959222
Wisconsin	1.00000000	1.00000000**
Wyoming	1.00000000	1.00000000**

* These states contain only SR PSUs in the 1990 sample design and have an implied first-stage factor of 1.000000.
** Race cells were collapsed.

SECOND-STAGE RATIO ADJUSTMENT

The second-stage ratio adjustment decreases the error in the great majority of sample estimates. Chapter 14 illustrates the amount of reduction in variance for key labor force estimates. The procedure is also believed to reduce the bias due to coverage errors (see Chapter 15). The procedure adjusts the weights for sample records within each rotation group to control the sample estimates for a number of geographic and demographic subgroups of the population to ensure that these sample-based estimates of population match independent population controls in each of these categories. These independent population controls are updated each month. Three sets of controls are used:

- The civilian noninstitutional population 16 years of age and older for the 50 states and the District of Columbia.

- Total national civilian noninstitutional population for 14 Hispanic and 5 non-Hispanic age-sex categories (see Table 10–2).

- Total national civilian noninstitutional population for 66 white, 42 Black, and 10 "Other" age-sex categories (see Table 10–3).

The adjustment is done separately for each of the eight rotation groups which comprise a monthly sample. Adjusting the weights to match one set of controls can cause differences in other controls, so an iterative process is used to simultaneously control all variables. Successive iterations begin with the weights as adjusted by all previous iterations. A total of six iterations is performed, which results in virtual consistency between the sample estimates and population controls. The three-way (state, Hispanic/sex/age, race/sex/age) raking ratio estimator is also known as iterative proportional fitting.

In addition to reducing the error in many CPS estimates and converging to the population controls within six iterations for most items, the raking ratio estimator has another desirable property. When it converges, this estimator minimizes the statistic

$$\sum_i W_{2i} \ln(W_{2i}/W_{1i}),$$

where

W_{2i} = the final weight for the ith sample record, and
W_{1i} = the weight for the ith record after the first-stage adjustment.

Thus, the raking ratio estimator adjusts the weights of the records so that the sample estimates converge to the population controls while, in some sense, minimally affecting the first-stage weights. The reference (Ireland and Kullback, 1968) provides more details on the properties of raking ratio estimation.

Table 10–2. **Initial Cells for Hispanic/Non-Hispanic by Age and Sex**

Ages	Hispanic[1]	
	Male	Female
0-5		
6-13		
*14		
*15		
16-19		
20-29		
30-49		
50+		

* No distinction is made between sexes.
[1]Hispanics may be of any race.

Ages	Non-Hispanic
*0-5	
*6-13	
*14	
*15	
*16+	

*No distinction is made between sexes.

Table 10–3. **Initial Cells for Black/White/Other by Age, Race, and Sex**

Age	Black male	Black female
0-1		
2-3		
4-5		
6-7		
8-9		
10-11		
12-13		
14		
15		
16-17		
18-19		
20-24		
25-29		
30-34		
35-39		
40-44		
45-49		
50-54		
55-59		
60-64		
65+		

Age	White male	White female
0		
1		
2		
3		
4		
5		
6		
7		
8		
9		
10-11		
12-13		
14		
15		
16		
17		
18		
19		
20-24		
25-26		
27-29		
30-34		
35-39		
40-44		
45-49		
50-54		
55-59		
60-62		
63-64		
65-67		
68-69		
70-74		
75+		

Ages	Other[1] male	Other[1] female
0-5		
6-13		
*14		
*15		
16-44		
45+		

*No distinction is made between sexes.
[1]Other includes American Indian, Eskimo, Aleut, Asian, and Pacific Islander.

Sources of Independent Controls

The independent population controls used in the second-stage ratio adjustment are prepared by projecting forward the population figures derived from the 1990 decennial census using information from a variety of other data sources that account for births, deaths, and net migration. Subtracting estimated numbers of resident Armed Forces personnel and institutionalized persons from the resident population gives the civilian noninstitutional population. Estimates of net census undercount, determined from the Post-Enumeration Survey, are added to the population projections. One should note that, prepared in this manner, the controls are themselves estimates. However, they are derived independently of the CPS and provide useful information for adjusting the sample estimates. See Appendix D for more details on sources and derivation of the independent controls.

Computing Initial Second-Stage Ratio Adjustment Factors

As mentioned before, the second-stage adjustment involves a three-way rake:

- State

- Hispanic/sex/age

- Race/sex/age

In order to prevent the second-stage adjustment from increasing the variance of the sample estimates, estimation cells which have zero estimates (i.e., no sample respondents) or extremely large or small adjustment factors are identified and combined (or collapsed) with others. No collapsing is done for the state rake. Prior to iteration 1 for the Hispanic/sex/age and the race/sex/age rakes, initial adjustment factors are computed by rotation group for each of the cells shown in Tables 10–2 and 10–3.

For any particular cell j, the initial factor is computed as:

$$F_{jk} = C_j / E_{jk}$$

where

C_j = the population control for cell j (divided by 8, because the raking is done for each rotation group)

E_{jk} = the CPS estimate for cell j and rotation group k after the first-stage ratio adjustment.

These initial factors are not used in the estimates. They are only used to determine whether any cells need to be collapsed. A cell is combined with adjacent (next higher or next lower) age cells in the same race/sex or ethnicity/sex category if:

- It contains no sample respondents (i.e., $E_{jk} = 0$).

- Its initial factor is less than or equal to 0.6.

- Its initial factor is greater than or equal to 2.0.

Collapsing cells continues until none of the three criteria are met or all available cells have been collapsed. Once cells are collapsed, all cell definitions are maintained through all six iterations of the raking procedure. In a typical month approximately 10 cells require collapsing.

Raking

For each iteration of each rake an adjustment factor is computed for each cell and applied to the estimate of that cell. The factor is the population control divided by the estimate of the current iteration for the particular cell. The three steps are repeated through six iterations. The following simplified example begins after one state rake. The example shows the raking for two cells in an ethnicity rake and two cells in a race rake. Age/sex cells and one race cell (see Tables 10–2 and 10–3) have been collapsed here for simplification.

Iteration 1:
    State rake
    Hispanic rake
    Race rake

Example of Raking Ratio Adjustment
Raking Estimates by Ethnicity and Race

E_s = Estimate from CPS sample after state rake
E_e = Estimate from CPS sample after ethnicity rake
E_r = Estimate from CPS sample after race rake
F_e = Ratio adjustment factor for ethnicity
F_r = Ratio adjustment factor for race

Iteration 1 of the Ethnicity Rake

	Non-Hispanic	Hispanic	Population controls
Non-Black	$E_s = 650$ $F_e = \dfrac{1050}{650+180} = 1.265$ $E_e = E_sF_e = 822$	$E_s = 150$ $F_e = \dfrac{250}{150+20} = 1.471$ $E_e = E_sF_e = 221$	1000
Black	$E_s = 180$ $F_e = \dfrac{1050}{650+180} = 1.265$ $E_e = E_sF_e = 228$	$E_s = 20$ $F_e = \dfrac{250}{150+20} = 1.471$ $E_e = E_sF_e = 29$	300
Population controls	1050	250	1300

Iteration 1 of the Race Rake

	Non-Hispanic	Hispanic	Population controls
Non-Black	$E_e = 822$ $F_r = \dfrac{1000}{822+221} = .959$ $E_r = E_eF_r = 788$	$E_e = 221$ $F_r = \dfrac{1000}{822+221} = .959$ $E_r = E_eF_r = 212$	1000
Black	$E_e = 228$ $F_r = \dfrac{300}{228+29} = 1.167$ $E_r = E_eF_r = 266$	$E_e = 29$ $F_r = \dfrac{300}{228+29} = 1.167$ $E_r = E_eF_r = 34$	300
Population controls	1050	250	1300

Iteration 2 (repeat steps above beginning with sample cell estimates at the end of iteration 1)

.
.
.

Iteration 6

Note that the matching of estimates to controls for the race rake causes the cells to differ slightly from the controls for the ethnicity rake or previous rake. With each rake, these differences decrease when cells are matched to the controls for the most recent rake. For the most part, after six iterations the estimates for each cell have converged to the population controls for each cell. Thus, the weight for each record after the second-stage ratio adjustment procedure can be thought of as the weight for the record after the first-stage ratio adjustment multiplied by a series of 18 adjustment factors (six iterations of three rakes). The product of these 18 adjustment factors is called the second-stage ratio adjustment factor.

Weight After the Second-Stage Ratio Adjustment

At the completion of the second-stage ratio adjustment, the record for each person has a weight reflecting the product of:

(baseweight) x (special weighting adjustment factor) x (noninterview adjustment factor) x (first-stage ratio adjustment factor) x (second-stage ratio adjustment factor)

The weight after the second-stage ratio adjustment is also called the final weight. The estimates produced using the final weights are often called the first- and second-stage combined (FSC) estimates. The second-stage ratio adjustment factors also provide estimates of CPS coverage (see Chapter 16).

COMPOSITE ESTIMATOR[2]

Once each record has a final weight, an estimate of level for any given set of characteristics identifiable in the CPS can be computed by summing the final weights for all the sample cases that have that set of characteristics. The process for producing this type of estimate has been variously referred to as a Horvitz-Thompson estimator, a two-stage ratio estimator, or a simple weighted estimator. But the estimator actually used for the derivation of most official CPS labor force estimates that are based upon information collected every month from the full sample (in contrast to information collected in periodic supplements or from partial samples) is a composite estimator. Composite estimation for the CPS modifies the aggregated FSC estimates without adjusting the weights of individual sample records. This is a disadvantage for data users, since composite estimates for a particular month cannot be produced from a microdata file for only that month. However, research is being conducted into the possibility of using a composite estimator from which composite estimates could be produced from a single month's microdata file (Lent, Miller, and Cantwell, 1994).

In general, a composite estimate is a weighted average of several estimates. The composite estimate from the CPS has historically combined two estimates. The first of these is the FSC estimate described above. The second consists of the composite estimate for the preceding month and an estimate of the change from the preceding to the current month. The estimate of the change is based upon data from that part of the sample which is common to the two months (about 75 percent). The higher month-to-month correlation between estimates from the same sample units tends to reduce the variance of the estimate of month-to-month change. Although the average improvements in variance from the use of the composite estimator are

greatest for estimates of month-to-month change, improvements are also realized for estimates of change over other intervals of time and for estimates of levels in a given month (Breau and Ernst, 1983).

Prior to 1985, the two estimators described in the preceding paragraph were the only terms in the CPS composite estimator and were given equal weight. Since 1985, the weights for the two estimators have been unequal and a third term has been included, an estimate of the net difference between the incoming and continuing parts of the current month's sample. The formula for the composite estimate of a labor force level $Y_t^{/}$ is

$$Y_t^{/} = (1-K)\hat{Y}_t + K(Y_{t-1}^{/} + \triangle_t) + A\hat{\beta}_t$$

where

$$\hat{Y}_t = \sum_{i=1}^{8} x_{t,i}$$

$$\triangle_t = \frac{4}{3}\sum_{i \in s}(x_{t,i} - x_{t-1, i-1}) \text{ and}$$

$$\hat{\beta}_t = \sum_{i \notin s} x_{t,i} - \frac{1}{3}\sum_{i \in s} x_{t,i}$$

i   =   1,2,...,8 (month in sample).

$x_{t,i}$ =   sum of weights after second-stage ratio adjustment of respondents in month t, and month in sample i with characteristic of interest.

S =   {2,3,4,6,7,8}.

In the formula above, $\hat{Y}_t$ is the current month's FSC estimate, $\triangle_t$ is the estimate of change from rotation groups common to months t and t-1, and $\hat{\beta}_t$ is the estimate of the net difference between the incoming and continuing part of the current month's sample. The third term, $\hat{\beta}_t$, was added primarily because it was found to keep composited and uncomposited estimates closer while leaving intact the variance reduction advantages of the composite estimator. It also is often characterized as an adjustment term for the bias associated with rotation groups or time in sample in the CPS (see Chapter 15). Historically, the incoming parts of the sample (the rotation groups in their first or fifth month-in-sample) have tended to have higher labor force participation and unemployment rates than the continuing parts (Bailar, 1975). The third term offsets, to some degree, the effect of this bias on the estimate of change in the second term. If there was no such bias, then the expected value of $\hat{\beta}_t$ would be zero.

The values of the constants K (K=0.4) and A (A=0.2) were chosen to be approximately optimal for reducing variances of estimates of labor force characteristics. They reflect a compromise of optimal values across a variety of characteristics and types of estimates (Kostanich and Bettin, 1986).

---

[2]See Appendix I for update.

The compositing formula is applied only to estimated levels. Other types of composite estimates (rates, percents, means, medians) are computed from the component composited levels (e.g., an unemployment rate is computed by taking the composited unemployment level as a percentage of the composited labor force level).

SEASONAL ADJUSTMENT

A time series is a sequence of observations (measurements or estimates) of a particular measurable phenomenon over time. The use and interpretation of many time series, including many of the monthly labor force time series that are based on the CPS, are enhanced by the process of seasonal adjustment. The objective of seasonal adjustment is to measure and remove from time series the effects of normal seasonality caused by such things as weather, holidays, and school schedules. Seasonality can account for much of the observed month-to-month change in estimates, such as those for employment and unemployment, and can obscure the underlying movements associated with long-term cycles and trends that are of great economic significance for most users. For example, the unadjusted CPS levels of employment and unemployment in June are consistently much higher than those for May because of the influx of students into the labor force. If the only change that occurred in the unadjusted estimates between May and June approximated the normal seasonal change, then the seasonally adjusted estimates for the 2 months should be about the same, indicating that essentially no change occurred in the underlying business cycle and trend even though there may have been a large change in the unadjusted data. Changes that do occur in the seasonally adjusted series reflect changes not associated with normal seasonal change and should provide information about the direction and magnitude of changes in the behavior of trend and business cycle effects. They may, however, also reflect the effects of sampling error and other irregularities, which are not removed by the seasonal adjustment process. Change in the seasonally adjusted series can and occasionally should be in a direction opposite the movement in the unadjusted series.

Refinements of the methods used for seasonal adjustment have been under development for decades. The procedure used since 1980 for the seasonal adjustment of the labor force series is the X-11 ARIMA program, developed by Statistics Canada in the late 1970s as an extension and improvement of the widely used X-11 method developed at the U.S. Census Bureau in the 1960s. The X-11 approach to seasonal adjustment is univariate and nonparametric and involves the iterative application of a set of moving averages that can be summarized as one lengthy weighted average (Dagum, 1983). Nonlinearity is introduced by a set of rules and procedures for identifying and reducing the effect of "extreme values." In most uses of X-11 seasonal adjustment, including that for CPS-based labor force series, the seasonality is estimated as evolving rather than fixed over time. A detailed description of the program is given in Scott and Smith (1974).

The current official practice for the seasonal adjustment of the labor force series involves the running of all directly adjusted series through X-11 ARIMA twice each year, after receipt of June and December data, with 6 months of projected factors drawn from each run and used in the subsequent 6 months, and historical revisions drawn from the end-of-year run. This practice allows, among other things, the publication of the seasonal factors prior to their use.

Seasonally adjusted estimates of many labor force series, including the levels of the civilian labor force, employment, and unemployment and all unemployment rates, are derived indirectly by arithmetically combining or aggregating the series directly adjusted with X-11 ARIMA. For example, the overall unemployment rate is computed using 12 directly adjusted components—unemployment, agricultural employment, and nonagricultural employment by sex and by age (16-19 and 20+). The principal reason for doing such indirect adjustment is that it ensures that the major seasonally adjusted totals will be arithmetically consistent with at least one set of components. If the totals were directly adjusted along with the components, such consistency would generally not occur because X-11 is not a sum- or ratio-preserving procedure. It is not generally appropriate to apply factors computed for an aggregate series to the components of the aggregate because various components tend to have significantly different patterns of seasonal variation.

For up-to-date information and a more thorough discussion on the seasonal adjustment of the labor force series, see the January issues of **Employment and Earnings** (U.S. Department of Labor).

MONTHLY STATE AND SUBSTATE ESTIMATES

Employment and unemployment estimates for states and local areas are key indicators of local economic conditions. Under a Federal-state cooperative program, monthly estimates of the civilian labor force and unemployment are prepared for some 6,950 areas, including all states, metropolitan areas, counties, cities of 25,000 population or more, and all cities and towns in New England. The Bureau of Labor Statistics is responsible for the concepts, definitions, technical procedures, validation, and publication of the estimates which are prepared by state employment security agencies.

The state and area estimates are used by a wide variety of customers. Federal programs base allocations to states and areas on the data, as well as eligibility determinations for assistance. State and local governments use the estimates for planning and budgetary purposes and to determine the need for local employment and training services. Private industry and individuals use the data to compare and assess labor market developments in states and substate areas.

The underlying concepts and definitions of all labor force data developed for state and substate areas are consistent with those of the Current Population Survey (CPS). Annual average estimates for all states are derived directly from the CPS. In addition, monthly estimates for 11 large states (California, Florida, Illinois, Massachusetts, Michigan, New Jersey, New York, North Carolina, Ohio, Pennsylvania, and Texas) and two substate areas (New York City and the Los Angeles-Long Beach metropolitan area) also come directly from the CPS; these areas have a sufficiently large sample to meet the BLS standard of reliability. For the remaining 39 states and the District of Columbia, where the CPS sample is not large enough to produce reliable monthly estimates, data are produced using a signal-plus-noise time series models. These models combine current and historical data from the CPS, the Current Employment Statistics (CES) program, and State unemployment insurance (UI) systems to produce estimates that reflect each state's individual economy.

Estimates for substate labor market areas (other than the two direct-use CPS areas) are produced through a building block approach which uses data from several sources, including the CPS, CES, state unemployment insurance systems, and the decennial census to create estimates which are then adjusted to the state model-based measures of employment and unemployment. Below the labor market area level, estimates are prepared for all counties and cities with populations of 25,000 or more using disaggregation techniques based on decennial census and annual population estimates and current UI statistics.

Estimates for States

The employment and unemployment estimates for 11 large states (California, Florida, Illinois, Massachusetts, Michigan, New Jersey, New York, North Carolina, Ohio, Pennsylvania, and Texas) are sufficiently reliable to be taken directly from the Current Population Survey (CPS) on a monthly basis. These states are termed "direct-use" states.

For the 39 smaller states and the District of Columbia (called "nondirect-use" states) models based on a signal-plus-noise approach (Bell and Hillmer, 1990; Tiller, 1992) are used to develop employment and unemployment estimates.[3] The model of the signal is a time series model of the true labor force which consists of three components: a variable coefficient regression, a flexible trend, and a flexible seasonal component. The regression components are based on historical and current relationships found within each state's economy as reflected in the different sources of data that are available for each state—the CPS, the Current Employment Statistics (CES) survey, and the

unemployment insurance (UI) system. The noise component of the models explicitly accounts for autocorrelation in the CPS sampling error and changes in the average magnitude of the error. While all the state models have important components in common, they differ somewhat from one another to better reflect individual state labor force characteristics. See Appendix E for more details on the models.

Two models—one for the employment-to-population ratio and one for the unemployment rate—are used for each state. The employment-to-population ratio, rather than the employment level, and the unemployment rate, rather than the unemployment level, are estimated primarily because the hyperparameters used in the state-space models are easier to estimate and compare across states for ratios than levels.

The employment-to-population ratio models estimate the monthly CPS employment-to-population ratio using the state's monthly CPS data and employment data from the CES. The models also include trend and seasonal components to account for different seasonal patterns in the CPS not captured by the CES series. The seasonal component accounts for the seasonality in the CPS not explained by the CES while the trend component adjusts for long-run systematic differences between the two series (Evans, Tiller, and Zimmerman, 1993).

The unemployment rate models estimate the monthly CPS unemployment rate using the state's monthly unemployment insurance claims data, along with trend and seasonal components (Zimmerman, Evans, and Tiller, 1993). Once the estimates are developed from the models, levels are calculated for the employment, unemployment, and labor force.

State estimates are seasonally adjusted using the X-11 ARIMA seasonal-adjustment procedure twice a year and new seasonal adjustment factors are created. The entire series is used as an input to the seasonal adjustment process; however, for programming purposes only the last 5 years of data are replaced.

Benchmarking and Population Controls

Once each year, when new population controls are available from the Census Bureau, CPS labor force estimates for all states and the District of Columbia are adjusted to these controls. The model-based estimates for the 39 states and the District of Columbia are reestimated incorporating the recontrolled CPS labor force estimates (see above), and these (monthly) estimates are adjusted, or benchmarked, to the newly population-controlled annual average CPS estimates. The benchmarking technique uses a procedure called the Denton Method (Denton, 1971) which adjusts the annual average of the model-based employment and unemployment levels to equal the corresponding CPS annual averages, while preserving, as much as possible, the original monthly seasonal pattern of the model-based estimates.

---

[3]Because of the January 1996 sample reduction, estimates for **all** states are based on a model.

Substate estimates for previous years are also revised on an annual basis. The updates incorporate any changes in the inputs, such as revisions to establishment-based employment estimates, corrections in claims data, and updated historical relationships. The revised estimates are then readjusted to the latest (benchmarked) state estimates of employment and unemployment.

PRODUCING OTHER LABOR FORCE ESTIMATES

In addition to basic weighting to produce estimates for persons, several "special-purpose" weighting procedures are performed each month. These include:

- Weighting to produce estimates for households and families.

- Weighting to produce estimates from data based on only 2 of 8 rotation groups (outgoing rotation weighting for the quarter sample data).

- Weighting to produce labor force estimates for veterans and nonveterans (veterans weighting).

- Weighting to produce estimates from longitudinally-linked files (longitudinal weighting).

Most of these special weights are based on the final weight after the second-stage ratio adjustment. Some also make use of composited estimates. In addition, consecutive monthly estimates are often averaged to produce quarterly or annual average estimates. Each of these procedures is described in more detail below.

Family Weight

Family weights are used to produce statistics on families and family composition. They also provide the basis for household weights. The family weight is derived from the final weight of the reference person in each household. In most households, it is exactly the reference person's weight. However, when the reference person is a married man, for purposes of family weights, he is given the same weight as his wife. This is done so that weighted tabulations of CPS data by sex and marital status show an equal number of married women and married men with their spouses present. If the CPS final weights were used for this tabulation (without any further adjustment), the estimated numbers of married women and married men would not be equal, since the second-stage ratio adjustment tends to increase the weights of males more than the weights of females. The wife's weight is usually used as the family weight, since CPS coverage ratios for women tend to be higher and subject to less month-to-month variability than those for men.

Household Weight

The same household weight is assigned to every person in the same household and is equal to the family weight of the household reference person. The household weight is used to produce estimates at the household level, such as the number of households headed by a female or the number of occupied housing units.

Outgoing Rotation Weights (Quarter Sample Data)

Some items in the CPS questionnaire are asked only in households due to rotate out of the sample temporarily or permanently after the current month. These are the households in the rotation groups in their fourth or eighth month-in-sample, sometimes referred to as the "outgoing" rotation groups. Items asked in the outgoing rotations include those on discouraged workers (through 1993), earnings (since 1979), union affiliation (since 1983), and industry and occupation of second jobs of multiple jobholders (beginning in 1994). Since the data are collected from only one-fourth of the sample each month, these estimates are averaged over 3 months to improve their reliability, and published quarterly.

Since 1979, most CPS files have included separate weights for the outgoing rotations. These weights were generally referred to as "earnings weights" on files through 1993, and are generally called "outgoing rotation weights" on files for 1994 and subsequent years. In addition to ratio adjustment to independent population controls (in the second stage), these weights also reflect additional constraints that force them to sum to the composited estimates of employment, unemployment, and not in labor force each month. An individual's outgoing rotation weight will be approximately four times his or her final weight.

To compute the outgoing rotation adjustment factors, the final CPS weights of the appropriate records in the two outgoing rotation groups are tallied. CPS composited estimates from the full sample for the labor force categories of employed wage and salary workers, other employed, unemployed, and not in labor force by age, race and sex are used as the controls. The adjustment factor for a particular cell is the ratio of the control total to the weighted tally from the outgoing rotation groups.

Collapsing is performed with an adjacent cell (the next higher or lower age group) if:

- A cell has no sample respondents,

- A composited control total is less than or equal to zero, or

- The corresponding adjustment factor is less than or equal to 2.0 or greater than or equal to 8.0 (i.e., one-half or twice the normal adjustment factor of 4).

If a cell requires collapsing, it is collapsed with another cell (the next higher or lower age group) in the same sex, race, and labor force category. The adjustment factors are

recomputed after all collapsing has been performed. The outgoing rotation weights are obtained by multiplying the outgoing ratio adjustment factors by the final CPS weights. For consistency, an outgoing rotation group weight equal to four times the basic CPS family weight is assigned to all persons in the two outgoing rotation groups who were not eligible for this special weighting (military personnel and persons aged 15 and younger).

Production of monthly, quarterly, and annual estimates using the quarter sample data and the associated weights is completely parallel to production of uncomposited, simple weighted estimates from the full sample—the weights are summed and divided by the number of months used. The composite estimator is not applicable for these estimates because there is no overlap between the quarter samples in consecutive months. Because the outgoing rotations are all independent samples within any consecutive 12-month period, averaging of these estimates on a quarterly and annual basis realizes relative reductions in variance greater than those achieved by averaging full sample estimates.

Family Outgoing Rotation Weight

The family outgoing rotation weight is analogous to the family weight computed for the full sample, except that outgoing rotation weights are used, rather than the final weights from the second-stage ratio adjustment.

Veterans' Weights

Since 1986, CPS interviewers have collected data on veteran status from all respondents. Veterans' weights are calculated for all CPS respondents based on their veteran status. This information is used to produce tabulations of employment status for veterans and nonveterans.

The process begins with the final weights after the second-stage ratio adjustment. Each respondent is classified as a veteran or a nonveteran. Veterans' records are classified into six cells: (1) male, pre-Vietnam veteran; (2) female, pre-Vietnam veteran; (3) male, Vietnam veteran; (4) female, Vietnam veteran; (5) male, other veteran; and (6) female, other veteran. The cell definitions change throughout the decade as the minimum age for nonpeacetime veterans increases.

The final weights for CPS veterans are tallied into sex/age/type-of-veteran cells using the classifications described above. Separate ratio adjustment factors are computed for each cell, using independently established monthly counts of veterans provided by the Department of Veterans Affairs. The ratio adjustment factor is the ratio of the independent control total to the sample estimate. The final weight for each veteran is multiplied by the appropriate adjustment factor to produce the veteran's weight.

To compute veterans' weights for nonveterans, a table of composited estimates is produced from the CPS data by sex, race (White/non-White), labor force status (unemployed, employed, and not in the labor force), and age. The

veterans' weights produced in the previous step are tallied into the same cells. The estimated number of veterans is then subtracted from the corresponding cell entry for the composited table to produce nonveterans control totals. The final weights for CPS nonveterans are tallied into the same sex/race/labor force status/age cells. Separate ratio adjustment factors are computed for each cell, using the nonveterans controls derived above. The factor is the ratio of the nonveterans control total to the sample estimate. The final weight for each nonveteran is multiplied by the appropriate factor to produce the nonveterans' weight. A table of labor force estimates by age status for veterans and nonveterans is published each month.

Longitudinal Weights

For many years, the month-to-month overlap of 75 percent of the sample households has been used as the basis for estimating monthly "gross flow" statistics. The difference or change between consecutive months for any given level or "stock" estimate is an estimate of net change that reflects a combination of underlying flows in and out of the group represented. For example, the month-to-month change in the employment level is the number of people who went from not being employed in the first month to being employed in the second month minus the number who made the opposite transition. The gross flow statistics provide estimates of these underlying flows and can provide useful insights to analysts beyond those available in the stock data.

The estimation of monthly gross flows, and any other longitudinal use of the CPS, begins with a longitudinal matching of the microdata (or person-level) records within the rotation groups common to the months of interest. Each matched record brings together all the information collected in those months for a particular individual. The CPS matching procedure uses the household identifier and person line number as the keys for matching. Prior to 1994, it was also necessary to check other information and characteristics, such as age and sex, for consistency to verify that the match based on the keys was almost certainly a valid match. Beginning with 1994 data, the simple match on the keys provides an essentially certain match. Because the CPS does not follow movers (rather, the sample addresses remain in sample according to rotation pattern), and because not all households are successfully interviewed every month they are in sample, it is not possible to match interview information for all persons in the common rotation groups across the months of interest. The highest percentage of matching success is generally achieved in the matching of consecutive months, where between 90 and 95 percent of the potentially matchable records (or about 67 to 71 percent of the full sample) can usually be matched. The use of CATI and CAPI since 1994 has also introduced dependent interviewing which eliminated much of the erratic differences in response between pairs of months.

On most CPS files for 1994 forward, there is a longitudinal weight which allows users to estimate gross labor force flows by simply summing up the longitudinal weights after matching. These longitudinal weights reflect the technique that had been used prior to 1994 to inflate the gross flow estimates to appropriate population levels. That technique simply inflates all estimates or final weights by the ratio of the current month's population controls to the sum of the final weights for the current month in the matched cases by sex. Although the technique does provide estimates consistent with the population levels for the stock data in the current month, it does not force consistency with labor force stock levels in either the current or the previous month, nor does it control for the effects of the bias and sample variation associated with the exclusion of movers, differential noninterview in the matched months, the potential for the compounding of classification errors in flow data, and the particular rotations that are common to the matched months. There have been a number of proposals for improving the estimation of gross labor force flows, but none have yet been adopted in official practice. See *Proceedings of the Conference on Gross Flows in Labor Force Statistics* (U.S. Department of Commerce and U.S. Department of Labor, 1985) for information on some of these proposals and for more complete information on gross labor force flow data and longitudinal uses of the CPS.

Averaging Monthly Estimates

CPS estimates are frequently averaged over a number of months. The most commonly computed averages are (1) quarterly, which provide four estimates per year by grouping the months of the calendar year in nonoverlapping intervals of three, and (2) annual, combining all 12 months of the calendar year. Quarterly and annual averages of uncomposited data can be computed by summing the weights for all of the months contributing to each average and dividing by the number of months involved. Averages for calculated cells, such as rates, percents, means, and medians, are computed from the averages for the component levels, not by averaging the monthly values (e.g., a quarterly average unemployment rate is computed by taking the quarterly average unemployment level as a percentage of the quarterly average labor force level, not by averaging the three monthly unemployment rates together).

Although such averaging multiplies the number of interviews contributing to the resulting estimates by a factor approximately equal to the number of months involved in the average, the sampling variance for the average estimate is actually reduced by a factor substantially less than that number of months. This is primarily because the CPS rotation pattern and resulting month-to-month overlap in sample units ensure that estimates from the individual months are not independent. The reduction in sampling error associated with the averaging of CPS estimates over adjacent months was studied using 12 months of data collected beginning January 1987 (Fisher and McGuinness, 1993). That study showed that characteristics for which the month-to-month correlation is low, such as unemployment, are helped considerably by such averaging, while characteristics for which the correlation is high, such as employment, benefit less from averaging. For unemployment, variances of national estimates were reduced by about one-half for quarterly averages and about one-fifth for annual averages.

REFERENCES

Bailar, B., (1975) "The Effects of Rotation Group Bias on Estimates from Panel Surveys," **Journal of the American Statistical Association,** Vol. 70, pp. 23-30.

Bell, W.R., and Hillmer, S.C. (1990), "The Time Series Approach to Estimation for Repeated Surveys," **Survey Methodology,** 16, pp. 195-215.

Breau, P. and Ernst, L., (1983) "Alternative Estimators to the Current Composite Estimator," **Proceedings of the Section on Survey Research Methods,** American Statistical Association, pp. 397-402.

Copeland, K. R., Peitzmeier, F. K., and Hoy, C. E. (1986), "An Alternative Method of Controlling Current Population Survey Estimates to Population Counts," **Proceedings of the Survey Research Methods Section,** American Statistical Association, pp.332-339.

Dagum, E.B. (1983), **The X-11 ARIMA Seasonal Adjustment Method,** Statistics Canada, Catalog No. 12-564E.

Denton, F.T. (1971), "Adjustment of Monthly or Quarterly Series to Annual Totals: An Approach Based on Quadratic Minimization," **Journal of the American Statistical Association,** 64, pp. 99-102.

Evans, T.D., Tiller, R.B., and Zimmerman, T.S. (1993), "Time Series Models for State Labor Force Estimates," **Proceedings of the Survey Research Methods Section,** American Statistical Association, pp. 358-363.

Fisher, R. and McGuinness R. (1993), "Correlations and Adjustment Factors for CPS (VAR80 1)." Internal Memorandum for Documentation, January 6th, Demographic Statistical Methods Division, U.S. Census Bureau.

Hansen, M. H., Hurwitz, W. N., and Madow, W. G.(1953), **Sample Survey Methods and Theory,** Vol. II, New York: John Wiley and Sons.

Harvey, A.C. (1989), **Forecasting, Structural Time Series Models and the Kalman Filter,** Cambridge: Cambridge University Press.

Ireland, C.T., and Kullback, S. (1968), "Contingency Tables With Given Marginals," **Biometrika,** 55, pp. 179-187.

Kostanich, D. and Bettin, P. (1986), "Choosing a Composite Estimator for CPS," presented for presentation at the International Symposium on Panel Surveys, Washington, DC.

Lent, J., Miller, S., and Cantwell, P. (1994), "Composite Weights for the Current Population Survey," **Proceedings of the Survey Research Methods Section,** American Statistical Association, pp. 867-872.

Oh, H.L., and Scheuren, F., (1978), "Some Unresolved Application Issues in Raking Estimation," **Proceedings of the Section on Survey Research Methods,** American Statistical Association, pp. 723-728.

Proceedings of the Conference on Gross Flows in Labor Force Statistics (1985), U.S. Department of Commerce and U.S. Department of Labor.

Scott, J.J., and Smith, T.M.F. (1974), "Analysis of Repeated Surveys Using Time Series Methods," **Journal of the American Statistical Association,** 69, pp. 674678.

Thompson, J.H., (1981), "Convergence Properties of the Iterative 1980 Census Estimator," **Proceedings of the American Statistical Association on Survey Research Methods,** American Statistical Association, pp. 182185.

Tiller, R.B. (1989), "A Kalman Filter Approach to Labor Force Estimation Using Survey Data," **Proceedings of the Survey Research Methods Section,** American Statistical Association, pp. 16-25.

Tiller, R.B.(1992), "Time Series Modeling of Sample Survey Data From the U.S. Current Population Survey," **Journal of Official Statistics,** 8, pp, 149-166.

U.S. Department of Labor, Bureau of Labor Statistics (Published monthly), **Employment and Earnings,** Washington, DC: Government Printing Office.

Young, A.H., (1968), "Linear Approximations to the Census and BLS Seasonal Adjustment Methods," **Journal of the American Statistical Association,** 63, pp. 445-471.

Zimmerman, T.S., Evans, T.D., and Tiller, R.B. (1994), "State Unemployment Rate Time Series," **Proceedings of the Survey Research Methods Section,** American Statistical Association, pp. 1077-1082.

Chapter 11.
Current Population Survey Supplemental Inquiries

INTRODUCTION

In addition to providing data on the labor force status of the population, the Current Population Survey (CPS) is used to collect data for a variety of studies on the entire U.S. population and specific population subsets. Such studies keep the nation informed of the economic and social well being of its people and are conducted for use by Federal and state agencies, private foundations, and other organizations. Supplemental inquiries take advantage of several special features of the CPS: large sample size and general purpose design; highly skilled, experienced interviewing and field staff; and generalized processing systems that can easily accommodate the inclusion of additional questions.

Some CPS supplemental inquiries are conducted annually, others every other year, and still others on a one-time basis. The frequency and recurrence of a supplement depend on what best meets the needs of the supplement's sponsor. It is important that any supplemental inquiry meet strict criteria discussed in the next section.

Producing supplemental data from the CPS involves more than just including additional questions. Separate data processing is required to edit responses for consistency and to impute missing values. Another weighting method is often necessary because the supplement targets a different universe from the basic CPS. A supplement can also engender a different level of response or cooperation from respondents.

CRITERIA FOR SUPPLEMENT INQUIRIES

As a basis for undertaking supplements for federal agencies or other sponsors, a number of criteria which determine the acceptability of such work have been developed and refined over the years. These criteria were developed by the Census Bureau in consultation with the Bureau of Labor Statistics (BLS).

The staff of the Census Bureau, working with the sponsoring agency, develops the survey design, including the methodology, questionnaires, pretesting options, interviewer instructions and processing requirements. The Census Bureau provides a written description of the statistical properties associated with each supplement. The same standards of quality that apply to the basic CPS apply to the supplements.

The following criteria are considered before undertaking a supplement:

1. The subject matter of the inquiry must be in the public interest.

2. The inquiry must not have an adverse effect on the CPS or other Census Bureau programs. The questions must not cause respondents to question the importance of the survey and result in losses of response or quality. It is essential that the image of the Census Bureau as the objective fact finder for the Nation is not damaged. Other important functions of the Census Bureau, such as the decennial censuses or the economic censuses, must not be affected in terms of quality or response rates or in congressional acceptance and approval of these programs.

3. The subject matter must be compatible with the basic CPS survey. This is to avoid introducing a concept that could affect the accuracy of responses to the basic CPS information. For example, a series of questions incorporating a revised labor force concept that could inadvertently affect responses to the standard labor force items would not be included.

4. The inquiry must not slow down the work of the basic survey or impose a response burden that may affect future participation in the basic CPS. In general, the supplemental inquiry must not add more than 10 minutes of interview time per respondent or 25 minutes per household. All conflicts or competition for the use of Census Bureau staff or facilities that arise in dealing with the supplemental inquiry are resolved by giving the basic CPS first priority. The Census Bureau will not jeopardize the schedule for completing CPS or other Census Bureau work to favor completing a supplemental inquiry within some specified time frame.

5. The subject matter must not be overly sensitive. This criterion is imprecise, and its interpretation has changed over time. For example, the subject of birth expectations, once considered sensitive, has been included as a CPS supplemental inquiry.

6. It must be possible to meet the objectives of the inquiry through the survey method. That is, it must be possible to translate the survey objectives into meaningful questions, and the respondent must be able to supply the information required to answer the questions.

7. If the supplemental information is to be collected during the CPS interview, the inquiry must be suitable for the personal visit/telephone procedures used in the CPS.

8. All data must abide by the Census Bureau's enabling legislation, which, in part, ensures that no information will be released that can identify an individual. Requests for name, address, social security number, or other information that can directly identify an individual will not be included. In addition, information that could be used to indirectly identify an individual with a high probability of success (e.g., small geographic areas in conjunction with income or age) will be suppressed.

9. The cost of supplements must be borne by the sponsor, regardless of the nature of the request or the relationship of the sponsor to the ongoing CPS.

The questionnaires developed for the supplement are subject to the Census Bureau's pretesting policy. This policy was established in conjunction with other sponsoring agencies to encourage questionnaire research aimed at improving data quality.

Even though the proposed inquiry is compatible with the criteria given in this section, the Census Bureau does not make the final decision regarding the appropriateness or desirability of the supplemental survey. The Office of Management and Budget, through its Statistical Policy Division, reviews the proposal to make certain it meets government wide standards regarding the need for the data and appropriateness of the design and ensures that the survey instruments, strategy, and response burden are acceptable.

RECENT SUPPLEMENTAL INQUIRIES

The scope and type of CPS supplemental inquiries vary considerably from month to month and from year to year. Generally, in any given month, a respondent who is selected for the supplement is asked the additional questions that comprise the supplemental inquiry after completing the regular part of the CPS. Table 11-1 summarizes CPS supplemental inquiries that were conducted between September 1994 and December 2001.

The Housing Vacancy Supplement (HVS) is unusual in that it is the only supplement that is conducted every month. This supplement collects additional information (e.g., number of rooms, plumbing, and rental/sales price) on housing units identified as vacant in the basic CPS. Probably the most widely used supplement is the Annual Demographic Survey (ADS) which is conducted every March. This supplement collects data on work experience, several sources of income, migration, household composition, health insurance coverage, and receipt of noncash benefits.

The basic CPS weighting is not always appropriate for supplements, since supplements tend to have higher non-response rates. In addition, supplement universes are generally different from the basic CPS universe. Thus, some supplements require weighting procedures different from the basic CPS. These variations are described for two of the major supplements, the Housing Vacancy Survey and the Annual Demographic Survey, in the following sections.

Housing Vacancy Survey Supplement

Description of supplement

The HVS is a monthly supplement to the CPS sponsored by the Census Bureau. The supplement is administered when the CPS encounters a unit in sample that is intended for year-round or seasonal occupancy and is currently vacant, or occupied by persons with a usual residence elsewhere. The interviewer asks a reliable respondent (e.g., the owner, a rental agent, or a knowledgeable neighbor) questions on year built; number of rooms, bedrooms, and bathrooms; how long the housing unit has been vacant; the vacancy status (for rent, for sale, etc); and when applicable, the selling price or rent amount.

The purpose of the HVS is to provide current information on the rental and homeowner vacancy rates, home ownership rates, and characteristics of units available for occupancy in the United States as a whole, geographic regions, and inside and outside metropolitan areas. The rental vacancy rate is a component of the index of leading economic indicators which is used to gauge the current economic climate. Although the survey is performed monthly, data for the nation and for Northeast, South, Midwest, and West regions are released quarterly and annually. The data released annually include information for states and large metropolitan areas.

Calculation of vacancy rates

The HVS collects data on year-round and seasonal vacant units. Vacant year-round units are those intended for occupancy at any time of the year, even though they may not be in use year-round. In resort areas, a housing unit which is intended for occupancy on a year-round basis is considered a year-round unit; those intended for occupancy only during certain seasons of the year are considered seasonal. Also, vacant housing units held for occupancy by migratory workers employed in farm work during the crop season are classified as seasonal. The rental and homeowner vacancy rates are the most prominent HVS statistics. The vacancy rates are determined using information collected by the HVS and CPS since the formulas use both vacant and occupied housing units.

The rental vacancy rate is calculated as the ratio of vacant year-round units for rent to the sum of renter occupied units, vacant year-round units rented but awaiting occupancy, and vacant year-round units for rent.

Table 11–1. **Current Population Survey Supplements September 1994 - December 2001**

Title	Month	Purpose	Sponsor
Housing Vacancy	Monthly	Provide quarterly data on vacancy rates and characteristics of vacant units.	Census
Health/Pension	September 1994	Provide information on health/pension coverage for persons 40 years of age and older. Information includes benefit coverage by former as well as current employer and reasons for noncoverage, as appropriate. Amount, cost, employer contribution, and duration of benefits are also measured. Periodicity: As requested.	PWBA
Lead Paint Hazards Awareness	December 1994 June 1997	Provide information on the current awareness of the health hazards associated with lead-based paint. Periodicity: As requested.	HUD
Contingent Workers	February 1995, 1997, 1999, 2001	Provide information on the type of employment arrangement workers have on their current job and other characteristics of the current job such as earnings, benefits, longevity, etc., along with their satisfaction with and expectations for their current jobs. Periodicity: Biennial.	BLS
Annual Demographic Supplement	March 1995 - 2001	Data concerning work experience, several sources of income, migration, household composition, health insurance coverage, and receipt of noncash benefits. Periodicity: Annual.	Census/BLS
Food Security	April 1995, September 1996, April 1997, August 1998, April 1999, September 2000, April 2001, December 2001	Data that will measure hunger and food security. It will provide data on food expenditures, access to food, and food quality and safety.	FNS
Race and Ethnicity	May 1995, July 2000	Use alternative measurement methods to evaluate how best to collect these types of data.	BLS/Census
Marital History	June 1995	Information from ever married persons on marital history. Periodicity: Quinquennial.	Census/BLS
Fertility	June 1995, 1998, 2000	Data on the number of children that women aged 15-44 have ever had and the children's characteristics. Periodicity: Quinquennial.	Census/BLS
Educational Attainment	July 1995	Tested several methods of collecting these data. Tested both the current method (highest grade completed or degree received) and the old method (highest grade attended and grade completed).	BLS/Census
Veterans	August 1995, August 1997, September 1999, August 2001	Data for veterans of the United States on Vietnam-theatre status, service-connected income, effect of a service-connected disability on current labor force participation, and participation in veterans' programs. Periodicity: Biennial.	BLS
School Enrollment	October 1994, 1995, 1996, 1997, 1998, 1999, 2000, 2001	Provide information on school enrollment, junior or regular college attendance, and high school graduation. Periodicity: Annual.	BLS/Census/NCES
Tobacco Use	September 1995, January 1996, May 1996, September 1998, January 1999, May 1999, January 2000, May 2000, June 2001, November 2001	Data for 15+ population on current and former use of tobacco products; restrictions of smoking in workplace for employed persons; and personal attitudes toward smoking; a repeat of supplements in 1995/1996. Supplements will be repeated in 1998/1999. Periodicity: As requested.	NCI
Displaced Workers	February 1996, 1998, 2000	Provide data on workers who lost a job in the last 5 years because of plant closing, shift elimination, or other work-related reason. Periodicity: Biennial.	BLS
Job Tenure/Occupational Mobility	February 1996, 1998, 2000	Collect data that will measure an individual's tenure with his/her current employer and in his/her current occupation. Periodicity: As requested.	BLS
Child Support	April 1996, 1998, 2000	Identify households with absent parents and provide data on child support arrangements, visitation rights of absent parent, amount and frequency of actual versus awarded child support, and health insurance coverage. Data are also provided on why child support was not received and/or awarded. April data will be matched to March data.	OCSE
Voting and Registration	November 1994, 1996, 1998, 2000	Provide demographic information on persons who registered and did not register to vote. Also measures number of persons who voted and reasons for not registering. Periodicity: Biennial.	Census
Work Schedule/ Home-Based Work	May 1997, 2001	Provide information about multiple job holdings and work schedules and telecommuters who work at a specific remote site.	BLS
Telephone Availability	March, July & November 1994-2001	Collect data on whether there is a phone in the HU, on which to contact the person; is a telephone interview acceptable.	FCC
Computer Use/Internet Use	August 2000, September 2001, November 1994, October 1997	Obtain information about household access to computers and the use of the internet or worldwide web.	NTIA

The homeowner vacancy rate is calculated as the ratio of vacant year-round units for sale to the sum of owner occupied units, vacant year-round units sold but awaiting occupancy, and vacant year-round units for sale.

Weighting procedure

Since the HVS universe differs from the CPS universe, the HVS records require a different weighting procedure from the CPS records. The HVS records are weighted by the CPS basic weight, the CPS special weighting factor, and two HVS adjustments. (Refer to Chapter 10 for a description of the two CPS weighting adjustments.) The two HVS adjustments are referred to as the HVS first-stage ratio adjustment and the HVS second-stage ratio adjustment.

The HVS first-stage ratio adjustment is comparable to the CPS first-stage ratio adjustment in that it reduces the contribution to variance from the sampling of PSUs. The adjustment factors are based on 1990 census data. For each state, they are calculated as the ratio of the state-level census count of vacant year-round housing units in all NSR PSUs to the state-level estimate, using census counts, of vacant year-round housing units in NSR PSUs.

The first-stage adjustment factors are applied to both vacant year-round and seasonal housing units in NSR PSUs.

The HVS second-stage ratio adjustment, which applies to vacant year-round and seasonal housing units in SR and NSR PSUs, is calculated as the ratio of the weighted CPS interviewed housing units after CPS second-stage ratio adjustment to the weighted CPS interviewed housing units after CPS first-stage ratio adjustment.

The cells for the HVS second-stage adjustment are calculated within each month-in-sample by census region and type of area (metropolitan/nonmetropolitan, central city/balance of MSA, and urban/rural). This adjustment is made to all eligible HVS records.

The final weight for each HVS record is determined by calculating the product of the CPS basic weight, the CPS special weighting factor, the HVS first-stage ratio adjustment, and the HVS second-stage ratio adjustment. Note that the occupied units in the denominator of the vacancy rate formulas use a different final weight since the data come from the CPS. The final weight applied to the renter and owner-occupied units is the CPS household weight. (Refer to Chapter 10 for a description of the CPS household weight.)

Annual Demographic Supplement (March Supplement)

Description of supplement

The ADS is sponsored by the Census Bureau and the Bureau of Labor Statistics (BLS). The Census Bureau has collected data in the ADS since 1947. From 1947 to 1955,

the ADS took place in April and from 1956 to 2001 the ADS took place in March. In 2002, a sample increase was implemented requiring more time for data collection. Thus, additional ADS interviews are now taking place in February and April. However, even with this sample increase, most of the data collection still occurs in March.

The supplement collects data on family characteristics, household composition, marital status, migration, income from all sources, information on weeks worked, time spent looking for work or on layoff from a job, occupation and industry classification of the job held longest during the year, health insurance coverage, and receipt of noncash benefits. A major reason for conducting the ADS around the month of March is to obtain better income data. It was thought that since March is the month before the deadline for filing federal income tax returns, respondents were likely to have recently prepared tax returns or be in the midst of preparing such returns and could report their income more accurately than at any other time of the year.

The ADS sample consists of the March CPS sample, plus additional CPS households identified in the prior November and following April CPS samples. Starting in 2002, the eligible ADS sample households are:

1. The entire March CPS sample.

2. Hispanic households — identified in November (from all month-in-sample (MIS) rotation groups) and in April (MIS 1,5).

3. Non-Hispanic non-White households — identified in November (MIS 1, 5-8) and April (MIS 1,5).

4. Non-Hispanic White households with children 18 years or younger — identified in November (MIS 1, 5-8) and April (MIS 1,5).

Prior to 2002, only the November CPS households containing at least one person of Hispanic origin were added to the ADS. The added households in 2002, along with a general sample increase in selected states (see Appendix J), are collectively known as the State Children's Health Insurance Program (SCHIP) sample expansion. The added households improve the reliability of the ADS estimates for the Hispanic households, non-Hispanic non-White households, and non-Hispanic White households with children 18 years or younger.

Because of the characteristics of CPS sample rotation (see Chapter 3), the additional sample from the November and April CPS rotation groups are completely different from those in the March CPS. The additional sample cases increase the effective sample size of the ADS compared to the March CPS sample alone. The ADS sample includes 18 MIS rotation groups for Hispanic households, 15 MIS rotation groups for non-Hispanic non-White households, 15 MIS rotation groups for non-Hispanic White households with children 18 years or younger, and 8 MIS rotation groups for all other households.

Table 11-2. **Summary of ADS Interview Month**

CPS month/Hispanic status		Mover								Nonmover							
		Month in sample								Month in sample							
		1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
November	Hispanic[1]	NI[2]								March							
	Non-Hispanic[3]	NI[2]								Feb.	NI[2]			Feb.	Feb./Apr.		
March	Hispanic[1]	March								March							
	Non-Hispanic[3]	March								March							
April	Hispanic[1]	Apr.	NI[2]		Apr.	NI[2]				Apr.	NI[2]		Apr.	NI[2]			
	Non-Hispanic[3]	Apr.	NI[2]		Apr.	NI[2]				Apr.	NI[2]		Apr.	NI[2]			

[1]Hispanics may be of any race.
[2]NI - Not interviewed for the ADS.
[3]The non-Hispanic group includes both non-Hispanic non-Whites and non-Hispanic Whites with children ≤ 18 years old.

The March and April ADS eligible cases are administered the ADS questionnaire in those respective months (see Table 11-2). The April cases are classified as "split path" cases because they receive the ADS, while other households receive the supplement scheduled for April. The November eligible Hispanic households are administered the ADS questionnaire in March, regardless of rotation group (MIS 1-8). (Note that November MIS 5-8 households have already completed all 8 months of interviewing for the CPS before March, and the November MIS 1-4 households have an extra contact scheduled for the ADS before the 5th interview of the CPS later in the year.)

The November eligible non-Hispanic households (in MIS 1, 5-8) are administered the ADS questionnaire in either February or April. November ADS eligible cases in MIS 1 and 5 are interviewed for the CPS in February (in MIS 4 and 8, respectively) so the ADS questionnaire is administered in February. (These are also split path cases, since other households in the rotation groups get the regular supplement scheduled for February). The November MIS 6-8 eligible cases are split between the February and April CPS interviewing months.

Mover households, defined as households with a different reference person when compared to the November CPS interview, are found at the time of the ADS interview. Mover households identified from the November eligible sample are removed from the ADS sample. Mover households identified in the March and April eligible samples receive the ADS questionnaire.

The ADS sample universe is slightly different from the CPS. The CPS completely excludes military personnel while the ADS includes military personal who live in households with at least one other civilian adult. These differences require the ADS to have a different weighting procedure from the regular CPS.

Weighting procedure

Prior to weighting, missing supplement items are assigned values based on hot deck imputation, a system that uses a statistical matching process. Values are imputed even when all the supplement data are missing. Thus, there is no separate adjustment for households that respond to the basic survey but not to the supplement. The ADS records are weighted by the CPS basic weight, the CPS special weighting factor, the CPS noninterview adjustment, the CPS first-stage ratio adjustment, and the CPS second-stage ratio adjustment procedure. (Refer to Chapter 10 for a description of these adjustments.) Also there is an additional ADS noninterview adjustment for the November ADS sample, a SCHIP Adjustment Factor, a family equalization adjustment, and an adjustment for Armed Forces members.

The November eligible sample, the March eligible sample, and the April eligible sample are weighted separately up to the second-stage weighting adjustment. The samples are then combined so that one second-stage adjustment procedure is performed. The flowchart in Figure 11-1 illustrates the weighting process for the ADS sample.

Households from November eligible sample. The households from the November eligible sample start with their basic CPS weight as calculated in November, modified by their November CPS special weighting factor and November CPS noninterview adjustment. At this point, a second noninterview adjustment is made for those November eligible households that are still occupied, but an interview could not be obtained in the February, March, or April CPS. Then, the November ADS sample weights are adjusted by the CPS first-stage adjustment ratio and finally, the weights are adjusted by the SCHIP Adjustment Factor.

The ADS noninterview adjustment for November eligible sample. The second noninterview adjustment is applied to the November eligible sample households to reflect noninterviews of occupied housing units that occur in the February, March, or April CPS. If a noninterviewed household is actually a Mover household, it would not be eligible for interview. Since the mover status of noninterviewed households is not known, we assume that the proportion of

Figure 11-1. Diagram of the ADS Weighting Scheme

mover households is the same for interviewed and noninterviewed households. This is reflected in the noninterview adjustment. With this exception, the noninterview adjustment procedure is the same as described in Chapter 10. The weights of the interviewed households are adjusted by the noninterview factor as described below. At this point, the noninterviews and those November Mover households are removed from any further weighting of the ADS. The noninterview adjustment factor, F_{ij}, is computed as follows:

$$F_{ij} = \frac{Z_{ij} + N_{ij} + B_{ij}}{Z_{ij} + B_{ij}}$$

where:

$Z_{ij} =$ the weighted number of November eligible sample households interviewed in the February, March, or April CPS in cell j of cluster i.

$N_{ij} =$ the weighted number of November eligible sample occupied, noninterviewed housing units in the February, March, or April CPS in cell j of cluster i.

$B_{ij} =$ the weighted number of November eligible sample Mover households identified in the February, March, or April CPS in cell j of cluster i.

The weighted counts used in this formula are those after the November CPS noninterview adjustment is applied. The clusters refer to the variously defined regions that compose the United States. These include clusters for the Northeast, Midwest, South, and West, as well as clusters for particular cities or smaller areas. Within each of these clusters is a pair of residence cells. These could be (1) Central City and balance of MSA, (2) MSA and non-MSA, or (3) Urban and Rural, depending on the type of cluster.

SCHIP adjustment factor for the November eligible sample. The SCHIP adjustment factor is applied to Nonmover eligible households that contain residents who are Hispanic, non-Hispanic non-White, and non-Hispanic Whites with children 18 years or younger to compensate for the increased sample in these demographic categories. Hispanic households receive a SCHIP adjustment factor of 8/18 and non-Hispanic non-White households and non-Hispanic White households with children 18 years or younger receive a SCHIP adjustment factor of 8/15. (See Table 11-3.) After this adjustment is applied, the November ADS sample is ready to be combined with the March and April eligible samples for the application of the second-stage ratio adjustment.

Households from April eligible sample. The households from the April eligible sample start with their basic CPS weight as calculated in April, modified by their April CPS special weighting factor, the April CPS noninterview adjustment, and the SCHIP adjustment factor. After the SCHIP adjustment factor is applied, the April eligible sample is ready to be combined with the November and March eligible samples for the application of the second-stage ratio adjustment.

SCHIP adjustment factor for the April eligible sample. The SCHIP adjustment factor is applied to April eligible households that contain residents who are Hispanic, non-Hispanic non-Whites, or non-Hispanic Whites with children 18 years or younger to compensate for the increased sample in these demographic categories regardless of mover status. Hispanic households receive a SCHIP adjustment factor of 8/18 and non-Hispanic non-White households and non-Hispanic White households with children 18 years or younger receive a SCHIP adjustment factor of 8/15. Table 11-3 summarizes these weight adjustments.

Households from the March eligible sample. The March eligible sample households start with their basic CPS weight, modified by their CPS special weighting factor, the March CPS noninterview adjustment, the March CPS first-stage ratio adjustment (as described in Chapter 10), and the SCHIP adjustment factor.

SCHIP adjustment factor for the March eligible sample. The SCHIP adjustment factor is applied to the March eligible nonmover households that contain residents who are Hispanic, non-Hispanic non-White, and non-Hispanic Whites with children 18 years or younger to compensate for the increased sample in these demographic categories. Hispanic households receive a SCHIP adjustment factor of 8/18 and non-Hispanic non-White households and non-Hispanic White resident households with children 18 years or younger receive a SCHIP adjustment factor of 8/15. Mover households and all other households receive a SCHIP adjustment of 1. Table 11-3 summarizes these weight adjustments.

Combined household eligible sample from November, March, and April CPS. At this point, the eligible sample from November, March, and April are combined. The remaining adjustments are applied to this combined sample file.

ADS second-stage ratio adjustment. The second-stage ratio adjustment adjusts the ADS estimates so that they agree with independent age, sex, race, and ethnicity population controls as described in Chapter 10. The same procedure, used for CPS, is used for the ADS.

Additional ADS weighting. After the ADS weight (the weight through the second-stage procedure) is determined, the next step is to determine the final ADS weight. There are two more weighting adjustments applied to the ADS sample cases. One is for family equalization. Without this adjustment there would be more married men than married women. Weights, mostly of males, are adjusted to

Table 11-3. **Summary of SCHIP Adjustment Factor**

CPS month/Hispanic status		Mover								Nonmover							
		Month in sample								Month in sample							
		1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
November	Hispanic[1]	0[2]								8/18							
	Non-Hispanic[3]	0[2]								8/15	0[2]			8/15	8/15		
March	Hispanic[1]	1								8/18							
	Non-Hispanic[3]	1								8/15							
April	Hispanic[1]	8/18	0[2]			8/18	0[2]			8/18	0[2]			8/18	0[2]		
	Non-Hispanic[3]	8/15	0[2]			8/15	0[2]			8/15	0[2]			8/15	0[2]		

[1]Hispanics may be of any race.
[2]Zero weight indicates the cases are ineligible for the ADS.
[3]The non-Hispanic group includes both non-Hispanic non-Whites and non-Hispanic Whites with children ≤ 18 years old.

give a husband and wife the same weight, while maintaining the overall age/race/sex/Hispanic control totals. The second adjustment is for members of the Armed Forces.

Family equalization. The family equalization procedure categorizes adults (at least 15 years old) into seven groups based on sex and household composition:

1. Female partners in female/female unmarried partner households
2. All other civilian females
3. Married males, spouse present
4. Male partners in male/female unmarried partner households
5. Other civilian male heads
6. Male partners in male/male unmarried partner households
7. All other civilian males

Three different methods, dependent on the household composition, are used to assign the ADS weight to other members of the household. The methods are 1) assigning the weight of the householder to the spouse or partner, 2) averaging the weights of the householder and partner, or 3) computing a ratio adjustment factor and multiplying the factor by the ADS weight.

Armed forces. Male and female members of the Armed Forces living off post or living with their families on post are included in the ADS as long as there is at least one civilian adult living in the same household, whereas the CPS excludes all Armed Forces members. Households with no civilian adults in the household, i.e., households with all Armed Forces members, are excluded from the ADS. The weights assigned to the Armed Forces members, included in the ADS, is the same weight civilians receive through the SCHIP adjustment. Control totals, used in the second-stage factor, do not include Armed Forces members so the Armed Force members do not go through the second-stage ratio adjustment. During family equalization, a male Armed Forces member with a spouse or partner is reassigned the weight of his spouse/partner.

SUMMARY

Although this discussion focuses on only two CPS supplements, the HVS and the ADS, every supplement has its own unique objectives. The additional questions, edits, and imputations are tailored to each supplement's data needs. For many supplements this also means altering the weighting procedure to reflect a different universe, account for a modified sample, or adjust for a higher rate of nonresponse. The weighting revisions discussed here for HVS and ADS are only indicative of the types of modifications that might be used for a supplement.

Chapter 12.
Data Products From the Current Population Survey

INTRODUCTION

Information collected in the Current Population Survey (CPS) is made available by both the Bureau of Labor Statistics and the Census Bureau through broad publication programs which include news releases, periodicals, and reports. CPS-based information is also available on magnetic tapes, CD-ROM, and computer diskettes and can be obtained online through the Internet. This chapter lists many of the different types of products currently available from the survey, describes the forms in which they are available, and indicates how they can be obtained. This chapter is not intended to be an exhaustive reference for all information available from the CPS. Furthermore, given the rapid ongoing improvements occurring in computer technology, it can be expected that greater numbers of the CPS-based products will be electronically accessible in the future.

BUREAU OF LABOR STATISTICS

Each month, employment and unemployment data are published initially in *The Employment Situation* news release about 2 weeks after data collection is completed. The release includes a narrative summary and analysis of the major employment and unemployment developments together with tables containing statistics for the principal data series. The news release is also available electronically on the Internet and can be accessed at http://stats.bls.gov:80/newsrels.htm.

Subsequently, more detailed statistics are published in *Employment and Earnings,* a monthly periodical. The detailed tables provide information on the labor force, employment, and unemployment by a number of characteristics, such as age, sex, race, marital status, industry, and occupation. Estimates of the labor force status and detailed characteristics of selected population groups not published on a monthly basis, such as Vietnam-era veterans and Hispanics[1] are published every quarter. Data are also published quarterly on usual median weekly earnings classified by a variety of characteristics. In addition, the January issue of *Employment and Earnings* provides annual averages on employment and earnings by detailed occupational categories, union affiliation, and employee absences.

About 25,000 of the monthly labor force data series plus quarterly and annual averages are maintained in LABSTAT, the BLS public database, on the Internet. They can be accessed from http://stats.bls.gov/datahome.htm. In most cases, these data are available from the inception of the series through the current month. Approximately 250 of the most important estimates from the CPS are presented monthly and quarterly on a seasonally adjusted basis. The CPS is used also for a program of special inquiries to obtain detailed information from particular segments or for particular characteristics of the population and labor force. About four such special surveys are made each year. The inquiries are repeated annually in the same month for some topics, including the earnings and total incomes of individuals and families (published by the Census Bureau); the extent of work experience of the population during the calendar year; the marital and family characteristics of workers; the employment of school-age youth, high school graduates and dropouts, and recent college graduates; and the educational attainment of workers. Surveys are also made periodically on subjects such as contingent workers, job tenure, displaced workers, and disabled veterans.

Generally, the persons who provide information for the monthly CPS questions also answer the supplemental questions. Occasionally, the kind of information sought in the special surveys requires the respondent to be the person about whom the questions are asked. The results of these special surveys are first published as news releases and subsequently in the *Monthly Labor Review* or BLS reports.

In addition to the regularly tabulated statistics described above, special data can be generated through the use of the CPS individual (micro) record files. These files contain records of the responses to the survey questionnaire for all individuals in the survey. While the microdata can be used simply to create additional cross-sectional detail, an important feature of their use is the ability to match the records of specific individuals at different points in time during their participation in the survey. (The actual identities of these individuals are protected on all versions of the files made available to noncensus staff.) By matching these records, data files can be created which lend themselves to some limited longitudinal analysis and the investigation of short-run labor market dynamics. An example is the statistics on gross labor force flows, which indicate how many persons move among the labor force status categories each month. Microdata files are available for all months since January 1976 and for various months in prior years. These data are made available on magnetic tape, CD-ROM, or diskette.

---

[1]Hispanics may be of any race.

Annual averages from the CPS for the four census regions and nine divisions, the 50 states and the District of Columbia, 50 large metropolitan areas, and 17 central cities are published annually in *Geographic Profile of Employment and Unemployment.* Data are provided on the employed and unemployed by selected demographic and economic characteristics.

Table 12–1 provides a summary of the CPS data products available from BLS.

U.S. CENSUS BUREAU

The U.S. Census Bureau has been analyzing data from the Current Population Survey and reporting the results to the public for over five decades. The reports provide information on a recurring basis about a wide variety of social, demographic, and economic topics. In addition, special reports on many subjects have also been produced. Most of these reports have appeared in 1 of 3 series issued by the Census Bureau: P-20, Population Characteristics; P-23, Special Studies; and P-60, Consumer Income. Many of the reports are based on data collected as part of the March demographic supplement to the CPS. However, other reports use data from supplements collected in other months (as noted in the listing below). A full inventory of these reports as well as other related products is documented in: *Subject Index to Current Population Reports and Other Population Report Series, CPR P23-192,* which is available from the Government Printing Office, or the Census Bureau. Most reports have been issued in paper form; more recently, some have been made available on the Internet (http://www.census.gov). Generally, reports are announced by press release, and are released to the public via the Census Bureau Public Information Office.

Census Bureau Report Series

P-20, Population Characteristics. Regularly recurring reports in this series include topics such as geographic mobility, educational attainment, school enrollment (October supplement), marital status, households and families, Hispanic origin, the Black population, fertility (June supplement), voter registration and participation (November supplement), and the foreign-born population.

P-23, Special Studies. Information pertaining to special topics, including one-time data collections, as well as research on methods and concepts are produced in this series. Examples of topics include computer ownership and usage, child support and alimony, ancestry, language, and marriage and divorce trends.

P-60, Consumer Income. Regularly recurring reports in this series include information concerning families, individuals, and households at various income and poverty levels, shown by a variety of demographic characteristics. Other reports focus on health insurance coverage and other noncash benefits.

In addition to the population data routinely reported from the CPS, Housing Vacancy Survey (HVS) data are collected from a sample of vacant housing units in the Current Population Survey (CPS) sample. Using these data, quarterly and annual statistics are produced on rental vacancy rates and home ownership rates for the United States, the four census regions, location inside and outside metropolitan areas (MAs), the 50 states and the District of Columbia, and the 75 largest MAs. Information is also made available on national home ownership rates by age of householder, family type, race, and Hispanic origin. A press release is issued each quarter as well as quarterly and annual data tables on the Internet.

Supplement Data Files

Public use microdata files containing supplement data are available from the Census Bureau. These files contain the full battery of basic labor force and demographic data along with the supplement data. A standard documentation package containing a record layout, source and accuracy statement, and other relevant information is included with each file. (The actual identities of the individuals surveyed are protected on all versions of the files made available to noncensus staff.) These files can be purchased through the Customer Services Branch of the Census Bureau and are available in either tape or CD-ROM format. The CPS homepage is the other source for obtaining these files. Eventually, we plan to add most historical files to the site along with all current and future files.

Table 12–1. **Bureau of Labor Statistics Data Products From the Current Population Survey**

Product	Description	Periodicity	Source	Cost
News Releases				
College Enrollment and Work Activity of High School Graduates	An analysis of the college enrollment and work activity of the prior year's high school graduates by a variety of characteristics	Annual	October CPS supplement	Free ([1])
Contingent and Alternative Employment Arrangements	An analysis of workers with ''contingent'' employment arrangements (lasting less than 1 year) and alternative arrangements including temporary and contract employment by a variety of characteristics	Biennial	January CPS supplement	Free ([1])
Displaced Workers	An analysis of workers who lost jobs in the prior 3 years due to plant or business closings, position abolishment, or other reasons by a variety of characteristics	Biennial	February CPS supplement	Free ([1])
Employment Situation of Vietnam-Era Veterans	An analysis of the work activity and disability status of persons who served in the Armed Forces during the Vietnam era	Biennial	September CPS supplement	Free ([1])
Job Tenure of American Workers	An analysis of employee tenure by industry and a variety of demographic characteristics	Biennial	February CPS supplement	Free ([1])
State and Regional Unemployment	An analysis of state and regional employment and unemployment	Annual	CPS annual averages	Free ([1])
The Employment Situation	Seasonally adjusted and unadjusted data on the Nation's employed and unemployed workers by a variety of characteristics	Monthly ([2])	Monthly CPS	Free ([1])
Union Membership	An analysis of the union affiliation and earnings of the Nation's employed workers by a variety of characteristics	Annual	Monthly CPS; outgoing rotation groups	Free ([1])
Usual Weekly Earnings of Wage and Salary Workers	Median usual weekly earnings of full- and part-time wage and salary workers by a variety of characteristics	Quarterly ([3])	Monthly CPS; outgoing rotation groups	Free ([1])
Work Experience of the Population	An examination of the employment and unemployment experience of the population during the entire preceding calendar year by a variety of characteristics	Annual	March CPS supplement	Free ([1])
Periodicals				
Employment and Earnings	A monthly periodical providing data on employment, unemployment, hours, and earnings for the Nation, states, and metropolitan areas	Monthly ([3])	CPS; other surveys and programs	$35.00 domestic; $43.75 foreign
Monthly Labor Review	A monthly periodical containing analytical articles on employment, unemployment, and other economic indicators, book reviews, and numerous tables of current labor statistics	Monthly	CPS; other surveys and programs	$29.00 domestic; $36.25 foreign
Other Publications				
A Profile of the Working Poor	An annual report on workers whose families are in poverty by work experience and various characteristics	Annual	March CPS supplement	Free
Geographic Profile of Employment and Unemployment	An annual publication of employment and unemployment data for regions, states, and metropolitan areas by a variety of characteristics	Annual	CPS annual averages	$9.00
Issues in Labor Statistics	Brief analysis of important and timely labor market issues	Occasional	CPS; other surveys and programs	Free

Table 12–1. **Bureau of Labor Statistics Data Products From the Current Population Survey**—Con.

Product	Description	Periodicity	Source	Cost
Microdata Files				
Job Tenure and Occupational Mobility		Biennial	February CPS supplement	([4])
Displaced Workers		Biennial	February CPS supplement	([4])
Contingent Work		Biennial	January CPS supplement	([4])
Annual Demographic Survey		Annual	March CPS supplement	([4])
Veterans		Biennial	September CPS supplement	([4])
School Enrollment		Annual	October CPS supplement	([4])
Work Schedules/Home-Based Work		Occasional	May CPS supplement	([4])
Annual Earnings (outgoing rotation groups)		Annual	Annual CPS	([4])
Time Series (Macro) Files				
National Labor Force Data		Monthly	Labstat([5])	([1])
Regional Data		Monthly	Labstat([5])	([1])
Unpublished Tabulations				
National Labor Force Data		Monthly	Microfiche	Free
Regional, State, and Area Data		Monthly	Microfiche	Free

[1] Accessible from the Internet (http://stats.bls.gov:80/newsrels.htm).
[2] About 3 weeks following period of reference.
[3] About 5 weeks after period of reference.
[4] Diskettes ($80); cartridges ($165-$195); tapes ($215-$265); and CD-ROMs ($150).
[5] Electronic access via the Internet (http://stats.bls.gov).

Note: Prices noted above are subject to change.

Chapter 13.
Overview of Data Quality Concepts

INTRODUCTION

It is far easier to put out a figure than to accompany it with a wise and reasoned account of its liability to systematic and fluctuating errors. Yet if the figure is … to serve as the basis of an important decision, the accompanying account may be more important than the figure itself. John W. Tukey (1949, p. 9)

The quality of any estimate based on sample survey data can and should be examined from two perspectives. The first is based on the mathematics of statistical science, and the second stems from the fact that survey measurement is a production process conducted by human beings. From both perspectives, survey estimates are subject to error, and to avoid misusing or reading too much into the data, we should use them only after their potential error of both sorts has been examined relative to the particular use at hand.

In this chapter, we give an overview of these two perspectives on data quality, discuss their relationship to each other from a conceptual viewpoint, and define a number of technical terms. The definitions and discussion are applicable to all sample surveys, not just the Current Population Survey (CPS). Succeeding chapters go into greater detail about the specifics as they relate to CPS.

QUALITY MEASURES IN STATISTICAL SCIENCE

The statistical theory of finite population sampling is based on the concept of repeated sampling under fixed conditions. First, a particular method of selecting a sample and aggregating the data from the sample units into an estimate of the population parameter is specified. The method for sample selection is referred to as the sample design (or just the design). The procedure for producing the estimate is characterized by a mathematical function known as an estimator. After the design and estimator have been determined, a sample is selected and an estimate of the parameter is computed. The difference between the value of the estimate and the population parameter is referred to here as the sampling error, and it will vary from sample to sample (Särndal, Swensson, and Wretman, 1992, p. 16).

Properties of the sample design-estimator methodology are determined by looking at the distribution of estimates that results from taking all possible samples that could be selected using the specified methodology. The expected (or mean) value of the squared sampling errors over all possible samples is known as the mean squared error. The mean squared error is generally accepted as the standard overall measure of the quality of a proposed design-estimator methodology.

The mean value of the individual estimates is referred as the expected value of the estimator. The difference between the expected value of a particular estimator and the value of the population parameter is known as sampling bias. When the bias of the estimator is zero, the estimator is said to be unbiased. The mean value of the squared difference of the values of the individual estimates and the expected value of the estimator is known as the sampling variance of the estimator. The variance measures the magnitude of the variation of the individual estimates about their expected value while the mean squared error measures the magnitude of the variation of the estimates about the value of the population parameter of interest. It can be shown that the mean squared error can be decomposed into the sum of the variance and the square of the bias. Thus, for an unbiased estimator, the variance and the mean squared error are equal.

Quality measures of a design-estimator methodology expressed in this way, that is, based on mathematical expectation assuming repeated sampling, are inherently grounded on the assumption that the process is correct and constant across sample repetitions. Unless the measurement process is uniform across sample repetitions, the mean squared error is not by itself a full measure of the quality of the survey results.

The assumptions associated with being able to compute any mathematical expectation are extremely rigorous and rarely practical in the context of most surveys. For example, the basic formulation for computing the true mean squared error requires that there be a perfect list of all units in the universe population of interest, that all units selected for a sample provide all the requested data, that every interviewer be a clone of an ideal interviewer who follows a predefined script exactly and interacts with all varieties of respondents in precisely the same way, and that all respondents comprehend the questions in the same way and have the same ability to recall from memory the specifics needed to answer the questions.

Recognizing the practical limitations of these assumptions, sampling theorists continue to explore the implications of alternative assumptions which can be expressed in terms of mathematical models. Thus, the mathematical expression for variance has been decomposed in various

ways to yield expressions for statistical properties that include not only sampling variance, but simple response variance (a measure of the variability among the possible responses of a particular respondent over repeated administrations of the same question) (Hansen, Hurwitz, and Bershad, 1961) and correlated response variance, one form of which is interviewer variance (a measure of the variability among responses obtained by different interviewers over repeated administrations). Similarly, when a particular design-estimator fails over repeated sampling to include a particular set of population units in the sampling frame or to ensure that all units provide the required data, bias can be viewed as having components such as coverage bias, unit nonresponse bias, or item nonresponse bias (Groves, 1989). For example, a survey administered solely by telephone could result in coverage bias for estimates relating to the total population if the nontelephone households were different than the telephone households with respect to the characteristic being measured (which almost always occurs).

One common theme of these types of models is the decomposition of total mean squared error into one set of components resulting from the fact that estimates are based on a sample of units rather than the entire population (sampling error) and another set of components due to alternative specifications of procedures for conducting the sample survey (nonsampling error). (Since nonsampling error is defined negatively, it ends up being a catch-all term for all errors other than sampling error.) Conceptually, nonsampling error in the context of statistical science has both variance and bias components. However, when total mean squared error is decomposed mathematically to include a sampling error term and one or more other ''nonsampling error'' terms, it is often difficult to categorize such terms as either variance or bias. The term nonsampling error is used rather loosely in the survey literature to denote mean squared error, variance, or bias in the precise mathematical sense and to imply error in the more general sense of process mistakes (see next section).

Some nonsampling error components which are conceptually known to exist have yet to be expressed in practical mathematical models. Two examples are the bias associated with the use of a particular set of interviewers and the variance associated with the selection of one of the numerous possible sets of questions. In addition, the estimation of many nonsampling errors — and sampling bias — is extremely expensive and difficult or even impossible in practice. The estimation of bias, for example, requires knowledge of the truth, which may be sometimes verifiable from records (e.g., number of hours paid for by employer) but often is not verifiable (e.g., number of hours actually worked). As a consequence, survey organizations typically concentrate on estimating the one component of total mean squared error for which practical methods have been developed — variance.

Since a survey is generally conducted only once with one specific sample of units, it is impossible to compute the actual sampling variance. In simple cases, it is frequently possible to construct an unbiased estimator of variance. In the case of complex surveys like CPS, estimators have been developed which typically rely on the proposition — usually well-grounded — that the variability among estimates based on various subsamples of the one actual sample is a good proxy for the variability among all the possible samples like the one at hand. In the case of CPS, 160 subsamples or replicates are used in variance estimation for the 1990 design. (For more specifics, see Chapter 14.) It is important to note that the estimates of variance resulting from the use of this and similar methods are not merely estimates of sampling variance. The variance estimates include the effects of some nonsampling errors, such as response variance and intra-interviewer correlation. On the other hand, users should be aware of the fact that for some statistics these estimates of standard error might be significant underestimates of total error, an important consideration when making inferences based on survey data.

To draw conclusions from survey data, samplers rely on the theory of finite population sampling from a repeated sampling perspective: If the specified sample design-estimator methodology were implemented repeatedly and the sample size sufficiently large, the probability distribution of the estimates would be very close to a normal distribution.

Thus, one could safely expect 90 percent of the estimates to be within two standard errors of the mean of all possible sample estimates (standard error is the square root of the estimate of variance) (Gonzalez et al., 1975; Moore, 1997). However, one cannot claim that the probability is .90 that the true population value falls in a particular interval. In the case of a biased estimator due to nonresponse, undercoverage, or other types of nonsampling error, confidence intervals may not cover the population parameter at the desired 90 percent rate. In such cases, a standard error estimator may indirectly account for some elements of nonsampling error in addition to sampling error and lead to confidence intervals having greater than the nominal 90 percent coverage. On the other hand, if the bias is substantial, confidence intervals can have less than the desired coverage.

QUALITY MEASURES IN STATISTICAL PROCESS MONITORING

The process of conducting a survey includes numerous steps or components, such as defining concepts, translating concepts into questions, selecting a sample of units from what may be an imperfect list of population units, hiring and training interviewers to ask persons in the sample unit the questions, coding responses to questions into predefined categories, and creating estimates which take into account the fact that not everyone in the population of interest had a chance to be in the sample and not all of those in the sample elected to provide responses. It is a

process where the possibility exists at each step of making a mistake in process specification and in deviating during implementation from the predefined specifications.

For example, we now recognize that the initial labor force question used in CPS for many years (''What were you doing most of last week--...'') was problematic to many respondents (see Chapter 6). Moreover, many interviewers tailored their presentation of the question to particular respondents, for example, saying ''What were you doing most of last week—working, going to school, etc.?'' if the respondent was of school age. Having a problematic question is a mistake in process specification; varying question wording in a way not prespecified is a mistake in process implementation.

Errors or mistakes in process contribute to nonsampling error in that they would contaminate results even if the whole population were surveyed. Parts of the overall survey process which are known to be prone to deviations from the prescribed process specifications and thus could be potential sources of nonsampling error in the CPS are discussed in Chapter 15, along with the procedures put in place to limit their occurrence.

A variety of quality measures have been developed to describe what happens during the survey process. These measures are vital for managers and staff working on a survey as reflecting process quality, but they can also be useful to users of the various products of the survey process — both individual responses and their aggregations into statistics, since they can aid in determining a particular product's potential limitations and whether it is appropriate for the task at hand. Chapter 16 contains a discussion of quality indicators and, in a few cases, their potential relationship to nonsampling errors.

SUMMARY

The quality of estimates made from any survey, including CPS, is a function of innumerable decisions made by designers and implementers. As a general rule of thumb, designers make decisions aimed at minimizing mean squared error within given cost constraints. Practically speaking, statisticians are often compelled to make decisions on sample designs and estimators based on variance alone; however, in the case of CPS, the availability of external population estimates and data on rotation group bias makes it possible to do more than that. Designers of questions and data collection procedures tend to focus on limiting bias and assume the specification of exact question wording and ordering will naturally limit the introduction of variance. Whatever the theoretical focus of the designers, the accomplishment of the goal is heavily dependent upon those responsible for implementing the design.

Implementers of specified survey procedures, like interviewers and respondents, are presumably concentrating on doing the best job they know how to do given time and knowledge constraints. Consequently, process monitoring through quality indicators, such as coverage and response rates, is necessary to determine when additional training or revisions in process specification are needed. Continuing process improvement must be a vital component of survey management if the quality goals set by designers are to be achieved.

REFERENCES

Gonzalez, M. E., Ogus, J. L., Shapiro, G., and Tepping, B. J. (1975), ''Standards for Discussion and Presentation of Errors in Survey and Census Data,'' **Journal of the American Statistical Association,** 70, No. 351, Part II, 5-23.

Groves, R. M. (1989), **Survey Errors and Survey Costs,** New York: John Wiley & Sons.

Hansen, M. H., Hurwitz, W. N., and Bershad, M. A. (1961), ''Measurement Errors in Censuses and Surveys,'' **Bulletin of the International Statistical Institute**, 38(2), pp. 359-374.

Moore, D. S. (1997), **Statistics Concepts and Controversies**, 4th Edition, New York: W. H. Freeman.

Särndal, C., Swensson, B., and Wretman, J. (1992), **Model Assisted Survey Sampling**, New York: Springer-Verlag.

Tukey, J. W. (1949), ''Memorandum on Statistics in the Federal Government,'' **American Statistician**, 3, No. 1, pp. 6-17; No. 2, pp. 12-16.

Chapter 14.
Estimation of Variance

INTRODUCTION

The following two objectives are considered in estimating variances of the major statistics of interest for the Current Population Survey (CPS):

1. Estimate the variance of the survey estimates for use in various statistical analyses.

2. Evaluate the effect of each of the stages of sampling and estimation on the overall precision of the survey estimates for the evaluation of the survey design.

CPS variance estimates take into account the magnitude of the sampling error as well as the effects of some nonsampling errors, such as response variance and intra-interviewer correlation. Certain aspects of the CPS sample design, such as the use of one sample PSU per nonself-representing stratum and the use of systematic sampling within PSUs, make it impossible to obtain a completely unbiased estimate of the total variance. The use of ratio adjustments in the estimation procedure also contributes to this problem. Although imperfect, the current variance estimation procedure is accurate enough for all practical uses of the data, as well as reflecting the effects of the separate stages of sample selection and estimation on the total variance. Variance estimates of selected characteristics and tables, which also show the effects of estimation steps on variances, are presented at the end of this chapter.

VARIANCE ESTIMATES BY THE REPLICATION METHOD

Replication methods are able to provide satisfactory estimates of variance for a wide variety of designs using probability sampling, even when complex estimation procedures are used. This method requires that the sample selection, the collection of data, and the estimation procedures be independently carried through (replicated) several times. The dispersion of the resulting estimates can be used to measure the variance of the full sample.

The Method

Obviously, one would not consider repeating the entire Current Population Survey several times each month simply to obtain variance estimates. A practical alternative is to draw a set of random subsamples from the full sample surveyed each month, using the same principles of selection as are used for the full sample, and to apply the regular CPS estimation procedures to these subsamples. We refer to these subsamples as replicates. The number of replicates to use is based on a trade-off between cost and the reliability of the variance estimator; that is, increasing the number of replicates decreases the variance of the variance estimator.

Prior to the design introduced after the 1970 census, variance estimates were computed using 40 replicates. The replicates were subjected to only the second-stage ratio adjustment for the same age-sex-race categories used for the full sample at the time. The noninterview and first-stage ratio adjustments were not replicated. Even with these simplifications, limited computer capacity allowed for the computation of variances for only 14 characteristics. For the 1970 design, an adaptation of the Keyfitz method of calculating variances was used. These variance estimates were derived using the Taylor approximation, dropping terms with derivatives higher than the first. By 1980, improvements in computer memory capacity allowed for the calculation of variance estimates for many characteristics with replication of all stages of the weighting through compositing. Note that the seasonal adjustment has not been replicated. A study from an earlier design indicated that the seasonal adjustment of CPS estimates had relatively little impact on the variances; however, it is not known what impact this adjustment would have on the current design variances.

Starting with the 1980 design, variances were computed using a modified balanced half-sample approach. The sample was divided to form 48 replicates that retained all the features of the sample design, for example, the stratification and the within-PSU sample selection. For total variance, a pseudo-first-stage design was imposed on the CPS by dividing large self-representing (SR) PSUs into smaller areas called Standard Error Computation Units (SECUs) and combining small nonself-representing (NSR) PSUs into paired strata or pseudostrata. One NSR PSU was selected randomly from each pseudostratum for each replicate. Forming these pseudostrata was necessary since the first stage of the sample design has only one NSR PSU per stratum in sample. However, pairing the original strata for variance estimation purposes creates an upward bias in the variance estimator. For self-representing PSUs each SECU was divided into two panels, and one panel was selected for each replicate. One column of a 48 by 48 Hadamard orthogonal matrix was assigned to each SECU or pseudostratum. The unbiased weights were multiplied by replicate factors of 1.5 for the selected panel and 0.5 for

the other panel in the SR SECU or NSR pseudostratum (Fay, Dippo, and Morganstein, 1984). Thus the full sample was included in each replicate, but differing weights for the half-samples were determined by the matrix. These 48 replicates were processed through all stages of the CPS weighting through compositing. The estimated variance for the characteristic of interest was computed by summing a squared difference between each replicate estimate ($\hat{Y}_r$) and the full sample estimate ($\hat{Y}_0$) The complete formula[1] is

$$Var(\hat{Y}_0) = \frac{4}{48} \sum_{r=1}^{48} (\hat{Y}_r - \hat{Y}_0)^2.$$

Due to costs and computer limitations, variance estimates were calculated for only 13 months (January 1987 through January 1988) and for about 600 estimates at the national level. Replication estimates of variances at the subnational level were not reliable because of the small number of SECUs available (Lent, 1991). Based on the 13 months of variance estimates, generalized sampling errors (explained later) were calculated. (See Wolter 1985 or Fay 1984, 1989 for more details on half-sample replication for variance estimation.)

METHOD FOR ESTIMATING VARIANCE FOR 1990 DESIGN

The general goal of the current variance estimation methodology, the method in use since July 1995, is to produce consistent variances and covariances for each month over the entire life of the design. Periodic maintenance reductions in the sample size and the continuous addition of new construction to the sample complicated the strategy needed to achieve this goal. However, research has shown that variance estimates are not adversely affected as long as the cumulative effect of the reductions is less than 20 percent of the original sample size (Kostanich, 1996). Assigning all future new construction sample to replicates when the variance subsamples are originally defined provides the basis for consistency over time in the variance estimates.

The current approach to estimating the 1990 design variances is called successive difference replication. The theoretical basis for the successive difference method was discussed by Wolter (1984) and extended by Fay and Train (1995) to produce the successive difference replication method used for the CPS. The following is a description of the application of this method. Successive USUs[2] (ultimate

sampling units) formed from adjacent hit strings (see Chapter 3) are paired in the order of their selection to take advantage of the systematic nature of the CPS within-PSU sampling scheme. Each USU usually occurs in two consecutive pairs; for example, (USU1, USU2), (USU2, USU3), (USU3, USU4), etc. A pair then is similar to a SECU in the 1980 design variance methodology. For each USU within a PSU, two pairs (or SECUs) of neighboring USUs are defined based on the order of selection — one with the USU selected before and one with the USU selected after it. This procedure allows USUs adjacent in the sort order to be assigned to the same SECU, thus better reflecting the systematic sampling in our variance estimator. Also, the large increase in the number of SECUs and in the number of replicates (160 vs. 48) over the 1980 design increases the precision of the variance estimator.

Replicate Factors for Total Variance

Total variance is composed of two types of variance, the variance due to sampling of housing units within PSUs (within-PSU variance) and the variance due to the selection of a subset of all NSR PSUs (between-PSU variance). Replicate factors are calculated using a 160 by 160[3] Hadamard orthogonal matrix. To produce estimates of total variance, replicates are formed differently for SR and NSR sample. Note that between-PSU variance cannot be estimated directly using this methodology. Rather, it is the difference between the estimates of total variance and within-PSU variance. NSR strata are combined into pseudostrata within each state, and one NSR PSU from the pseudostratum is randomly assigned to each panel of the replicate as in the 1980 design variance methodology. Replicate factors of 1.5 or 0.5 adjust the weights for the NSR panels. These factors are assigned based on a single row from the Hadamard matrix and are further adjusted to account for the unequal sizes of the original strata within the pseudostratum (Wolter, 1985). In most cases these pseudostrata consist of a pair of strata except where an odd number of strata within a state requires that a triplet be formed. In this case, two rows from the Hadamard matrix are assigned to the pseudostratum resulting in replicate factors of about 0.5, 1.7, and 0.8; or 1.5, 0.3, and 1.2 for the three PSUs. All USUs in a pseudostratum are assigned the same row number(s).

For SR sample, two rows of the Hadamard matrix are assigned to each pair of USUs creating replicate factors, f_r for r = 1,...,160

$$f_{ir} = 1 + (2)^{-\frac{3}{2}} a_{i+1,r} - (2)^{-\frac{3}{2}} a_{i+2,r}$$

where $a_{i,r}$ equals a number in the Hadamard matrix (+1 or -1) for the ith USU in the systematic sample. This formula yields replicate factors of approximately 1.7, 1.0, or 0.3.

---

[1]Usually balanced half-sample replication uses replicate factors of 2 and 0 with the formula,

$$Var(\hat{Y}_o) = \frac{1}{k} \sum_{r=1}^{k} (\hat{Y}_r - \hat{Y}_o)^2.$$

where k is the number of replicates. The factor of 4 in our variance estimator is the result of using replicate factors of 1.5 and 0.5.

[2]An ultimate sampling unit is usually a group of four neighboring housing units.

---

[3]Rows 1 and 81 have been dropped from the matrix.

As in the 1980 methodology, the unbiased weights (baseweight x special weighting factor) are multiplied by the replicate factors to produce unbiased replicate weights. These unbiased replicate weights are further adjusted through the noninterview adjustment, the first-stage ratio adjustment, the second-stage ratio adjustment, and compositing just as the full sample is weighted. A variance estimator for the characteristic of interest is a sum of squared differences between each replicate estimate ($\hat{Y}_r$) and the full sample estimate ($\hat{Y}_o$) The formula is

$$\text{Var}(\hat{Y}_o) = \frac{4}{160} \sum_{r=1}^{160} (\hat{Y}_r - \hat{Y}_o)^2.$$

Note that the replicate factors 1.7, 1.0, and 0.3 for the self-representing portion of the sample were specifically constructed to yield "4" in the above formula in order that the formula remains consistent between SR and NSR areas (Fay and Train, 1995).

Replicate Factors for Within-PSU Variance

The above variance estimator can also be used for within-PSU variance. The same replicate factors used for total variance are applied to SR sample. For NSR sample, alternate row assignments are made for USUs to form pairs of USUs in the same manner that was used for the SR assignments. Thus for within-PSU variance all USUs (both SR and NSR) have replicate factors of approximately 1.7, 1.0, or 0.3.

The successive difference replication method is used to calculate total national variances and within-PSU variances for some states and metropolitan areas. Improved reliability for within-PSU subnational variance estimates is expected due to the increased number of SECUs over the 1980 design. The reliability of between-PSU variance estimates for subnational variance estimates (e.g., state estimates) and the reliability of national estimates of variance should be about the same as for the 1980 design. For more detailed information regarding the formation of replicates, see the internal Census Bureau memorandum (Gunlicks, 1996).

VARIANCES FOR STATE AND LOCAL AREA ESTIMATES

For estimates at the national level, total variances are estimated from the sample data by the successive difference replication method previously described. For local areas that are coextensive with one or more sample PSUs, a variance estimator can be derived using the methods of variance estimation used for the SR portion of the national sample. It is anticipated that these estimates of variance will be more reliable than the 1980 procedure for areas with comparable sample sizes due to the change in methodology. However, estimates for states and areas which have

substantial contributions from NSR sample areas have variance estimation problems that are much more difficult to resolve.

Complicating Factors for State Variances

Most states contain a small number of NSR sample PSUs. Pairing them into pseudostrata reduces the number of NSR SECUs still further and increases reliability problems. Also, the component of variance resulting from sampling PSUs can be more important for state estimates than for national estimates in states where the proportion of the population in NSR strata is larger than the national average.

Further, creating pseudostrata for variance estimation purposes introduces a between-stratum variance component that is not in the sample design, causing over-estimation of the true variance. The between-PSU variance which includes the between-stratum component is relatively small at the national level for most characteristics, but it can be much larger at the state level (Gunlicks, 1993; Corteville, 1996). Thus, this additional component should be accounted for when estimating state variances.

Research is in progress to produce improved state and local variances utilizing the within-PSU variances obtained from successive difference replication and perhaps some modeling techniques. These variances will be provided as they become available.

GENERALIZING VARIANCES

With some exceptions, the standard errors provided with published reports and public data files are based on generalized variance functions (GVFs). The GVF is a simple model that expresses the variance as a function of the expected value of the survey estimate. The parameters of the model are estimated using the direct replicate variances discussed above. These models provide a relatively easy way to obtain an approximate standard error on numerous characteristics.

Why Generalized Standard Errors Are Used

It would be possible to compute and show an estimate of the standard error based on the survey data for each estimate in a report, but there are a number of reasons why this is not done. A presentation of the individual standard errors would be of limited use, since one could not possibly predict all of the combinations of results that may be of interest to data users. Also, for estimates of differences and ratios that users may compute, the published standard errors would not account for the correlation between the estimates.

Most importantly, variance estimates are based on sample data and have variances of their own. The variance estimate for a survey estimate for a particular month generally

has less precision than the survey estimate itself. This means that the estimates of variance for the same characteristic may vary considerably from month-to-month or for related characteristics (that might actually have nearly the same level of precision) in a given month. Therefore, some method of stabilizing these estimates of variance, for example, by generalization or by averaging over time, is needed to improve their reliability.

Experience has shown that certain groups of CPS estimates have a similar relationship between their variance and expected value. Modeling or generalization may provide more stable variance estimates by taking advantage of these similarities.

The Generalization Method

The GVF that is used to estimate the variance of an estimated population total, X, is of the form

$$Var(\hat{X}) = aX^2 + bX \qquad (14.1)$$

where a and b are two parameters estimated using least squares regression. The rationale for this form of the GVF model is the assumption that the variance of $\hat{X}$ can be expressed as the product of the variance from a simple random sample for a binomial random variable and a "design effect." The design effect (deff) accounts for the effect of a complex sample design relative to a simple random sample. Defining P = X/N as the proportion of the population having the characteristic X, where N is the population size, and Q = 1-P, the variance of the estimated total $\hat{X}$, based on a sample of n individuals from the population, is

$$Var(\hat{X}) = \frac{N^2 PQ(deff)}{n} \qquad (14.2)$$

This can be written as

$$Var(\hat{X}) = - (deff)\left(\frac{N}{n}\right)\left(\frac{X^2}{N}\right) + (deff)\left(\frac{N}{n}\right) X .$$

Letting

$$a = -\frac{b}{N}$$

and

$$b = \frac{(deff)N}{n}$$

gives the functional form

$$Var(\hat{X}) = aX^2 + bX.$$

We choose

$$a = -\frac{b}{N}$$

where N is a control total so that the variance will be zero when X = N.

In generalizing variances, all estimates that follow a common model such as 14.1 (usually the same characteristics for selected demographic or geographic subgroups) are grouped together. This should give us estimates in the same group that have similar design effects. These design effects incorporate the effect of the estimation procedures, particularly the second stage, as well as the effect of the sample design. In practice, the characteristics should be clustered similarly by PSU, by USU, and among persons within housing units. For example, estimates of total persons classified by a characteristic of the housing unit or of the household, such as the total urban population, number of recent migrants, or persons of Hispanic[4] origin, would tend to have fairly large design effects. The reason for this is that these characteristics usually appear among all persons in the sample household and often among all households in the USU as well. On the other hand, lower design effects would result for estimates of labor force status, education, marital status, or detailed age categories, since these characteristics tend to vary among members of the same household and among households within a USU.

Notice also that for many subpopulations of interest, N is a control total used in the second-stage ratio adjustment. In these subpopulations, as X approaches N, the variance of X approaches zero, since the second-stage ratio adjustment guarantees that these sample population estimates match independent population controls[5] (Chapter 10). The GVF model satisfies this condition. This generalized variance model has been used since 1947 for the CPS and its supplements, although alternatives have been suggested and investigated from time to time (Valliant, 1987). The model has been used to estimate standard errors of means or totals. Variances of estimates based on continuous variables (e.g., aggregate expenditures, amount of income, etc.) would likely fit a different functional form better.

The parameters, a and b, are estimated by use of the model for relative variance

$$Vx^2 = a + \frac{b}{X},$$

where the relative variance (Vx^2) is the variance divided by the square of the expected value of the estimate. The a and b parameters are estimated by fitting a model to a group of related estimates and their estimated relative variances. The relative variances are calculated using the successive difference replication method.

The model fitting technique is an iterative weighted least squares procedure, where the weight is the inverse of the square of the predicted relative variance. The use of these weights prevents items with large relative variances from unduly influencing the estimates of the a and b parameters.

---

[4]Hispanics may be of any race.
[5]The variance estimator assumes no variance on control totals, even though they are estimates.

Usually at least a year's worth of data are used in this model fitting process and each group of items should comprise at least 20 characteristics with their relative variances, although occasionally fewer characteristics are used.

Direct estimates of relative variances are required for estimates covering a wide range, so that observations are available to ensure a good fit of the model at high, low, and intermediate levels of the estimates. It must be remembered that, by using a model to estimate the relative variance of an estimate in this way, we introduce some error, since the model may substantially and erroneously modify some legitimately extreme values. Generalized variances are computed for estimates of month-to-month change as well as for estimates of monthly levels. Periodically, the a and b parameters are updated to reflect changes in the levels of the population totals or changes in the ratio (N/n) which result from sample reductions. This can be done without recomputing direct estimates of variances as long as the sample design and estimation procedures are essentially unchanged (Kostanich, 1996).

How the Relative Variance Function is Used

After the parameters a and b of expression (14.1) are determined, it is a simple matter to construct a table of standard errors of estimates for publication with a report. In practice, such tables show the standard errors that are appropriate for specific estimates, and the user is instructed to interpolate for figures not explicitly shown in the table. However, many reports present a list of the parameters, enabling data users to compute generalized variance estimates directly. A good example is a recent monthly issue of **Employment and Earnings** (U.S. Department of Labor) from which the following table was taken.

Example:
The approximate standard error, $s_{\hat{X}}$, of an estimated monthly level $\hat{X}$ can be obtained with a and b from the above table and the formula

$$s_{\hat{X}} = \sqrt{a\hat{X}^2 + b\hat{X}}$$

Assume that in a given month there are an estimated 6 million unemployed men in the civilian labor force ($\hat{X} = 6,000,000$). From Table 14–1

$$a = -0.000015749 \text{ and } b = 2464.91 \text{ , so}$$

$$s_x = \sqrt{(-0.000015749)(6,000,000)^2 + (2464.91)(6,000,000)} \approx 119,000$$

An approximate 90-percent confidence interval for the monthly estimate of unemployed men is between 5,810,000 and 6,190,000 [or $6,000,000 \pm 1.6(119,000)$].

VARIANCE ESTIMATES TO DETERMINE OPTIMUM SURVEY DESIGN

The sample design and the estimation procedures used in the CPS have changed many times since the start of the

Table 14–1. **Parameters for Computation of Standard Errors for Estimates of Monthly Levels[1]**

Characteristic	a	b
Unemployment:		
Total or White....................	−0.000015749	2464.91
Black	0.000191460	2621.89
Hispanic origin[2]	−0.000098631	2704.53

[1]Parameters reflect variances for 1995. See Appendix H for variance estimates since January 1996.
[2]Hispanics may be of any race.

survey. These changes result from a continuing program of research and development with the objective of optimizing the use of the techniques and resources which are available at a given time. Changes also occur when reliability requirements are revised. Estimates of the components of variance attributable to each of the several stages of sampling and the effect of the separate steps in the estimation procedure on the variance of the estimate are needed to develop an efficient sample design.

The tables which follow show variance estimates computed, using replication methods, by type (total and within-PSU), and by stage of estimation. Averages over 6 months have been used to improve the reliability of the estimated monthly variances. The 6-month period, July-December 1995, was used for estimation because the sample design was unchanged throughout the period. Appendix H provides similar information on variances and components of variance after the January 1996 sample reduction.

Variance Components Due to Stages of Sampling

Table 14–2 indicates, for the first- and second-stage combined (FSC) estimate, how the several stages of sampling affect the total variance of each of the given characteristics. The FSC estimate is the estimate after the first- and second-stage ratio adjustments are applied. Within-PSU variance and total variance are computed as described earlier in this chapter. Between-PSU variance is estimated by subtracting the within-PSU variance from the total variance. Due to variation of the variance estimates, the between-PSU variance is sometimes negative. The far right two columns of Table 14–2 show the percentage within-PSU variance and the percentage between-PSU variance in the total variance estimate.

For all characteristics shown in Table 14–2, the proportion of the total variance due to sampling housing units within PSUs (within-PSU variance) is larger than that due to sampling a subset of NSR PSUs (between-PSU variance). In fact, for most of the characteristics shown, the within-PSU component accounts for over 90 percent of the total variance. For civilian labor force and not in labor force characteristics, almost all of the variance is due to sampling housing units within PSUs. For the total and White employed in agriculture, the within-PSU component still

Table 14–2. **Components of Variance for FSC Monthly Estimates**

[Monthly averages: July - Dec. 1995]

Civilian-noninstitutional population 16 years old and over	FSC[1] estimate (x 10^6)	Standard error (x 10^5)	Coefficient of variation (percent)	Percent of total variance	
				Within	Between
Unemployed...................................	7.30	1.38	1.89	98.4	1.6
White..	5.31	1.19	2.24	98.1	1.9
Black...	1.57	0.68	4.32	100.4	−0.4
Hispanic origin[2]	1.15	0.63	5.51	94.2	5.8
Teenage, 16-19.............................	1.35	0.57	4.25	98.7	1.3
Employed - Agriculture	3.47	1.49	4.30	63.2	36.8
White..	3.23	1.43	4.42	66.1	33.9
Black...	0.10	0.19	18.52	105.9	−5.9
Hispanic origin[2]...........................	0.60	0.73	12.27	83.4	16.6
Teenage, 16-19.............................	0.29	0.28	9.50	97.0	3.0
Employed - Nonagriculture.	122.34	3.34	0.27	92.8	7.2
White..	104.05	3.05	0.29	94.4	5.6
Black...	13.29	1.30	0.98	100.3	−0.3
Hispanic origin[2]...........................	10.72	1.43	1.34	86.3	13.7
Teenage, 16-19.............................	6.40	0.96	1.50	108.1	−8.1
Civilian labor force	133.11	2.89	0.22	99.0	1.0
White..	112.58	2.61	0.23	99.3	0.7
Black...	14.97	1.18	0.79	99.9	0.1
Hispanic origin[2]...........................	12.47	1.09	0.87	99.2	0.8
Teenage, 16-19.............................	8.04	0.98	1.22	103.5	−3.5
Not in labor force	65.96	2.89	0.44	99.0	1.0
White..	54.67	2.61	0.48	99.3	0.7
Black...	8.37	1.18	1.41	99.9	0.1
Hispanic origin[2]...........................	6.31	1.09	1.73	99.2	0.8
Teenage, 16-19.............................	6.61	1.00	1.52	101.5	−1.5

[1]First- and second-stage combined (FSC).
[2]Hispanics may be of any race.

accounts for 60 to 70 percent of the total variance, while the between-PSU component accounts for the remaining 30 to 40 percent. The between-PSU component is also larger (between 10 and 20 percent) for employed in agriculture, Hispanic origin, and employed in nonagriculture, Hispanic origin than for other characteristics shown in the table.

Because the FSC estimate of total civilian labor force and total not in labor force must add to the independent population controls (which are assumed to have no variance), the standard errors and variance components for these estimated totals are the same. They are not the same for the teenage category because the estimates do not completely converge to the control totals during the second-stage ratio adjustment.

TOTAL VARIANCES AS AFFECTED BY ESTIMATION

Table 14–3 shows how the separate estimation steps affect the variance of estimated levels by presenting ratios of relative variances. It is more instructive to compare ratios of relative variances than the variances themselves, since the various stages of estimation can affect both the level of an estimate and its variance (Hanson, 1978; Train, Cahoon, and Makens, 1978). The unbiased estimate uses the baseweight with special weighting factors applied. The

noninterview estimate includes the baseweights, the special weighting adjustment, and the noninterview adjustment. The first- and second-stage ratio adjustments are used in the estimate for the first- and second-stage combined (FSC).

In Table 14–3, the figures for unemployed show, for example, that the relative variance of the FSC estimate of level is 3.590×10^{-4} (equal to the square of the coefficient of variation in Table 14–2). The relative variance of the unbiased estimate for this characteristic would be 1.06 times as large. If the noninterview stage of estimation is also included, the relative variance is only 1.05 times the size of the relative variance for the FSC estimate of level. Including the first stage of estimation raises the relative variance factor to 1.13. The relative variance for total unemployed after applying the second stage without the first stage is about the same as the relative variance that results from applying the first- and second-stage adjustments.

The relative variance as shown in the last column of this table illustrates that the first-stage ratio adjustment has little effect on the variance of national level characteristics in the context of the overall estimation process. The first-stage adjustment was implemented in order to reduce variances of state-level estimates. Its actual effect on these state-level estimates has not yet been examined.

The second-stage adjustment, however, appears to greatly reduce the total variance, as intended. This is

Table 14–3. Effects of Weighting Stages on Monthly Relative Variance Factors

[Monthly averages: July - Dec. 1995]

Civilian-noninstitutional population 16 years old and over	Relative variance of FSC estimate of level (x 10[-4])	Relative variance factor[1]			
		Unbiased estimator	Unbiased estimator with–		
			NI[2]	NI & FS[3]	NI & SS[4]
Unemployed.	3.590	1.06	1.05	1.13	1.00
White.	5.004	1.11	1.11	1.16	1.00
Black.	18.626	1.30	1.28	1.28	1.00
Hispanic origin[5].	30.347	1.14	1.14	1.19	1.00
Teenage, 16-19.	18.092	1.13	1.13	1.14	0.99
Employed - Agriculture	18.460	0.98	0.97	1.04	0.96
White.	19.514	0.98	0.97	1.03	0.96
Black.	342.808	1.03	1.02	1.04	1.01
Hispanic origin[5].	150.548	1.01	0.98	1.06	0.95
Teenage, 16-19.	90.235	1.08	1.08	1.10	1.00
Employed - Nonagriculture	0.075	4.71	4.47	6.85	0.96
White.	0.086	5.54	5.44	7.55	0.97
Black.	0.957	6.73	6.60	6.61	1.00
Hispanic origin[5].	1.783	3.19	3.18	3.27	0.95
Teenage, 16-19.	2.264	2.03	2.00	2.05	1.00
Civilian labor force	0.047	7.28	6.84	10.98	0.99
White.	0.054	8.96	8.74	12.33	0.99
Black.	0.623	9.87	9.62	9.64	1.01
Hispanic origin[5].	0.764	7.03	6.95	7.41	0.99
Teenage, 16-19.	1.497	2.69	2.64	2.74	0.99
Not in labor force	0.192	2.79	2.67	4.15	0.99
White.	0.228	3.22	3.16	4.50	0.99
Black.	1.994	4.21	4.11	3.88	1.01
Hispanic origin[5].	2.984	2.99	2.92	3.09	0.99
Teenage, 16-19.	2.312	2.04	2.03	2.23	1.00

[1]Relative variance factor is the ratio of the relative variance of the specified level to the relative variance of the FSC level.
[2]NI = Noninterview.
[3]FS = First-stage.
[4]SS = Second-stage.
[5]Hispanics may be of any race.

especially true for characteristics that comprise high proportions of age, sex, or race/ethnicity subclasses, such as White, Black, or Hispanic persons in the civilian labor force or employed in nonagricultural industries. Without the second stage, the relative variances of these characteristics would be 5, 6, or even 11 to 12 times as large. For smaller groups, such as unemployed and employed in agriculture, the second-stage adjustment does not have as dramatic an effect.

After the second-stage ratio adjustment, a composite estimator is used to improve estimates of month-to-month change by taking advantage of the 75 percent of the total sample that continues from the previous month (see Chapter 10). Table 14–4 compares the variance and relative variance of the composited estimates of level to those of the FSC estimates. For example, the estimated variance of the composited estimate of unemployed persons of Hispanic origin is 3.659×10^9. The variance factor for this characteristic is 0.92, implying that the variance of the composited estimate is 92 percent of the variance of the estimate after the first and second stages. The relative variance, which takes into account the estimate of the number of people with this characteristic, is 0.94 times the size of the relative variance of the FSC estimate. The two factors are similar for most characteristics, indicating that compositing tends to have a small effect on the level of most estimates.

DESIGN EFFECTS

Table 14-5 shows the design effects for the total variance for selected labor force characteristics. A design effect (deff) is the ratio of the variance from complex sample design or a sophisticated estimation method to the variance of a simple random sample design. The design effects in this table were computed by solving equation 14.2 for deff and replacing N/n in the formula with an estimate of the national sampling interval. Estimates of P and Q were obtained from the 6 months of data.

For unemployed, the design effect is 1.314 for the uncomposited (FSC) estimate and 1.229 for the composited estimate. This means that, for the same number of sample cases, the design of the CPS (including the sample selection, weighting, and compositing) increases the variance by about 23 percentage points over the variance of an unbiased estimate based on a simple random sample. On the other hand, for the civilian labor force the design of the CPS decreases the variance by about 24 percentage points. Note that the design effects for composited estimates are generally lower than those for the FSC estimates indicating again the tendency of the compositing to reduce the variance of most estimates.

Table 14–4. **Effect of Compositing on Monthly Variance and Relative Variance Factors**

[Monthly averages: July - Dec. 1995]

Civilian-noninstitutional population 16 years old and over	Variance of composited estimate of level (x 10^9)	Variance factor[1]	Relative variance factor[2]
Unemployed	17.690	0.93	0.94
White	13.003	0.93	0.94
Black	4.448	0.97	1.00
Hispanic origin[3]	3.659	0.92	0.94
Teenage, 16-19	3.279	1.00	1.01
Employed - Agriculture	20.789	0.93	0.94
White	18.781	0.93	0.93
Black	0.302	0.85	0.95
Hispanic origin[3]	4.916	0.91	0.88
Teenage, 16-19	0.713	0.92	0.93
Employed - Nonagriculture	94.359	0.85	0.84
White	79.145	0.85	0.85
Black	14.489	0.85	0.85
Hispanic origin[3]	18.466	0.90	0.90
Teenage, 16-19	8.508	0.92	0.92
Civilian labor force	69.497	0.83	0.83
White	58.329	0.85	0.85
Black	12.037	0.86	0.87
Hispanic origin[3]	10.448	0.88	0.88
Teenage, 16-19	8.607	0.89	0.89
Not in labor force	69.497	0.83	0.83
White	58.329	0.85	0.85
Black	12.037	0.86	0.85
Hispanic origin[3]	10.448	0.88	0.88
Teenage, 16-19	9.004	0.89	0.88

[1]Variance factor is the ratio of the variance of a composited estimate to the variance of an FSC estimate.
[2]Relative variance factor is the ratio of the relative variance of a composited estimate to the relative variance of an FSC estimate.
[3]Hispanics may be of any race.

Note: Composite formula constants used: K = 0.4 and A = 0.2.

Table 14–5. **Design Effects for Total Monthly Variances**

[Monthly averages: July - Dec. 1995]

Civilian-noninstitutional population 16 years old and over	Design effects for total variance	
	After first and second stages	After compositing
Unemployed	1.314	1.229
White	1.318	1.227
Black	1.424	1.403
Hispanic origin[1]	1.689	1.574
Teenage, 16-19	1.183	1.191
Employed - Agriculture	3.153	2.958
White	3.091	2.875
Black	1.680	1.516
Hispanic origin[1]	4.370	3.903
Teenage, 16-19	1.278	1.188
Employed - Nonagriculture	1.144	0.966
White	0.904	0.770
Black	0.659	0.564
Hispanic origin[1]	0.976	0.880
Teenage, 16-19	0.723	0.664
Civilian labor force	0.913	0.760
White	0.672	0.576
Black	0.487	0.421
Hispanic origin[1]	0.491	0.433
Teenage, 16-19	0.606	0.540
Not in labor force	0.913	0.760
White	0.829	0.709
Black	0.841	0.723
Hispanic origin[1]	0.939	0.824
Teenage, 16-19	0.763	0.679

[1]Hispanics may be of any race.

REFERENCES

Corteville, J. (1996), "State Between-PSU Variances and Other Useful Information for the CPS 1990 Sample Design (VAR90-16)," Memorandum for Documentation, March 6th, Demographic Statistical Methods Division, U. S. Census Bureau.

Fay, R.E., (1984) "Some Properties of Estimates of Variance Based on Replication Methods," **Proceedings of the Section on Survey Research Methods**, American Statistical Association, pp. 495-500.

Fay, R. E., (1989) "Theory and Application of Replicate Weighting for Variance Calculations," **Proceedings of the Section on Survey Research Methods,** American Statistical Association, pp. 212-217.

Fay, R., Dippo, C., and Morganstein, D. (1984), "Computing Variances From Complex Samples With Replicate Weights," **Proceedings of the Section on Survey Research Methods,** American Statistical Association, pp. 489-494.

Fay, R., and Train, G. (1995), "Aspects of Survey and Model-Based Postcensal Estimation of Income and Poverty Characteristics for States and Counties," **Proceedings of the Section on Government Statistics,** American Statistical Association, pp. 154-159.

Gunlicks, C. (1993), "Overview of 1990 CPS PSU Stratification (S_S90_DC-11)," February 24, Memorandum for Documentation, Demographic Statistical Methods Division, U.S. Census Bureau.

Gunlicks, C. (1996), "1990 Replicate Variance System (VAR90-20)," Internal Memorandum for Documentation, June 4th, Demographic Statistical Methods Division, U.S. Census Bureau.

Hanson, R. H. (1978), The Current Population Survey: **Design and Methodology, Technical Paper 40,** Washington, DC: Government Printing Office.

Kostanich, D. (1996), "Proposal for Assigning Variance Codes for the 1990 CPS Design (VAR90-22)," Internal Memorandum to BLS/Census Variance Estimation Subcommittee, June 17th, Demographic Statistical Methods Division, U.S. Census Bureau.

Kostanich, D., (1996), "Revised Standard Error Parameters and Tables for Labor Force Estimates: 1994-1995 (VAR80-6)," Internal Memorandum for Documentation, February 23rd, Demographic Statistical Methods Division, U.S. Census Bureau.

Lent, J. (1991), "Variance Estimation for Current Population Survey Small Area Estimates," **Proceedings of the Section on Survey Research Methods,** American Statistical Association, pp. 11-20.

Train, G., Cahoon, L., and Makens, P. (1978), "The Current Population Survey Variances, Inter-Relationships, and Design Effects," **Proceedings of the Section on Survey Research Methods,** American Statistical Association, pp. 443-448.

U.S. Dept. of Labor, Bureau of Labor Statistics, **Employment and Earnings, 42,** p.152, Washington, DC: Government Printing Office.

Valliant, R. (1987), "Generalized Variance Functions in Stratified Two-Stage Sampling," **Journal of the American Statistical Association**, 82, pp. 499-508.

Wolter, K. (1984), "An Investigation of Some Estimators of Variance for Systematic Sampling," **Journal of the American Statistical Association**, 79, pp. 781-790.

Wolter, K. (1985), **Introduction to Variance Estimation,** New York: Springer-Verlag.

Chapter 15.
Sources and Controls on Nonsampling Error

INTRODUCTION

Nonsampling error can enter the survey process at any point or stage, and many of these errors are not readily identifiable. Nevertheless, the presence of these errors can affect both the bias and variance components of the total survey error. The effect of nonsampling error on the estimates is difficult to measure accurately. For this reason, the most appropriate strategy is to examine the potential sources of nonsampling error and take steps to prevent these errors from entering the survey process. This chapter discusses the various sources of nonsampling error and the measures taken to control their presence in the Current Population Survey (CPS).

Many sources of nonsampling error were studied and well-charted in the pre-January 1994 CPS (e.g., see Brooks and Bailar, 1978; McCarthy, 1978), but the full extent of nonsampling error in the post-January 1994 CPS is not nearly so well examined. It is also unclear how much previous research on nonsampling error in the CPS is applicable to the redesigned instrument with computerized data collection. Although the intent of the new questionnaire and automation was to reduce nonsampling error, new sources of error may emanate from these methodological advances. Some potential sources of nonsampling error can be investigated with the available CPS and reinterview data, but there is only about 2 years of data available for analyses and comparisons at the time of this writing, and this may not be sufficient for a full analysis.

Nevertheless, it is clear that there are two main types of nonsampling error in the CPS. The first is error imported from other frames or sources of information, such as Decennial Census omissions. The second type is considered preventable, such as when the sample is not completely representative of the intended population, within-household omissions, respondents not providing true answers to a questionnaire item, or errors produced during the processing of the survey data.

Chapter 16 discusses the presence of CPS nonsampling error, but not the effect of the error on the estimates. The present chapter, however, focuses on the sources and operational efforts used to control the occurrence of error in the survey processes. Each section discusses a procedure aimed to reduce either coverage, nonresponse, response, or data processing errors. Despite the effort to treat each control as a separate entity, they nonetheless affect the survey in a general way. For example, training, even if focused on a specific problem, can have broadreaching total survey effects.

An understanding of the sources and causes of nonsampling error is necessary to design procedures to control the effects of these errors on the estimates. An important source of nonsampling error that is difficult to understand and control arises out of the interaction of the people involved in the survey. The methods by which administrators, analysts, programmers, and field personnel interact and exchange information are a potential source of nonsampling error. One attempt to reduce this source of nonsampling error is demonstrated by the use of Deming's quality management procedures in the 1990 sample redesign (Waite, 1991). It emphasizes the relationship between process design and quality improvement. It is a nonbureaucratic approach that has all involved persons working together toward a common goal. Each person is accountable and responsible for the verification of the consistency, reliability, and accuracy of his/her output before it is used as input to a succeeding process or stage. The quality of all previous processes affects the quality of all succeeding processes. Communication across all stages is enhanced by the inclusion of the entire staff involved in the operations, planning, and implementation of the processes up-front, even before the writing of specifications.

Included in the notions of interaction and exchange of information across all personnel levels and processes are feedback, continuous improvement, the introduction of automation and computerization, and training. Feedback, when accompanied by corrective measures, is vital and must flow in all directions between the field representative, the regional office, headquarters, and the data preparation office in Jeffersonville, Indiana. (Focus groups are an example of an effective source of feedback.) Continuous improvement in every process and stage of the CPS is a requirement, not an option. Also, automation and computerization are major factors in the control of nonsampling error, greatly improving uniformity of operations, as well as data access and control. The degree of automation is continuously increasing, thus affecting the amount and allocation of responsibilities. Finally, training is a continuing task in each regional office, the Jeffersonville office, and headquarters, and it is a major factor in controlling the occurrence of nonsampling error. (See Appendix F for a more detailed discussion on training procedures in the regional offices.)

This chapter deals with many sources and many controls for nonsampling error. It covers coverage error, error from nonresponse, response error, and processing errors. Many of these types of errors interact, and they can occur

anywhere in the survey process. Although it is believed that some of the nonsampling errors are small, their full effects on the survey estimates are unknown. The CPS attempts to prevent such errors from entering the survey process or tries to keep them as small as possible.

SOURCES OF COVERAGE ERROR

Coverage error exists when a survey does not completely represent the population of interest. When conducting a sample survey, a primary goal is to give every unit (for example, person or housing unit) in the target universe a known probability of being selected into the sample. When this occurs, the survey is said to have 100 percent coverage. This is rarely the case, however, since errors can enter the system during almost any phase of the survey process. A bias in the survey estimates results if characteristics of units erroneously included or excluded from the survey differ from those correctly included in the survey. Historically in the CPS, the net effect of coverage errors has been a loss in population (resulting from undercoverage).

The primary sources of CPS coverage error are:

1. **Frame omission.** Erroneous frame omission of housing units occurs when the list of addresses is incomplete. This can occur in all of the four sampling frames (i.e., unit, permit, area, and group quarters; see Chapter 3). Since these erroneously omitted housing units cannot be sampled, undercoverage of the target population results. Reasons for frame omissions are:

   • Census Address List: The census address list may be incomplete or some units are not locatable. This occurs when the census lister fails to canvass an area thoroughly or misses units in multiunit structures.

   • New Construction: Some areas of the country are not covered by building permit offices and are not recanvassed (unit frame) to pick up new housing. New housing units in these areas have zero probability of selection.

   • Mobile Homes: Mobile homes that move into areas that are not recanvassed (unit frame) also have a zero probability of selection.

2. **Misclassification of housing units.** Erroneous frame inclusions of housing units occur when housing units outside an appropriate area boundary are misclassified as inside the boundary. Other erroneous inclusions occur when a single unit is recorded as two units through faulty application of the housing unit definition.

   Since erroneously included housing units can be sampled, and if the errors are not detected and corrected, overcoverage of the target population results.

   Misclassification also occurs when an occupied housing unit is treated as vacant.

3. **Within-housing unit omissions.** Undercoverage of persons can arise from failure to list all usual residents of a housing unit on the questionnaire and from misclassifying a household member as a nonmember.

4. **Within-housing unit inclusions.** Overcoverage can occur because of the erroneous inclusion of persons on the roster for a housing unit. For instance, persons with a usual residence elsewhere are treated as members of the sample housing unit.

Other sources of coverage error are omission of homeless persons from the frame, unlocatable addresses from building permit lists, conversion of nonresidential to residential units, and missed housing units due to the start dates for the sampling of permits.[1] Newly constructed group quarters are another possible source of undercoverage. If noticed on the permit lists, the general rule is to purposely remove these group quarters; if not noticed and the field representative discovers a newly constructed group quarters, the case is stricken from the roster of sample addresses and never visited again for CPS. For example, dormitories are group quarters in the 1990 decennial census and are included in the group quarters frame. They are picked up in the other frames, not in the permit frame.

The sampling frames are designed to limit missed housing units in the census frame from affecting coverage of the CPS frames (see Chapter 3). Through various studies, it has been determined that the number of housing unit misses is small; this is because of the way the sampling frames are designed and because of the routines in place to ensure that the listing of addresses is performed correctly. A measure of potential undercoverage is the coverage ratio which incorporates the completeness of the coverage because of frame omissions, misclassification of housing units, and within-housing unit omissions and inclusions. For example, the coverage ratio for the total population, at least 16 years of age, is approximately 92 percent (January 1997). In general, there is low coverage for Blacks (about 83 percent), with Black males of ages 25-34 having the lowest coverage. Females have higher coverage. See Chapter 16 for further discussion of coverage error.

CONTROLLING COVERAGE ERROR

This section focuses on those processes during the sample design and sample preparation stages which aim to control housing unit omissions. Other sources of undercoverage can be viewed as response error and are described in a later section of this chapter.

---

[1]Sampling of permits began with those issued in 1989; the month varied depending on the size of structure and region. Housing units whose permits were issued before the start month in 1989 and not built by the time of the census may be missed.

Sample Design

The CPS sample selection is coordinated with the samples of nine other demographic surveys. These surveys undergo the same verification processes, so the discussion in the rest of this section is in no way unique to the CPS.

Sampling verification basically consists of testing and production. They are different in that the intent of the testing phase is to design the processes better to improve quality, and the intent of the production phase is to verify or inspect quality to improve the processes. Both phases are coordinated across the various surveys. Sampling intervals, expected sample sizes, and the premise that a housing unit is selected for only one survey for 10 years are continually verified.

The testing phase tests and verifies the various sampling programs before any sample is selected, ensuring that the programs work individually and collectively. Smaller data sets with unusual observations test the performance of the system in extreme situations. For example, unusual circumstances encountered in the 1990 decennial census are used to test the 1990 CPS sampling procedures. These circumstances include PSUs with no sample, crews of maritime vessels, group quarters, and Indian Reservations. Programs are then written to verify that procedures are in place to handle the problems. Also, actual data from the 1990 decennial census are used to test and verify the sampling processes.

The production phase of sampling verification treats output verification, especially expected sample sizes, focusing on whether the system ran successfully on actual data. The following are illustrations and do not represent any particular priority or chronology:

1. PSU probabilities of selection are verified; for example, their sum should be one.

2. Edits are done on file contents, checking for blank fields, out-of-range data, etc.

3. When applicable, the files are coded to check for consistency; for example, the number of PSUs "in" should equal the number of PSUs "out" at PSU selection.

4. Information is centralized; for example, the sampling rates for all states and across all surveys are located in one parameter file.

5. As a form of overall consistency verification, the output at several stages of sampling is compared to that of the former CPS design.

Sample Preparation

Sample preparation activities, in contrast to those of sample design, are ongoing. The listing sheet review and check-in of sample are monthly production processes, and the listing check is an ongoing field process.

Listing sheet review. The listing sheet review is a check on each month's listings to keep the nonsampling error as low as possible. Its aim is to ensure that the interviews will be conducted at the correct units and that all units have one and only one chance of selection. This review also plays a role in the verification of the sampling, subsampling, and relisting. At present, the timing of the review is such that not all errors detected in the listing sheet review are able to be corrected in time for incorporation into CPS. However, automation of the review is a major advancement towards improving the timing of the process and, thus, a more accurate CPS frame.

Listing sheet reviews for the unit, permit, area, and group quarters samples are performed in varying degrees by both the regional offices and the Jeffersonville office. The fact that these reviews are performed across different parts of the Census Bureau organization is, in itself, a means of controlling nonsampling error.

In terms of the reviews performed by the regional offices, of utmost importance is the review of field representative explanations made in the remarks column or footnotes section of the listing sheet. Other processes that the regional offices follow to verify and correct the listing sheets for samples from each type of frame follow.

After a field representative makes an initial visit either to a multiunit address from a unit frame or to a sample from a permit frame, the Unit/Permit Listing Sheet is reviewed. This review occurs the first time the basic address or sample from a permit frame is in sample for any survey. However, if major changes occur at an address on a subsequent visit, a review of the revised listing is done. If there is evidence that the field representative encountered a special or unusual situation, the materials are compared to the instructions in the Listing and Coverage Manual for Field Representatives (see Chapter 4) to ensure that the situation was handled correctly. Depending on whether it is a permit frame sample or a multiunit address from a unit frame, the following are verified:

1. Were correct entries made on the listing sheet when either fewer or more units were listed than expected?

2. Was CPS the first survey to interview at the multiunit address?

3. Were new listing sheets prepared?

4. Were no units listed?

5. Was an extra unit discovered?

6. Was a sample unit added without a serial number?

7. Did the unit designation change?

8. Was a sample unit demolished or condemned?

Errors and omissions are corrected and brought to the attention of the field representative. This may involve contacting the field representative for more information.

The regional office review of Area and Group Quarters Listing Sheets is performed on newly listed or updated samples from area or group quarters frames. Basically, this review is for completeness of identification.

As stated previously, listing sheet reviews are also performed by the Jeffersonville office. The following are activities involved in the reviews for the unit sample:

1. Use the mailout control to check-in listing sheets from the regional offices and followup any that are missing.

2. If subsampling was done, review for accuracy.

3. If any large multiunit addresses were relisted, review for accuracy.

4. If there were any address changes, key the changes into the address updates system for future samples at the same basic address.

5. Keep tallies by regional office and survey of when any of situations 2 through 4 above were encountered; provide headquarters with these tallies on a quarterly basis.

In terms of the review of listing sheets for permit frame samples, Jeffersonville checks to see whether the number of units is more than expected and whether there are any changes to the address. If there are more units than expected, Jeffersonville determines whether the additional units already had a chance of selection. If so, the regional office is contacted with instructions for updating the listing sheet. If there are any address changes, Jeffersonville provides updates to headquarters for potential future samples at the same basic address.

For the area and group quarters sample, the listing sheets remain in the regional offices. However, Jeffersonville receives a summary of the sampling that was done by each regional office ensuring that, given the sampling rate, the regional office sampled the appropriate number of units. If not, the regional office is contacted and corrections are made.

The majority of listing sheets are found without any inaccuracies during the regional office and Jeffersonville reviews across all frame types. When errors or omissions are detected, they usually occur in batches, signifying a misinterpretation of instructions by a regional office or a newly hired field representative. Therefore, because of these low error rates during the listing sheet review, the intent and value of this review are being revisited.

Check-in of sample. Depending on the sampling frame and mode of interview, the monthly process of check-in of sample can occur many times as the sample cases progress through the regional offices, the Jeffersonville office, and headquarters. Check-in of sample describes the processes by which the regional offices verify that the field representatives receive all the sample cases that they are supposed to and that all are completed and returned. Since the CPS is now conducted entirely by either computer assisted personal or telephone interview, a closer control of the tracking of a sample case exists than was ever possible. Control systems are in place to control and verify the sample count. The regional offices use these systems to control the sample during the period when it is active, and all discrepancies must be resolved before the office is able to certify that the workload in the database is correct. Since these systems aid in the balancing of the workload, time and effort can then be spent elsewhere.

Listing check. The purpose of the regional offices' listing check is to provide assurance that quality work is done on the listing operation and to provide feedback to the field representatives about their performance, with possible retraining when a field representative misunderstands or misapplies a listing concept. In the field, it checks a sample of each field representative's work to see whether he/she began the listing of units at the correct location, listed the units correctly, stayed in the appropriate geographic boundaries, and included all the units. Since it is not a part of the monthly production cycle and it can occur either before or after an interview, its intent is not to capture and correct data. Rather, it is an attempt, over time, to counteract any misunderstanding of listing and coverage procedures.

Since January 1993, the listing check has been a separate operation from reinterview; it is generally combined with some other personal visit field operation, such as observation. To avoid a bias in the designation of the units to be interviewed, the listing is performed without any knowledge of which units will be in the interview or reinterview. A listing check is performed only on area frame samples for field representatives who list or update. These field representatives are randomly checked once during the fiscal year, and depending on how close assignments are to each other, at least two assignments are checked to better evaluate the field representative's listing skills. The check of the unit frame sample was dropped when address samples were converted to samples from the unit frame in the 1990 Sample Redesign. This was done since little of the unit frame sample requires listing. Also, because permit and group quarters samples constitute such a small part of the universe, no formal listing check is done for sample in these frames.

SOURCES OF ERROR DUE TO NONRESPONSE

There are a variety of sources of nonresponse error in the CPS. Noninterviews could include housing units that are vacant, not yet built, demolished, or are not residential living units; however, these types of noninterviews are not considered to be nonresponses since they are out-of-scope and not eligible for interview. (See Chapter 7 for a description of the types of noninterviews.) Nonresponse error does occur when households that are eligible for interview are not interviewed for some reason: a respondent refuses to participate in the survey, is incapable of

completing the interview, or is not available or not contacted by the interviewer during the survey period, perhaps due to work schedules or vacation. These household noninterviews are called Type A noninterviews.

There are additional kinds of nonresponse in the CPS. Individual persons within the household may refuse to be interviewed, resulting in person nonresponse. Person nonresponse has not been much of a problem in the CPS because any responsible adult in the household is able to report for other persons in the household as a proxy reporter. Also, panel nonresponse exists when those persons who live in the same household during the entire time they are in the CPS sample may not agree to be interviewed each of the 8 months. Thus, panel nonresponse can be important if the CPS data are used longitudinally. Finally, some respondents who complete the CPS interview may be unable or unwilling to answer specific questions resulting in item nonresponse. Imputation procedures are implemented for item nonresponse. However, because there is no way of ensuring that the errors of item imputation will balance out, even on an expected basis, item nonresponse also introduces potential bias into the estimates.

For a sense of the magnitude of the error due to nonresponse, the household noninterview rate was 6.34 percent in January 1997. For item nonresponse, it is interesting to note allocation rates of 0.3 percent for the characteristic of Labor Force Status, 1.4 percent for Race, and 1.7 percent for Occupation, all using January 1997 weighted data. (See Chapter 16 for discussion on various quality indicators of nonresponse error.)

CONTROLLING ERROR DUE TO NONRESPONSE

Field Representative Guidelines[2]

Response/nonresponse rate guidelines have been developed for field representatives to help ensure the quality of the data collected. Maintaining high response rates is of primary importance, and response/nonresponse guidelines have been developed with this in mind. These guidelines, when used in conjunction with other sources of information, are intended to assist supervisors in identifying field representatives needing performance improvement. A field representative whose response rate, household noninterview rate (Type A), or minutes per case fall below the fully acceptable range based on one quarter's work is considered in need of additional training and development, and the CPS supervisor takes appropriate remedial action. National and regional response performance data are also provided to permit the regional office staff to judge whether their activities are in need of additional attention.

---

[2]See Appendix F for a detailed discussion, especially in terms of the performance evaluation system.

Summary Sheets

Another way to control nonresponse error is the production and review of summary sheets. Although produced by headquarters after the release of the monthly data products, they are used to detect changes in historical response patterns. Also, since they are distributed throughout headquarters and the regional offices, other indications of data quality and consistency can be focused upon. Contents of selected summary report tables are as follows: noninterview rates by regional office; monthly comparisons to prior year; noninterview-to-interview conversion rates; resolution status of computer assisted telephone interview cases; interview status by month-in-sample; daily transmittals; and coverage ratios.

Headquarters and Regional Offices Working as a Team

As detailed in a Methods and Performance Evaluation Memorandum (Reeder, 1997), the Census Bureau and the Bureau of Labor Statistics formed an interagency work group to examine CPS nonresponse in detail. One goal was to share possible reasons and solutions for the declining CPS response rates. A list of 31 questions was prepared for the regional offices to help aid in the understanding of CPS field operations, to solicit and share the regional offices' views on the causes of the increasing nonresponse rates, and to evaluate methods to decrease these rates. All of the answers provide insight on CPS operations that may affect nonresponse and followup procedures for household noninterviews, but a listing of a few will suffice:

- The majority of regional offices responded that there is written documentation of the followup process for CPS household noninterviews;

- The standard process was that a field representative must let the regional office know about a possible household noninterview as soon as possible;

- Most regions attempt to convert confirmed refusals under certain circumstances;

- All regions provide monthly feedback to their CPS field representatives on their household noninterview rates; and

- About half of the offices responded that they provide specific regional based training/activities for field representatives on converting or avoiding household noninterviews.

Response to one question deserves special attention: Most offices use letters in a consistent manner for followup of noninterviews. Most regional offices also include informational brochures tailored to the respondent with the letters.

SOURCES OF RESPONSE ERROR

The survey interviewer asks a question and collects a response from the respondent. Response error exists if the response is not the true answer. Reasons for response error include:

- The respondent misinterprets the question, tries to look better by inflating or deflating the response, does not know the true answer and guesses (for example, recall effects), or chooses a response randomly;

- The interviewer reads the question incorrectly, does not follow the appropriate skip pattern, misunderstands or misapplies the questionnaire, or records the wrong answer;

- The data collection modes (for example, personal visit and telephone) elicit different response rates, as well as the frequency of interview. Studies have shown that response rates generally decline as the length of time in survey increases; and

- The questionnaire does not elicit correct responses, due to a nonuser-friendly format, complicated skip patterns, and difficult coding procedures.

Thus, response error can arise from many sources. The survey instrument, the mode of data collection, the interviewer, and the respondent are the focus of this section[3], as well as their interactions as discussed in The Reinterview Program.

In terms of magnitude, measures of response error are obtainable through the reinterview program; specifically, Chapter 16 discusses the index of inconsistency. It is a ratio of the estimated simple response variance to the estimated total variance, arising from sampling and simple response variance. When identical responses are obtained from trial to trial, both the simple response variance and the index of inconsistency are zero. Theoretically, the index has a range of 0 to 100. For example, the index of inconsistency for the labor force characteristic of unemployed for February through July 1994 is considered moderate at 30.8.

CONTROLLING RESPONSE ERROR

The Survey Instrument[4]

The survey instrument involves the CPS questionnaire, the computer software that runs the questionnaire, and the mode by which the data are collected. The modes are personal visits or telephone calls made by field representatives and telephone calls made by interviewers at centralized telephone centers. Regardless of the mode, the questionnaire and the software are basically the same (See Chapter 7.)

Software. Changes in the early 1990s resulted in the use of computer assisted interviewing technology in the CPS. Designed in an automated environment, this technology allows very complex skip patterns and other procedures which combine data collection, data input, and a degree of in-interview consistency editing into a single operation.

This technology provides an automatic selection of questions for each interview. The screens display response options, if applicable, and information about what to do next. The interviewer does not have to worry about skip patterns, with the possibility of error. Appropriate proper names, pronouns, verbs, and reference dates are automatically filled into the text of the questions. If there is a refusal to answer a demographic item, that item is not asked again in later interviews; rather, it is longitudinally allocated. This exhibits the trade-off of nonsampling error existing for the item versus the possibility of a total noninterview. The instrument provides opportunities for the field representative to review and correct any incorrect/inconsistent information before the next series of questions is asked, especially in terms of the household roster. In later months, the instrument passes industry and occupation information forward to be verified and corrected. In addition to reducing response and interviewer burden, this also avoids erratic variations in industry and occupation codes among pairs of months for people who have not changed jobs, but describe their industry and occupation differently in the 2 months.

The questionnaire. Beginning in January 1994, the CPS used a questionnaire in which the wording was different from what was collected previously for almost every labor force concept. One of the objectives in redesigning the CPS questionnaire was to reduce the potential for response error in the questionnaire-respondent-interviewer interaction and to improve measurement of CPS concepts. The approaches used to lessen the potential for response error (i.e., enhanced accuracy) were: shorter and clearer question wording, splitting complex questions into two or more questions, building concept definitions into question wording, reducing reliance on volunteered information, explicit and implicit strategies for the respondent to provide numeric data, and the use of precoded response categories for open-ended questions. Although they were also used with the pencil and paper questionnaire, interviewer notes recorded at the end of the interview are critical to obtaining reliable and accurate responses.

Mode of data collection. As stated in Chapters 7 and 16, the first and fifth months' interviews are done in person whenever possible, while the remaining interviews may be conducted via telephone either by the field interviewer or

---

[3]Most discussion in this section is applicable whether the interview is conducted via computer assisted personal interview, telephone interview by the field representative or a centralized telephone facility, or computer assisted telephone interview.

[4]Many of the topics in this section are presented in more detail in Chapter 6.

an interviewer from a centralized telephone facility. Although each mode has its own set of performance guidelines that must be adhered to, similarities do exist. The controls detailed in *The Interviewer* of this section, and *Error Due to Nonresponse,* which are mainly directed at personal visits, are basically valid for the calls made from the centralized facility via the supervisor's listening in.

Continuous testing and improvements. Experience gained during the CPS redesign and research on questionnaire design has assisted in the development of methods for testing new or revised questions for the CPS. In addition to reviewing new questions to ensure that they will not jeopardize the collection of basic labor force information and to determine whether the questions are appropriate additions to a household survey about the labor force, the wording of new questions are tested to gauge whether respondents are correctly interpreting the questions. Chapter 6 provides an extensive listing of the various methods of testing and the Census Bureau developed a set of protocols for pretesting demographic surveys.

There is also interest in improving existing questions. The "don't know" and refusal rates for specific questions are monitored, inconsistencies caused by instrument-directed paths through the survey or instrument-assigned classifications are looked for during the estimation process, and interviewer notes recorded at the conclusion of the interview are reviewed. Also, focus groups with CPS interviewers and supervisors are periodically conducted.

Despite the benefits of adding new questions and improving existing ones, changes in the CPS are approached cautiously until the effects are measured and evaluated. When possible, methods to bridge differences caused by changes or techniques to avoid the disruption of historical series are included in the testing. In fact, for the 18 months prior to the implementation of the 1994 redesigned CPS, a parallel survey was conducted using the new methodology and, for the first 5 months after the redesigned survey was introduced, a parallel survey was conducted using the unrevised procedures. Results from the parallel survey were used to anticipate the effect the changes would have on the survey estimates, nonresponse rates, etc. (Kostanich and Cahoon, 1994; Polivka, 1994; and Thompson, 1994.) Comparable testing and parallel surveys were used for previous revisions to the questionnaire.

The Interviewer

Interviewer training, observation, monitoring, and evaluation are all methods used to control nonsampling error, arising from both the frame and data collection. For further discussion of field representative guidelines, see this chapter's Error Due to Nonresponse section and Appendix F. Group training and home study are continuing tasks in each regional office to control various nonsampling errors, and they are tailored to the types of duties and length of service of the interviewer. Observations, including those of the listing check, are extensions of classroom training and provide on-the-job training and on-the-job evaluation.

Field observation is one of the methods used by the supervisor to check and improve the performance of the field representative. It provides a uniform method for assessing the field representative's attitudes toward the job and use of the computer and evaluating the field representative's ability to apply CPS concepts and procedures during actual work situations. There are three types of observations: initial, general performance review, and special needs. Across all types, the observer stresses good interviewing techniques: asking questions as worded and in the order presented on the questionnaire, adhering to instructions on the instrument and in the manuals, knowing how to probe, recording answers in the correct manner and in adequate detail, developing and maintaining good rapport with the respondent conducive to an exchange of information, avoiding questions or probes that suggest a desired answer to the respondent, and determining the most appropriate time and place for the interview. The emphasis is on correcting habits which interfere with the collection of reliable statistics.

The Respondent: Self Versus Proxy

The CPS Interviewing Manual (see Chapter 4) states that any household member 15 years of age or older is technically eligible to act as a respondent for the household. The field representative attempts to collect the labor force data from each eligible individual; however, in the interests of timeliness and efficiency, any knowledgeable adult household member can provide the information. Also, the survey instrument is structured so that every effort is made to interview the same respondent every month. Just over one-half of the CPS labor force data are collected by self-response, and most of the remainder is collected by proxy from a household respondent. The use of a non-household member as a household respondent is only allowed in certain limited situations; for example, the household may consist of a single person whose physical or mental health does not permit a personal interview.

As a caveat, there has been a substantial amount of research into self versus proxy reporting, including research involving CPS respondents. Much of the research indicates that self-reporting is more reliable than proxy reporting, particularly when there are motivational reasons for self and proxy respondents to report differently. For example, parents may intentionally "paint a more favorable picture" of their children than fact supports. However, there are some circumstances in which proxy reporting is more accurate, such as in responses to certain sensitive questions.

Interviewer/Respondent Interaction

Rapport with the respondent is a means of improving data quality. This is especially true for personal visits, which

are required for months-in-sample one and five whenever possible. By showing a sincere understanding and interest in the respondent, a friendly atmosphere is created in which the respondent can talk honestly and openly. Interviewers are trained to ask questions exactly as worded and to ask every question. If the respondent misunderstands or misinterprets a question, the question is repeated as worded and the respondent is given another chance to answer; probing techniques are used if a relevant response is still not obtained. The respondent should be left with a friendly feeling towards the interviewer and the Census Bureau, clearing the way for future contacts.

The Reinterview Program - Quality Control and Response Error[5]

One of the objectives of the reinterview program is to evaluate individual field representative performance. Thus, it is a significant part of quality control. It checks a sample of the work of a field representative and identifies and measures aspects of the field procedures which may need improvement. It is also critical in the identification and prevention of data falsification.

The reinterview program also provides a measure of response error. Responses from first and second interviews at selected households are compared and differences are identified and analyzed. This helps to evaluate the accuracy of the original survey results; as a byproduct, instructions, training, and procedures are also evaluated.

SOURCES OF MISCELLANEOUS ERRORS

Data processing errors are one focus of this final section. Their sources can include data entry, industry and occupation coding, and methodologies for edits, imputations, and weighting. Also, the CPS Population Controls are not error-free; a number of approximations or assumptions are used in their derivations. Other sources are composite estimation and modeling errors which may arise from, for example, seasonally adjusted series for selected labor force data, and monthly model-based state labor force estimates.

CONTROLLING MISCELLANEOUS ERRORS

Parallel Survey

Among specific processes that are used to control nonsampling error in each of these sources, the parallel survey mentioned in *Response Error* crosses several sources. The 18 months of the parallel survey were used to test the survey instrument and the development of the processing

system. It ensured that the edits, imputations, and weighting program were working correctly. It often uncovered related problems, such as incorrect skip-patterns. (The Bureau of Labor Statistics was involved extensively in the review of the edits.)

Industry and Occupation Coding Verification

To be ultimately accepted into the CPS processing system, files containing records needing three-digit industry and occupation codes are electronically sent to the Jeffersonville office for the assignment of these codes (see Chapter 9). Once completed and transmitted back to headquarters, the remainder of the production processing, including edits, weighting, microdata file creation, and tabulations can begin.

Using on-line industry and occupation reference materials, the Jeffersonville coder enters three-digit numeric industry and occupation codes that represent the description for each case on the file. If the coder cannot determine the proper code, the case is assigned a referral code, which will later be coded by a referral coder. A substantial effort is directed at the supervision and control of the quality of this operation. The supervisor is able to turn the dependent verification setting *on* or *off* at any time during the coding operation. (In the *on* mode, a particular coder's work is to be verified by a second coder. Additionally, a 10 percent sample of each month's cases is selected to go through a quality assurance system to evaluate each coder's work. The selected cases are verified by another coder after the current monthly processing has been completed. Upon completion of the coding and possible dependent verification of a particular batch, all cases for which a coder assigned at least one referral code must be reviewed and coded by a referral coder.

Edits, Imputation, and Weighting

As detailed in Chapter 9, there are six edit modules: household, demographic, industry and occupation, labor force, earnings, and school enrollment. Each module establishes consistency between logically related items, assigns missing values using relational imputation, longitudinal editing, or cross-sectional imputation, and deletes inappropriate entries. Each module also sets a flag for each edit step that can potentially affect the unedited data.

Consistency is one of the checks used to control non-sampling error. Are the data logically correct, and are the data consistent within the month? For example, if a respondent says that he/she is a doctor, is he/she old enough to have achieved this occupation? Are the data consistent from the previous month and across the last 12 months? The imputation rates should normally stay about the same as the previous month's, taking seasonal patterns into account. Are the universes verified, and how consistent are the interview rates? In terms of the weighting, a check is

---

[5]Appendix G provides an overview of the design and methodology of the entire Reinterview Program.

made for zero or very large weights. If such outliers are detected, verification and possible correction follow. Another method to validate the weighting is to look at coverage ratios which should fall within certain historical bounds. A key working document used by headquarters for all of these checks is a three-sheet tabulation of monthly summary statistics that highlights the six edit modules and the weighting program for the past 12 months.

The CPS composite estimator is used for the derivation of most official CPS labor force estimates that are based upon information collected every month from the full sample (i.e., not collected in periodic supplements or from partial samples.) As detailed in Chapter 10, it modifies the aggregated estimates without adjusting the weights of individual sample records by combining the simple weighted estimate and the composite estimate for the preceding month plus an estimate of the change from the preceding to the current month.

Extensive verification at the time of program development was done to ensure that it was working correctly. Also, as part of the routine production processing at headquarters and as part of the routine check-in of the data by the Bureau of Labor Statistics, both compute identical tables in composited and uncomposited modes. These tables are then checked cell for cell to ensure that all uncomposited cells across the two agencies and that all composited cells across the two agencies are identical. If so, and they always have been, the data are deemed to have been computed correctly.

CPS Population Controls

Total and state-level CPS population controls are developed by the Census Bureau independently from the collection and processing of CPS data. These monthly independent projections of the population are used for the iterative, second-stage weighting of the CPS data. All of the estimates start with the last census, with administrative records and projection techniques providing updates. (See Appendix D for a detailed discussion of the methodologies.)

As a means of controlling nonsampling error throughout the processes, numerous internal consistency checks in the programming are performed. For example, input files containing age and sex detail are compared to independent files that give age and sex totals. Second, internal redundancy is intentionally built into the programs that process the files, as well as in files that contain overlapping/redundant data. Third, a clerical review (i.e., two-person review) of all handwritten input data is performed. A final and extremely important means of assuring that quality data are input into the CPS population controls is the continuous research into improvements in methods of making population estimates and projections.

Modeling Errors

It may be hypothesized that although modeling may indeed reduce some sampling and nonsampling errors, other nonsampling error due to the misspecification of the model may be introduced. Regardless, this final section focuses on a few of the methods of nonsampling error reduction, as applied to the seasonal adjustment programs, monthly model-based state labor force estimates, and composite estimation for selected labor force data. (See Chapter 10 for a discussion of these procedures.)

Changes that occur in a seasonally adjusted series reflect changes other than those arising from normal seasonal change; they are believed to provide information about the direction and magnitude of changes in the behavior of trend and business cycle effects. They may, however, also reflect the effects of sampling and nonsampling errors, which are not removed by the seasonal adjustment process. Research into the sources of these irregularities, specifically nonsampling error, can then lead to controlling their effects and even removal. The seasonal adjustment programs contain built-in checks as verification that the data are well-fit and that the modeling assumptions are reasonable. These diagnostic measures are a routine part of the output.

The processes for controlling nonsampling error during the production of monthly model-based state labor force estimates are very similar to those used for the seasonal adjustment programs. Built-in checks exist in the programs and, again, a wide range of diagnostics are produced that indicate the degree of deviation from the assumptions.

REFERENCES

Brooks, C.A. and Bailar, B.A. (1978), **An Error Profile: Employment as Measured by the Current Population Survey,** Statistical Policy Working Paper 3, Washington, D.C.: Subcommittee on Nonsampling Errors, Federal Committee on Statistical Methodology, U.S. Department of Commerce.

Kostanich, D.L. and Cahoon, L.S. (1994), **Effect of Design Differences Between the Parallel Survey and the New CPS,** CPS Bridge Team Technical Report 3, dated March 4.

McCarthy, P.J. (1978), **Some Sources of Error in Labor Force Estimates from the Current Population Survey,** Background Paper No. 15, Washington, D.C.: National Commission on Employment and Unemployment Statistics.

Polivka, A.E. (1994), **Comparisons of Labor Force Estimates from the Parallel Survey and the CPS during 1993: Major Labor Force Estimates,** CPS Overlap Analysis Team Technical Report 1, dated March 18.

Reeder, J.E. (1997), **Regional Response to Questions on CPS Type A Rates,** Bureau of the Census, CPS Office Memorandum No. 97-07, Methods and Performance Evaluation Memorandum No. 97-03, January 31, 1997.

Thompson, J. (1994), **Mode Effects Analysis of Major Labor Force Estimates,** CPS Overlap Analysis Team Technical Report 3, April 14,1994.

Waite, P.J. (1991), **Continuing the Project Development Plan,** Bureau of the Census internal memorandum, June 7, 1991.

Chapter 16.
Quality Indicators of Nonsampling Errors

(Updated coverage ratios, nonresponse rates, and other measures of quality can be found by clicking on "Quality Measures" at *www.bls.census.gov/cps*)

INTRODUCTION

Chapter 15 contains a description of the different sources of nonsampling error in the CPS and the procedures intended to limit those errors. In the present chapter, several important indicators of potential nonsampling error are described. Specifically, coverage ratios, response variance, nonresponse rates, mode of interview, time in sample biases, and proxy reporting rates are discussed. It is important to emphasize that unlike sampling error, these indicators show only the presence of potential nonsampling error, not the actual degree of nonsampling error present.

Nonetheless, these indicators of nonsampling error are regularly used to monitor and evaluate data quality. For example, surveys with high nonresponse rates are judged to be of low quality, but the actual nonsampling error of concern is not the nonresponse rate itself, but rather nonresponse bias, that is, how the respondents differ from the nonrespondents on the variables of interest. Although it is possible for a survey with a lower nonresponse rate to have greater nonresponse bias than a survey that has a higher nonresponse rate (if the nonrespondents differ to a greater degree from the respondents in the survey with the lower nonresponse rate), one would generally expect that greater nonresponse indicates a greater potential for bias. Unfortunately, while it is relatively easy to measure nonresponse rates, it is extremely difficult to measure or even estimate nonresponse bias. Thus, these indicators are simply a measurement of the potential presence of nonsampling errors. We are not able to quantify the effect the nonsampling error has on the estimates, and we do not know the combined effect of all sources of nonsampling error.

COVERAGE ERRORS

When conducting a sample survey, the primary goal is to give every person in the target universe a known probability of selection into the sample. When this occurs, the survey is said to have 100 percent coverage. This is rarely the case, however. Errors can enter the system during almost any phase of the survey process, from frame creation to interviewing. A bias in the survey estimates results when characteristics of persons erroneously included or excluded from the survey differ from those correctly included in the survey. Historically in the CPS, the net effect of coverage errors has been a loss in population (resulting from undercoverage).

Coverage Ratios

One way to estimate the coverage error present in a survey is to compute a coverage ratio. A coverage ratio is a ratio of the estimated number of persons in a specific demographic group from the survey over an independent population total for that group. The CPS coverage ratios are computed by dividing a CPS estimate using the weights after the first-stage ratio adjustment by the independent population controls used to perform the second-stage ratio adjustment. See Chapter 9 for more information on computation of weights. Population controls are not error free. A number of approximations or assumptions are required in deriving them. See Appendix D for details on how the controls are computed. Chapter 15 highlighted potential error sources in the population controls. Undercoverage exists when the coverage ratio is less than 1.0 and an overcoverage when the ratio is greater than 1.0. Table 16–1 shows the average monthly coverage ratios for January 1996 to April 1996.

It is estimated that the CPS has a 92.6 percent coverage rate for the first 4 months of 1996. This implies that the survey is missing approximately 7.4 persons per 100. This rate has been fairly consistent throughout the 1990s. It is difficult to compare this coverage rate to prior decades because of sample design changes and changes in the method that produces the population totals, but the total coverage rate was estimated at 96.3 percent during the mid-1970s (Hanson, 1978).

In terms of race, Whites have the highest coverage ratio (93.8 percent) while Blacks have the lowest (83.9 percent). Whites and Other[1] races are significantly different from Blacks, but not significantly different from each other. Black males age 20-29 have the lowest coverage ratio (66.2 percent) of any race/age group. Females across all races except Other race have higher coverage ratios than males. Hispanics[2] also have relatively low coverage rates. Historically, Hispanics and Blacks have lower coverage rates than Whites for each age group, particularly the 20-29 age group. This is by no fault of the interviewers or the CPS process. These lower coverage rates for minorities affect labor force estimates because persons that are missed by the CPS are almost surely on the average quite different from those that are included. The persons that are missed

---

[1]Other includes American Indian, Eskimo, Aleut, Asian, and Pacific Islander.
[2]Hispanics may be of any race.

Table 16–1. **CPS Coverage Ratios by Age, Sex, Race, and Ethnicity**

[Average for January 1996–April 1996]

Age, sex, race, and ethnicity	Average monthly coverage ratio	Age, sex, race, and ethnicity	Average monthly coverage ratio	Age, sex, race, and ethnicity	Average monthly coverage ratio	Age, sex, race, and ethnicity	Average monthly coverage ratio
Total[1]	0.9255						
White.	**0.9376**	**Black**	**0.8392**	**Other[2] race. . . . .**	**0.9255**	**Hispanic[3]**	**0.8331**
Male.	0.9163	Male	0.7727	Male	0.9119	Male	0.7841
16-19	0.8974	16-19	0.8328	16-19	0.8754	16-19	0.8693
20-29	0.8406	20-29	0.6616	20-29	0.8615	20-29	0.7485
30-39	0.9091	30-39	0.7104	30-39	0.8623	30-39	0.7790
40-59	0.9357	40-59	0.8383	40-59	0.9616	40-59	0.7904
60-69	0.9705	60-69	0.8735	60-69	1.0110	60-69	0.7899
70+.	0.9672	70+	0.8644	70+	1.0236	70+.	0.7982
Female.	0.9575	Female.	0.8931	Female.	0.9380	Female.	0.8820
16-19	0.9385	16-19	0.8676	16-19	0.9071	16-19	0.8658
20-29	0.9130	20-29	0.8170	20-29	0.8966	20-29	0.9015
30-39	0.9501	30-39	0.8650	30-39	0.9023	30-39	0.8920
40-59	0.9613	40-59	0.9374	40-59	0.9743	40-59	0.8766
60-69	0.9681	60-69	0.9541	60-69	1.0185	60-69	0.8310
70+.	1.0173	70+	0.9820	70+	1.0025	70+	0.8684

[1]The estimated standard errors on the Total, White, Black, Other Race, and Hispanic groups are 0.0058, 0.0064, 0.0180, 0.0322, and 0.0198, respectively.

[2]Other includes American Indian, Eskimo, Aleut, Asian, and Pacific Islander.

[3]Hispanics may be of any race.

Figure 16–1. **Average Yearly Noninterview and Refusal Rates for the CPS 1964-1996**

are accounted for in the CPS, but they are given labor force characteristics the same as for those that are included in the CPS. This produces bias in CPS estimates.

NONRESPONSE

As noted in Chapter 15, there are a variety of sources of nonresponse in the CPS, such as unit or household nonresponse, panel nonresponse, and item nonresponse. Unit nonresponse, referred to as Type A noninterviews, represents households that are eligible for interview but were not interviewed for some reason. Type A noninterviews occur because a respondent refuses to participate in the survey, is too ill or is incapable of completing the interview, is not available or not contacted by the interviewer, perhaps because of work schedules or vacation, during the survey period. Because the CPS is a panel survey, households who respond 1 month may not respond during a following month. Thus, there is also panel nonresponse in the CPS, which can become particularly important if CPS data are used longitudinally. Finally, some respondents who complete the CPS interview may be unable or unwilling to answer specific questions in the CPS resulting in some level of item nonresponse.

Type A Nonresponse

Type A noninterview rate. The Type A noninterview rate is calculated by dividing the total number of Type A households (refusals, temporarily absent, noncontacts, and other noninterviews) by the total number of eligible households (which includes Type As and interviewed households).

As can be seen in Figure 16–1, the noninterview rate for the CPS has remained relatively stable at around 4 to 5 percent for most of the past 35 years; however, there have been some changes during this time. It is readily apparent from Figure 16–1 that there was a major change in the CPS nonresponse rate in January 1994, which reflects the launching of the redesigned survey using computer assisted survey collection procedures. This rise is discussed below.

The end of 1995 and the beginning of 1996 also show a jump in Type A noninterview rates that were chiefly because of disruptions in data collection because of shutdowns of the Federal Government (see Butani, Kojetin, and Cahoon, 1996). The relative stability of the overall noninterview rate from 1960 to 1994 masks some underlying changes that have occurred. Specifically, the refusal portion of the noninterview rate has shown an increase over this same period with the bulk of the increase in refusals taking place from the early 1960s to the mid-1970s. In the late 1970s, there was a leveling off so that refusal rates were fairly constant until 1994. To compensate for this increase, there was a corresponding decrease in the rate of noncontacts and other noninterviews.

There also appears to be seasonal variation in both the overall noninterview rates and the refusal rates (see Figure 16–2). During the year, the noninterview and refusal rates have tended to increase after January until they reached a peak in March or April at the time of the income supplement. At this point, there was a steep drop in noninterview and refusal rates that extended below the starting point in January until they bottomed out in July or August. The rates then increased and approached the initial level. This pattern has been evident most years in the recent past and appears to be similar for 1994.

Effect of the transition to a redesigned survey with computer assisted data collection on noninterview rates. With the transition to the redesigned CPS questionnaire using computerized data collection in January 1994, there was a noticeable increase in the Type A nonresponse rates as can be seen in Figures 16–1 and 16–2. As part of this transition, there were several procedural changes in the collection of data for the CPS, and the adjustment to these new procedures may account for this increase. For example, the CAPI instrument now required the interviewers to go through the entire interview, while previously some interviewers may have conducted shortened interviews with reluctant respondents, obtaining answers to only a couple of critical questions. Another change in the data collection procedures was an increased reliance on using centralized telephone interviewing in CPS. Households not interviewed by the CATI centers by Wednesday night of interview week are recycled back to the field representatives on Thursday morning. These recycled CATI cases can present difficulties for the field representatives because there are only a few days left to make contact before the end of the interviewing period. As depicted in Figure 16-2, there has been greater variability in the monthly Type A nonresponse rates in CPS since the transition in January 1994. In addition, because of events such as the Federal Government shutdowns that affected nonresponse rates, it remains to be seen at what level the Type A nonresponse rates will stabilize for the new CPS. The annual overall Type A rate, the refusal rate, and noncontact rate (which includes temporarily absent households and other noncontacts) are shown in Table 16–2 for the period 1993-1996. Data beyond 1996 were not available at the time of this writing.

Table 16–2. **Components of Type A Nonresponse Rates, Annual Averages for 1993-1996**

[Percent distribution]

Nonresponse rate	1993	1994	1995	1996
Overall Type A.............	4.69	6.19	6.86	6.63
Noncontact................	1.77	2.30	2.41	2.28
Refusal	2.85	3.54	3.89	4.09
Other.....................	.13	.32	.34	.25

Panel nonresponse. Households are selected into the CPS sample for a total of 8 months in a 4-8-4 pattern as described in Chapter 3. Many families in these households

Figure 16–2. **Monthly Noninterview and Refusal Rates for the CPS 1984-1996**

Note: The break in the refusal rate indicates government furlough.

may not be in the CPS the entire 8 months because of moving (movers are not followed, but the new household members are interviewed). Those who live in the same household during the entire time they are in the CPS sample may not agree to be interviewed each month. Table 16–3 shows the percentage of households who were interviewed 0, 1, 2, … 8 times during the 8 months that they were eligible for interview during the period January 1994 to October 1995. These households represent seven rotation groups (see Chapter 3) that completed all of their rotations in the sample during this period. The vast majority of households, 82 percent, had completed interviews each month, and only 2 percent never participated. The remaining 16 percent participated to some degree (for further information see Harris-Kojetin and Tucker, 1997).

Effect of Type A Noninterviews on Labor Force Classification.

Although the CPS has monthly measures of Type A nonresponse, the total effect of nonresponse on labor force estimates produced from the CPS cannot be calculated from CPS data alone. It is the nature of nonresponse that we do not know what we would like to know from the nonrespondents, and therefore, the actual degree of bias because of nonresponse is unknown. Nonetheless, because the CPS is a panel survey, information is often available at some point in time from households that were nonrespondents at another point. Some assessment can be made of

the effect of nonresponse on labor force classification by using data from adjacent months and examining the month-to-month flows of persons from labor force categories to nonresponse as well as from nonresponse to labor force categories. Comparisons can then be made for labor force status between households that responded both months with households that responded 1 month, but failed to respond in the other month. However, the labor force status of persons in households that were nonrespondents for both months is unknown.

Table 16–3. **Percentage of Households[1] by Number of Completed Interviews During the 8 Months in the Sample**

[January 1994–October 1995]

Number of completed interviews	Percent
0	2.0
1	.5
2	.5
3	.6
4	2.0
5	1.2
6	2.5
7	8.9
8	82.0

[1]Includes only households in the sample all 8 months with only interviewed and Type A nonresponse interview status for all 8 months, i.e., households that were out of scope (e.g., vacant) for any month in sample were not included in these tabulations. Movers were not included in this tabulation.

Table 16–4. **Labor Force Status by Interview/Noninterview Status in Previous and Current Month**

[Average January–June 1997[1] percent distribution]

1st month labor force status	Interview in 2nd month	Nonresponse in 2nd month	Difference
Civilian labor force	65.98	68.51	2.53**
Employed..	62.45	63.80	1.35**
Unemployment rate	5.35	6.87	1.52**
2nd month labor force status	**Interview in 1st month**	**Nonresponse in 1st month**	**Difference**
Civilian labor force	65.79	67.41	1.62**
Employed..	62.39	63.74	1.35**
Unemployment rate	5.18	5.48	.30*

** p < .01   * <. 05

[1]From Tucker and Harris-Kojetin (1997).

Monthly labor force data were used for each consecutive pair of months for January through June 1997, separately for households that responded for each consecutive pair of months, and for households that responded only 1 month and were nonrespondents the other month (see Tucker and Harris-Kojetin, 1997). The top half of Table 16–4 shows the labor force classification in the first month for persons in households that were respondents the second month compared to persons who were in households that were noninterviews the second month. Persons from households that became nonrespondents had significantly higher rates of participation in the labor force, employment, and unemployment than persons from households that were respondents both months. The bottom half of Table 16–4 shows the labor force classification for the second month for persons in households that were respondents in the previous month compared to persons who were in households that were noninterviews the previous month. The pattern of differences is similar, but the magnitude of the differences is less. Because the overall Type A noninterview rate is quite small, the effect of the nonresponding households on the overall unemployment rate is also relatively small. However, it is important to remember that the labor force characteristics of the persons in households who never responded are not measured or included here. In addition, other nonsampling errors are also likely present, such as those due to repeated interviewing or month-in-sample effects (described later in this chapter).

Item Nonresponse

Another component of nonresponse is item nonresponse. Respondents may refuse or may be unable to answer certain items, but still respond to most of the CPS questions with only a few items missing. To examine the prevalence of item nonresponse in the CPS, counts were made of the refusals and "don't know" responses from 10 demographic items and 94 labor force items (due to skip patterns only some of these questions were asked of each respondent) for all interviewed cases from January to December 1994. The average levels of item-missing data

Table 16–5. **Percentage of CPS Items With Missing Data**

Item series	Percent
Demographic.....................................	1.54
Labor force	1.46
Industry and occupation	3.76
Earnings..	12.44

from most of the CPS labor force and demographic items are shown in Table 16–5. It is likely that the bias due to imputation for the demographic and labor force items is quite small, but there may be a concern for earnings data.

RESPONSE VARIANCE

Estimates of the Simple Response Variance Component

To obtain an unbiased estimate of the simple response variance, it is necessary to have at least two independent measurements of the characteristic for each person in a subsample of the entire sample using the identical measurement procedure on each trial. It is also necessary that responses to a second interview not be affected by the response obtained in the first interview.

Two difficulties occur in every attempt to measure the simple response variance by a reinterview. The first is lack of independence between the original interview and the reinterview; a person visited twice within a short period and asked the same questions may tend to remember his/her original responses and repeat them. A second difficulty is that the data collection methods used in the original interview and in the reinterview are seldom the same. Both of these difficulties exist in the CPS reinterview program data used to measure simple response variance. For example, the interviews in the response variance reinterview sample are conducted by telephone by senior interviewers or supervisory personnel. Also, some characteristics of the reinterview survey process itself introduce

Table 16–6. **Comparison of Responses to the Original Interview and the Reinterview**

Reinterview	Original interview		Total
	Unemployed	Not unemployed	
Unemployed	a	b	a + b
Not unemployed	c	d	c + d
Total	a + c	b + d	n = a + b + c + d

unknown biases in the estimates (e.g., a high noninterview rate.)

It is useful to consider a schematic representation of the results of the original CPS interview and the reinterview in estimating, say, the number of persons reported as unemployed (see Table 16-6). In this schematic representation, the number of identical persons having a difference between the original interview and the reinterview for the characteristic unemployed is given by (b + c). If these responses are two independent measurements for identical persons using identical measurement procedures, then an estimate of the simple response variance component is given by (b + c) /2n and

$$E\left(\frac{b + c}{2n}\right) = \sigma_R^2$$

and as in formula 14.3 (Hansen, Hurwitz, and Pritzker, 1964)[3]. We know that (b + c)/2n, using CPS reinterview data, underestimates σ_R^2, chiefly because conditioning sometimes takes place between the two responses for a specific person resulting in correlated responses, so that fewer differences (i.e., entries in the b and c cells) occur than would be expected if the responses were in fact independent.

Besides its effect on the total variance, the simple response variance is potentially useful to evaluate the precision of the survey measurement methods. It can be used to evaluate the underlying difficulty in assigning individuals to some category of a distribution on the basis of responses to the survey instrument. As such, it is an aid in determining whether a concept is sufficiently measurable by a household survey technique and whether the resulting survey data fulfill their intended purpose. For characteristics having ordered categories (e.g., number of hours worked last week), the simple response variance is helpful in determining whether the detail of the classification is too fine. To provide a basis for this evaluation, the estimated simple response variance is expressed, not in absolute terms, but as a proportion of the estimated total population variance. This ratio is called the index of inconsistency and has a theoretical range of 0.0 to 100.0 when expressed as a percentage (Hansen, Hurwitz, and Pritzker, 1964). The

denominator of this ratio, the population variance, is the sum of the simple response variance, σ_R^2, and the population sampling variance, σ_S^2. When identical responses are obtained from trial to trial, the simple response variance is zero and the index of inconsistency has a value of zero. As the variability in the classification of an individual over repeated trials increases (and the measurement procedures become less reliable), the value of the index increases until, at the limit, the responses are so variable that the simple response variance equals the total population variance. At the limit, if a single response from N individuals is required, equivalent information is obtained if one individual is randomly selected and interviewed N times, independently.

Two important inferences can be made from the index of inconsistency for a characteristic. One is to compare it to the value of the index that could be obtained for the characteristic by best (or preferred) set of measurement procedures that could be devised. The second is to consider whether the precision of the measurement procedures indicated by the level of the index is still adequate to serve the purposes for which the survey is intended. In the CPS, the index is more commonly used for the latter purpose. As a result, the index is used primarily to monitor the measurement procedures over time. Substantial changes in the indices that persist for several months result in review of field procedures to determine and remedy the cause.

Table 16–7 provides estimates of the index of consistency shown as percentages for selected labor force characteristics for the period of February through July 1994.

Table 16–7. **Index of Inconsistency for Selected Labor Force Characteristics February–July 1994**

Labor force characteristic	Estimate of index of inconsistency	90 percent confidence limits
Working, full time	9.4	8.4 to 10.6
Working, part time	22.6	20.3 to 25.0
With a job, not at work.................	30.1	24.5 to 37.0
Unemployed..........................	30.8	25.9 to 36.5
Not in labor force	8.1	7.1 to 9.2

MODE OF INTERVIEW

Incidence of Telephone Interviewing

As described in Chapter 6, the first and fifth months' interviews are typically done in person while the remaining interviews may be done over the telephone either by the field interviewer or an interviewer from a centralized telephone facility. Although the first and fifth interviews are supposed to be done in person, the entire interview may

---

[3]The expression (b+c)/n is referred to as the gross difference rate; thus, the simple response variance is estimated as one-half the gross difference rate.

not be completed in person, and the field interviewer may call the respondent back to obtain missing information. The CPS CAPI instrument records whether the last contact with the household was by telephone or personal visit. The percentage of CAPI cases from each month-in-sample that were completed by telephone are shown in Table 16–8. Overall, about 66 percent of the cases are done by telephone, with almost 85 percent of the cases in month-in-samples 2-4 and 6-8 done by telephone. Furthermore, a substantial percentage of cases in months 1 and 5 are obtained by telephone, despite standing instructions to the field interviewers to conduct personal visit interviews.

Table 16–8. **Percentage of Households With Completed Interviews With Data Collected by Telephone (CAPI Cases Only)**

[Percent]

Month-in-sample	Last contact with household was telephone (average, Jan.–Dec. 1995)	Majority of data collected by telephone (June 1996)
1......................................	9.6	15
2......................................	80.2	79
3......................................	83.1	83
4......................................	83.6	83
5......................................	23.5	34
6......................................	83.1	82
7......................................	84.1	84
8......................................	84.5	85

Because the indicator variable in the CPS instrument reflects only the last contact with the household, it may not be the best indicator of how most of the data were gathered from a household. For example, an interviewer may obtain information for several members of the household during the first month's personal visit, but may make a telephone call back to obtain the labor force data for the last household member resulting in the interview being recorded as a telephone interview. In June 1996, an additional item was added to the CPS instrument (for that month only) that asked interviewers whether the majority of the data for each completed case was obtained by telephone or personal visit. The results using this indicator are presented in the second column of Table 16–8. It was expected that in MIS 1 and 5 interviewers would have reported that the majority of the data was collected though personal visit more often than was revealed by the last contact with the household. This would seem likely because, as noted above, the last contact with a household may be a telephone call back to obtain missing information not collected at the initial personal visit. However, a higher percentage of cases were reported with the majority of the data being collected by telephone than the percentage of cases with the last contact with the household being by telephone for MIS 1 and 5. The explanation for this pattern of results is not clear at the present time.

Effects of Centralized Telephone Interviewing

With the implementation of the revised CPS in January 1994, there was an increased reliance on the use of the Census Bureau's centralized telephone centers for conducting CPS interviews. As noted in Chapter 4, only cases in CATI eligible PSUs can be sent to the CATI centers for interviewing. Furthermore, all cases within eligible PSUs were randomly assigned to the CATI panel or the control (CAPI) panels, so that meaningful comparisons could be made on the effects of centralized interviewing on estimates from CPS. All cases were interviewed by CAPI in MIS 1 and 5. In months 2-4 and 6-8, most, but not all of the cases in the CATI panel were interviewed in the centralized facilities, because some cases were still interviewed by field staff in CAPI for a variety of reasons. Nonetheless, to preserve the integrity of the research design, the results shown in Table 16–9 reflect the panel that each case was assigned to, which was also the actual mode of interview for the vast majority of cases

Table 16–9 shows the results of the comparisons of the CATI test and control panels. There were significant differences found between the CATI test and control panels in MIS 2-4 and 6-8 for all groups on the unemployment rate. There were no significant differences in the unemployment rate in MIS 1 and 5 (and none would be expected given random assignment); however, the observed differences were not necessarily equal to zero. These initial differences may have had some slight effects on the size of the differences observed in MIS 2-4 and 6-8 for some of the groups. There were no significant differences for the civilian labor force participation rate or the employment to population ratio. Although there are differences in unemployment rates between the centralized CATI and CAPI interviews as noted above[4], the interpretation of the findings is not as clear. One major difference between CATI and CAPI interviews is that CAPI interviews are likely to be done by the same interviewer all 8 MIS, while CATI interviews involve a definite change in interviewers from MIS 1 to 2 and from MIS 5 to 6. In fact, cases in the CATI facilities often have a different interviewer each month. However, it is not clear whether the results from the centralized CATI or the CAPI interviews are more accurate.

TIME IN SAMPLE

The rotation pattern of the CPS sample was described in detail in Chapter 3, and the use of composite estimation in CPS was discussed briefly in Chapter 9. The effects of interviewing the same respondents for CPS several times has been discussed for a long time (e.g., Bailar, 1975; Brooks and Bailar, 1978; Hansen, Hurwitz, Nisselson, and Steinberg, 1955; McCarthy, 1978; Williams and Mallows,

---

[4]These findings were also similar to those from a test conducted using the parallel survey (see Thompson, 1994).

Table 16–9. **Effect of Centralized Telephone Interviewing on Selected Labor Force Characteristics**

[Average January 1996 - December 1996]

Employment status and sex	MIS 1 and 5			MIS 2-4 and 6-8				
	CATI test	Control (CAPI)	Difference	CATI test	Control (CAPI)	Difference	Standard error	P-value
TOTAL POPULATION, 16 YEARS OLD AND OVER								
Civilian labor force	66.89	67.20	−0.31	66.47	66.44	0.03	1.09	.98
Employment to population ratio	63.20	63.60	−0.40	62.79	63.20	−0.41	1.06	.70
Unemployment rate	5.52	5.36	0.16	5.53	4.87	0.66	0.21	.00
MALES, 16 YEARS OLD AND OVER								
Civilian labor force	75.33	76.35	−1.02	74.96	75.31	−0.35	1.71	.84
Employment to population ratio	71.19	72.23	−1.03	70.85	71.63	−0.79	1.66	.64
Unemployment rate	5.49	5.40	0.09	5.49	4.88	0.61	0.28	.03
FEMALES, 16 YEARS OLD AND OVER								
Civilian labor force	59.50	59.19	0.31	59.04	58.68	0.36	1.18	.76
Employment to population ratio	56.20	56.05	0.14	55.75	55.83	−0.08	1.12	.94
Unemployment rate	5.55	5.30	0.25	5.57	4.85	0.72	0.21	.00
WHITE, 16 YEARS OLD AND OVER								
Civilian labor force	67.50	67.59	−0.08	67.16	66.89	0.28	1.35	.84
Employment to population ratio	64.22	64.37	−0.15	63.94	63.97	−0.02	1.33	.99
Unemployment rate	4.86	4.76	0.11	4.79	4.37	0.43	0.16	.01
WHITE MALES, 16 YEARS OLD AND OVER								
Civilian labor force	76.44	77.38	−0.93	76.14	76.31	−0.17	1.98	.93
Employment to population ratio	72.67	73.74	−1.07	72.52	72.98	−0.46	1.94	.81
Unemployment rate	4.94	4.70	0.24	4.76	4.37	0.39	0.23	.10
WHITE FEMALES, 16 YEARS OLD AND OVER								
Civilian labor force	59.42	58.80	0.62	59.07	58.43	0.64	1.35	.63
Employment to population ratio	56.59	55.97	0.62	56.22	55.88	0.34	1.29	.79
Unemployment rate	4.78	4.82	−0.05	4.83	4.36	0.47	0.20	.02
BLACKS, 16 YEARS OLD AND OVER								
Civilian labor force	63.30	65.44	−2.14	62.12	64.22	−2.10	4.22	.62
Employment to population ratio	56.94	58.85	−1.91	55.55	58.44	−2.89	3.99	.47
Unemployment rate	10.04	10.06	−.02	10.59	9.01	1.58	0.79	.04
BLACK MALES, 16 YEARS OLD AND OVER								
Civilian labor force	67.23	70.24	−3.01	66.45	69.21	−2.75	5.47	.61
Employment to population ratio	60.33	61.76	−1.43	58.73	62.07	−3.34	4.97	.50
Unemployment rate	10.26	12.08	−1.82	11.62	10.31	1.30	0.72	.07
BLACK FEMALES, 16 YEARS OLD AND OVER								
Civilian labor force	60.55	62.06	−1.51	59.03	60.75	−1.72	4.75	.72
Employment to population ratio	54.57	56.81	−2.24	53.27	55.90	−2.63	4.51	.56
Unemployment rate	9.87	8.46	1.41	9.75	7.97	1.79	0.83	.03

1970). It is possible to measure the effect of the time spent in the sample on labor force estimates from the CPS by creating a month-in-sample index which shows the relationships of all the month-in-sample groups. This index is the ratio of the estimate based on the sample units in a particular month-in-sample group to the average estimate from all eight month-in-sample groups combined, multiplied by 100. If an equal percentage of people with the characteristic are present in each month-in-sample group, then the index for each group would be 100. Table 16–10

shows indices for each group by the number of months they have been in the sample. The indices are based on CPS labor force data from September to December 1995. For the percentage of the total population that is unemployed, the index of 108.62 for the first month households indicates that the estimate from households in sample the first month is about 1.0862 times as great as the average overall eight month-in-sample groups; the index of 99.57 for unemployed for the second month households

indicates that it is about the same as the average overall month-in-sample groups.

It should not be assumed that estimates from one of the month-in-sample groups should be taken as the standard in the sense that it would provide unbiased estimates, while the estimates for the other seven would be biased. It is far more likely that the expected value of each group is biased—but to varying degrees. Total CPS estimates, which are the combined data from all month-in-sample groups (see Chapter 9), are subject to biases that are functions of the biases of the individual groups. Since the expected values vary appreciably among some of the groups, it follows that the CPS survey conditions must be different in one or more significant ways when applied separately to these subgroups.

One way in which the survey conditions differ among rotation groups is reflected in the noninterview rates. The interviewers, being unfamiliar with households in sample for the first time, are likely to be less successful in calling when a responsible household member is available. Thus, noninterview rates generally start above average with first month families, decrease with more time in sample, go up a little for households in sample the fifth month (after the 8 months the household was not in sample), and then decrease again in the final months. (This pattern may also reflect the prevalence of personal visit interviews conducted for each of the months-in-sample as noted above.) An index of the noninterview rate constructed in a similar manner to the month-in-sample group index can be seen in Table 16–11. This noninterview rate index clearly shows that the greatest proportion of noninterviews occurs for cases that are in the sample for the first month and the second largest proportion of noninterviews are due to cases returning to the sample in the fifth month. These

Table 16–10. **Month-In-Sample Bias Indexes (and Standard Errors) in the CPS for Selected Labor Force Characteristics**

[Average September - December 1995]

Employment status and sex	Month-in-sample							
	1	2	3	4	5	6	7	8
TOTAL POPULATION, 16 YEARS OLD AND OVER								
Civilian labor force	101.34	100.26	99.93	100.32	99.5	99.5	99.68	99.48
Standard error	(.27)	(.26)	(.25)	(.25)	(.28)	(.24)	(.29)	(.30)
Unemployment level	108.62	99.57	98.27	103.33	99.19	99.39	98.74	92.89
Standard error	(2.46)	(2.59)	(2.28)	(2.59)	(2.65)	(2.42)	(2.46)	(2.66)
MALES, 16 YEARS OLD AND OVER								
Civilian labor force	100.98	100.26	99.87	100.34	99.98	99.55	99.62	99.40
Standard error	(.31)	(.31)	(.31)	(.26)	(.34)	(.29)	(.32)	(.34)
Unemployment level	107.65	98.13	97.19	102.94	103.60	99.29	95.60	95.59
Standard error	(3.45)	(3.53)	(3.03)	(3.41)	(3.64)	(3.22)	(3.33)	(3.73)
FEMALES, 16 YEARS OLD AND OVER								
Civilian labor force	101.75	100.26	100.00	100.29	98.94	99.44	99.76	99.56
Standard error	(.42)	(.38)	(.38)	(.41)	(.39)	(.37)	(.45)	(.47)
Unemployment level	110.09	101.46	99.62	103.53	93.80	99.79	102.25	89.47
Standard error	(3.65)	(3.44)	(3.17)	(3.42)	(3.43)	(3.61)	(3.42)	(3.16)
BLACK, 16 YEARS OLD AND OVER								
Civilian labor force	102.66	102.82	100.04	98.95	99.19	99.03	99.11	98.21
Standard error	(1.05)	(.97)	(.99)	(.90)	(1.07)	(1.06)	(.90)	(1.04)
Unemployment level	116.50	104.13	103.44	106.24	94.56	85.48	92.83	96.82
Standard error	(5.84)	(5.41)	(5.29)	(6.17)	(5.46)	(5.14)	(4.94)	(6.00)
BLACK MALES, 16 YEARS OLD AND OVER								
Civilian labor force	101.17	101.40	100.00	100.88	100.31	98.73	99.99	97.52
Standard error	(1.45)	(1.34)	(1.16)	(1.17)	(1.29)	(1.28)	(1.29)	(1.52)
Unemployment level	114.46	100.86	100.47	109.39	97.38	80.47	93.32	103.65
Standard error	(9.48)	(8.33)	(7.49)	(8.97)	(8.74)	(7.19)	(6.97)	(8.95)
BLACK FEMALES, 16 YEARS OLD AND OVER								
Civilian labor force	104.00	104.13	100.06	97.17	98.16	99.32	98.31	98.86
Standard error	(1.40)	(1.31)	(1.40)	(1.25)	(1.45)	(1.38)	(1.27)	(1.30)
Unemployment level	117.30	107.78	106.40	104.24	91.10	90.86	92.62	89.71
Standard error	(8.72)	(7.00)	(7.30)	(7.92)	(6.76)	(6.86)	(7.42)	(7.49)

Table 16–11. **Month-In-Sample Indexes in the CPS for Type A Noninterview Rates January–December 1994**

Month-in-sample	Noninterview rate index
1	134.0
2	93.1
3	89.3
4	90.0
5	117.9
6	95.6
7	92.4
8	88.2

Table 16–12. **Percentage of CPS Labor Force Reports Provided by Proxy Reporters**

	Percent reporting
All interviews	
Proxy reports	49.68
Both self and proxy	0.28
MIS 1 and 5	
Proxy reports	50.13
Both self and proxy	0.11
MIS 2-4 and 6-8	
Proxy reports	49.53
Both self and proxy	0.35

noninterview changes are unlikely to be distributed proportionately among the various labor force categories (see Table 16–4). Consequently, it is reasonable to assume that the representation of labor force categories will differ somewhat among the month-in-sample groups and that these differences can affect the expected values of these estimates, which would imply that at least some of the month-in-sample bias can be attributed to actual differences in response probabilities among the month-in-sample groups (Williams and Mallows, 1970). Although the individual probabilities are not known, they can be estimated by nonresponse rates. However, other factors are also likely to affect the month-in-sample patterns.

PROXY REPORTING

Like many household surveys, the CPS seeks information about all persons in the household whether they are available for interview or not. CPS field representatives accept reports from responsible adults in the household (see Chapter 6 for a discussion of respondent rules) to provide information about all household members. Respondents who provide labor force information about other household members are called proxy reporters. Because some household members may not know or be able to provide accurate information about the labor force status and activities of some household members, nonsampling error may occur because of the use of proxy reporters.

The level of proxy reporting in the CPS had generally been around 50 percent in the past and continues to be so in the revised CPS. As can be seen in Table 16–12, the

month-in-sample has very little effect on the level of proxy reporting. Thus, whether the interview is more likely to be a personal visit (for MIS 1 and 5) or a telephone interview (MIS 2-4 and 6-8) has very little effect.

Although one can make overall comparisons of the data given by self-reporters and proxy reporters, there is an inherent bias in the comparisons because proxy reporters were the people more likely to be found at home when the field representative called or visited. For this reason, household members with self- and proxy reporters tend to differ systematically on important labor force and demographic characteristics. In order to compare the data given by self- and proxy reporters systematic studies must be conducted that control assignment of proxy reporting status. Such studies have not been carried out.

SUMMARY

This chapter contains a description of several quality indicators in the CPS, namely, coverage, noninterview rates, telephone interview rates, and proxy reporting rates. These rates can be used to monitor the processes of conducting the survey, and they indicate the potential for some nonsampling error to enter into the process. This chapter also includes information on the potential effects of nonresponse, centralized telephone interviewing, and the month-in-sample groups on CPS estimates. This research comes close to identifying the presence and effects of nonsampling error in the CPS, but the results are far from conclusive.

The full extent of nonsampling error in the CPS is unknown. Because of the number and scope of changes that have occurred in the questionnaire as well as in data collection, it is also currently unclear how much previous research on nonsampling error in the CPS is applicable to the redesigned instrument with computerized data collection. Furthermore, new sources of nonsampling error may have been introduced by the change in data collection methodology and with changes in questions. Although some potential sources of nonsampling error can be investigated with the currently available CPS data and reinterview data, the best source of information on nonsampling error is often information provided by an external or outside source or through a special study or experiment. At the time of this writing, there is currently about 3 years' data available for internal analyses and comparison.

REFERENCES

Bailar, B.A. (1975), " The Effects of Rotation Group Bias on Estimates from Panel Surveys," **Journal of the American Statistical Association, 70**, 23-30.

Brooks, C.A. and Bailar, B.A. (1978). **Statistical Policy Working Paper 3 An Error Profile: Employment as Measured by the Current Population Survey**, Subcommittee on Nonsampling Errors, Federal Committee on Statistical Methodology, U.S. Department of Commerce, Washington, DC.

Butani, S., Kojetin, B.A., and Cahoon, L. (1996), "Federal Government Shutdown: Options for Current Population Survey (CPS) Data Collection," Paper presented at the Joint Statistical Meetings of the American Statistical Association, Chicago, Illinois.

Hansen, M.H., Hurwitz, W.N., Nisselson, H., and Steinberg, J. (1955), "The Redesign of the Current Population Survey," **Journal of the American Statistical Association, 50**, 701-719.

Hansen, M.H., Hurwitz, W.N., and Pritzker, L. (1964), "Response Variance and Its Estimation," **Journal of the American Statistical Association, 59**, 1016-1041.

Hanson, R. H. (1978), **The Current Population Survey: Design and Methodology**, Technical Paper 40, Washington, DC: Government Printing Office.

Harris-Kojetin, B.A. and Tucker, C. (1997), "Longitudinal Nonresponse in the Current Population Survey," Paper presented at the 8th International Workshop on Household Survey Nonresponse, Mannheim, Germany.

McCarthy, P.J. (1978), **Some Sources of Error in Labor Force Estimates From the Current Population Survey**, Background Paper No. 15, National Commission on Employment and Unemployment Statistics, Washington, DC.

Thompson, J. (1994). "Mode Effects Analysis of Labor Force Estimates," **CPS Overlap Analysis Team Technical Report 3**, Bureau of Labor Statistics and U.S. Census Bureau, Washington, DC.

Tucker, C. and Harris-Kojetin, B.A. (1997), "The Impact of Nonresponse on the Unemployment Rate in the Current Population Survey," Paper presented at the 8th International Workshop on Household Survey Nonresponse, Mannheim, Germany.

Williams, W.H. and Mallows, C.L., (1970), "Systematic Biases in Panel Surveys Due to Differential Nonresponse," **Journal of the American Statistical Association, 65**, 1338-1349.

Appendix A.
Maximizing Primary Sampling Unit (PSU) Overlap

INTRODUCTION

The sampling of primary sampling units (PSUs) for the redesigned CPS was accomplished using linear programming techniques that maximized the overlap with the PSUs from the former 1980s design while maintaining the correct overall probabilities of selection for a probability proportional to size (pps) sampling scheme. This strategy enables the fullest use of field staff who are already hired and trained and results in substantial cost savings. Procedures are already in place in the previously existing PSUs for identifying and sampling new construction housing units. In addition to the cost savings, a large PSU overlap prevents overly large breaks in data series that can occur when many PSUs are changed with the introduction of a new design.

There were 379 nonself-representing (NSR) PSUs in the 1980 CPS design. (There were some small rotating PSUs, but only those still in sample at the end of the 1980 design are counted.) The maximizing procedure retained 249 (66 percent) of these NSR PSUs for the 1990 design. Had selection for the 1990 design been independent of the 1980 design, only 149 (39 percent) would have been retained.

Methods of solving the problem of maximizing PSU overlap, while maintaining the correct unconditional design probabilities, have evolved over several decades. A limited solution was presented by Keyfitz (1951) for one PSU per stratum pps designs similar to the CPS. The definitions of individual PSUs were held constant, and the strata were assumed identical. Only the PSU probabilities of selection were allowed to be different for the former and new designs, thus responding to relative PSU population size changes over a decade. Raj (1968) demonstrated that the problem solved by Keyfitz could be reformulated as a transportation problem, which is commonly solved using linear programming techniques. Causey, Cox, and Ernst (1985) generalized the transportation problem approach to allow PSU redefinitions, changes in stratum composition, and more than one sample PSU per stratum. The generalization assumes the former design sampling to be independent within strata, which does not hold for the CPS since the 1980s design itself had PSU overlap maximized from the earlier 1970s design. This and other practical considerations are addressed by various authors, including Ernst (1986, 1990).

To illustrate the concepts, consider the following example adapted from Causey et al. (1985). The PSUs and strata do not change across designs, and a particular stratum has three PSUs with probabilities $p_{1.}=.36$, $p_{2.}=.24$ and $p_{3.}=.4$ for the former design. Over the decade, PSUs 1 and 2 have grown in size relative to PSU 3. Were sampling done anew without regard for the former design, then $p_{.1}=.5$, $p_{.2}=.3$, and $p_{.3}=.2$ would be the probabilities for the new design. To obtain the maximum overlap of PSUs between the former and new designs, without restraint, we would arbitrarily choose for the new design any PSU that had been sampled for the former design. The implication of the arbitrary strategy for the example would be that PSU 3, which has been losing in relative size and is now the least populous PSU in the stratum, would retain the highest probability of selection (.4 from the former design). In general, the arbitrary strategy introduces bias by favoring declining PSUs over other PSUs, thus causing growing PSUs to be underrepresented.

Keyfitz's method of maximizing PSU overlap while maintaining the new design probabilities is intuitive. If PSU 1 was actually sampled for the former design, include it for the new design since the design probability has increased over the decade from $p_{1.}=.36$ to $p_{.1}=.5$. (We say the conditional probability of choosing PSU 1 for the new design, given that PSU 1 was sampled for the former design, is 1.0 or certainty.) Similarly, the design probability of PSU 2 has increased from $p_{2.}=.24$ to $p_{.2}=.3$, so if it was actually sampled for the former design then include it for the new design. (The conditional probability of choosing PSU 2 for the new design, given that PSU 2 was sampled for the former design, is 1.0 or certainty.) On the other hand, the design probability of selection for PSU 3 has decreased from $p_{3.}=.4$ to $p_{.3}=.2$. If PSU 3 was the one selected for the former design, the probability can be cut down to size by giving it a 1/2 chance of remaining for the new design and a 1/2 chance of being replaced. (The conditional probability of choosing PSU 3 for the new design, given that PSU 3 was sampled for the former design, is .5 or 1/2.) The overall probability of PSU 3 being in the sample for the new design is .4 x 1/2 = .2, which is exactly the probability required for the new design.

The following tables of conditional and joint probabilities have been partially completed. (A joint probability p_{ij} is the probability that the ith PSU was sampled for the former design and the jth PSU for the new design.) The question marks (?) indicate that if PSU 3 was sampled for the former design, but not conditionally reselected for the new design, then one of the other PSUs needs to be selected.

Conditional probabilities				Joint probabilities p_{ij}			
	New PSU j				New PSU j		
Former PSU i	j=1	j=2	j=3	Former PSU i	j=1	j=2	j=3
i=1 $p_1.$=.36	1.0	0	0	i=1 $p_1.$=.36	.36	0	0
i=2 $p_2.$=.24	0	1.0	0	i=2 $p_2.$=.24	0	.24	0
i=3 $p_3.$=.4	?	?	.5	i=3 $p_3.$=.4	?	?	.2
	$p_{.1}$=.5	$p_{.2}$=.3	$p_{.3}$=.2		$p_{.1}$=.5	$p_{.2}$=.3	$p_{.3}$=.2
					diagonal sum = .36 +.24 +.2 = .8		

The joint probability table is very easy to construct. The rows must add to the former design sampling probabilities and the columns must add to the new design probabilities. Also, the p_{ij} entries must all add to 1.0. The joint probabilities shown are the products of the initial former design probabilities times the conditional probabilities (diagonal entries: .36 x 1.0 = .36 for PSU 1; .24 x 1.0 = .24 for PSU 2; .4 x .5 = .2 for PSU 3). For the PSU 1 row, the off-diagonal joint elements are zero; that is, never choose a

different PSU for the new design if PSU 1 was sampled for the former design. However, in the PSU 1 column something is missing. There is a .36 chance that PSU 1 was sampled for the former design, then carried over to the new design, but to give PSU 1 the desired .5 unconditional probability for the new design, it needs an extra .14 in its column. Similarly, PSU 2 needs an extra .06 in its column. The completed tables follow.

Conditional probabilities				Joint probabilities p_{ij}			
	New PSU j				New PSU j		
Former PSU i	j=1	j=2	j=3	Former PSU i	j=1	j=2	j=3
i=1 $p_1.$=.36	1.0	0	0	i=1 $p_1.$=.36	.36	0	0
i=2 $p_2.$=.24	0	1.0	0	i=2 $p_2.$=.24	0	.24	0
i=3 $p_3.$=.4	.35	.15	.5	i=3 $p_3.$=.4	.14	.06	.2
	new $p_{.1}$=.5	$p_{.2}$=.3	$p_{.3}$=.2		$p_{.1}$= .5	$p_{.2}$=.3	$p_{.3}$=.2
					diagonal sum = .36+.24 +.2 = .8		

The rows and columns of the joint probability table now add properly. Any conditional probability can be derived by dividing the corresponding joint probability by the former design probability of the row ($p_{ij}/p_i.$). For example, .35=.14/.4, .15=.06/.4, and .5=.2/.4 in the PSU 3 row. Any joint probability can be derived by multiplying the corresponding joint probability by the former design probability of the row. Continuing with the PSU 3 row, .14 = .35 x .4, .06 = .15 x .4, and .2 = .5 x .4.

The diagonal sum of .8 on the joint probability table is the unconditional probability of including the same PSU in the sample of both the former and new designs. It is as large as possible given that additivity to both the former probabilities (rows) and unconditional new probabilities must be preserved. That is, the diagonal sum has been maximized subject to row and column constraints. A mathematical description of the problem is:

$$\text{maximize} \sum_{i=1}^{3} p_{ii}$$
$$\text{subject to} \sum_{j=1}^{3} p_{ij} = p_{i.} \; i=1,2,3 \quad \text{and} \sum_{i=1}^{3} p_{ij} = p_{.j} \; j=1,2,3$$

This is a special case of the maximizing/minimizing problem that is called a transportation problem. The p_{ij} of the general transportation problem can be any nonnegative

variable, and the $p_{i.}$ and $p_{.j}$ are prespecified constants.

$$\text{maximize} \sum_{i=1}^{n} \sum_{j=1}^{m} c_{ij} \, p_{ij}$$
$$\text{subject to} \sum_{j=1}^{m} p_{ij} = p_{i.} \; i=1,..,n \quad \text{and} \sum_{i=1}^{n} p_{ij} = p_{.j} \; j=1,..,m$$

$$\sum_{i=1}^{n} p_{i.} = \sum_{j=1}^{m} p_{.j}$$

It is this generalized form of the transportation problem that has been adapted for the CPS. For the 1990s design, PSUs were redefined. In nonself-representing areas, new 1990s design strata were formed. A new stratum can be formed from pieces of several already existing strata. Even a simple example indicates the type of difficulties that may arise when the stratum structure is changed. Suppose a stratum is unchanged, except that to one PSU a county has been added from a previously existing PSU from a different previously existing stratum. The new design unconditional probabilities of selection are no more difficult to calculate than before, but several factors complicate maximizing the overlap. Now two previously existing strata contribute to the new stratum. For the unchanged PSUs, there is still a 1-1 matchup between the former design and the new design, but the changed PSU is linked to two previously existing PSUs, each of which had a chance of selection.

Note that the added county had no independent chance of selection in the former design, but certainly contributed to the chance of selection of its previously existing PSU. Joint former/new design probabilities are determined that maximize the overlap between designs, taking into account such difficulties and maintaining the proper probability structure. A few numerical examples illustrate how this and other simple changes in the stratum PSU structure are dealt with.

ADDING A COUNTY TO A PSU

Suppose a stratum changes between the former and new designs only by adding a county to one of its three PSUs. To simplify notation, let the stratum in the former design consist of PSUs a, b, and c_{old} with probabilities of selection $p(a)=.36$, $p(b)=.24$, and $p(c_{old})=.4$. In the new design, the stratum consists of PSUs a, b, and c_{new} with probabilities of selection $p(a)=.5$, $p(b)=.3$, and $p(c_{new})=.2$. In response to a declining population, county c_{add} has been added to c_{old} to create c_{new}. Suppose c_{add} was part of former design PSU d, and let $p(d)=.2$ be the former design probability of selection for PSU d in its stratum. Also suppose c_{add} accounts for one-fourth of the new design selection size of c_{new} (c_{add} has size .05). The Former Design Probabilities column in the table below gives all of the former design selection possibilities and their probabilities that sum to 1: either a, b, or c_{old} was sampled; PSU d may or may not have been sampled. The joint probability entries are constrained to sum to the probabilities of selection for both the former design (rows) and the new design PSUs (columns). The computer programs that solve the general transportation problem ensure that the maximum sum of joint overlap probabilities is achieved (.77). The following reasoning can be used to arrive at the solution.

a only   Select PSU a with conditional probability 1.0, since the former design probability .288 is less than the new probability .5. Enter .288 as the joint probability.

a and d   We want to select PSU a, if possible, since its relative size (.5) is much larger than the size of the c_{add} county of PSU d (.05). Since the sum of former design probabilities (.288 + .072) is less than the new design probability .5, PSU a can be selected with conditional probability 1.0.

b only   Select PSU b since .192 is less than the new probability .3.

b and d   Select PSU b since its relative size (.3) is much larger than the c_{add} county size (.05), and .192 + .048 < .3.

c_{old} and d   Select PSU c_{new}. This possibility takes precedence over c_{old} only, since there is more overlap.

c_{old} only   It is advantageous to select c_{new}, but an extra joint probability of $.12=.2-.08$ is all that can be allowed (the column must add to .2). The corresponding conditional probability is $.375= .12/.32$. PSU c_{old} comprises three-fourths of PSU c_{new}, and that factor is entered in parentheses after the joint probability. The entries in the PSU a and b columns ensure the proper additivity of the joint probabilities, and the (0) after the joint probabilities indicate no overlap.

Former design probabilities	Conditional probabilities			Joint probabilities		
	New PSU			New PSU		
	a	b	c_{new}	a	b	c_{new}
p(a only) = .36(1-.2) =.288	1.0	0	0	.288 (1)	0	0
p(a and d)= .36(.2) =.072	1.0	0	0	.072 (1)	0	0
p(b only) = .24(1-.2) =.192	0	1.0	0	0	.192 (1)	0
p(b and d)= .24(.2) =.048	0	1.0	0	0	.048 (1)	0
p(c_{old} and d) = .4(.2) =.08	0	0	1.0	0	0	.08 (1)
p(c_{old} only)= .4(1-.2) =.32	.4375	.1875	.375	.14 (0)	.06 (0)	.12 (3/4)
	p(a)= .5	p(b)=.3	p(c_{new})=.2	p(a)= .5	p(b)=.3	p(c_{new})=.2

overlap sum = .288 +.072 + .192 + .048 + .08 + .12 (3/4) = .77

DROPPING A COUNTY FROM A PSU

Suppose a stratum changes between the former and new designs only by dropping a county from one of its three PSUs. Let the stratum in the former design consist of PSUs a, b, and c_{old} with probabilities of selection $p(a)=.36$, $p(b)=.24$, and $p(c_{old})=.4$. In the new design, the stratum consists of PSUs a, b, and c_{new} with probabilities of selection $p(a)=.5$, $p(b)=.3$, and $p(c_{new})=.2$. County c_{drop} has been dropped from c_{old} to create c_{new}. Note that dropping

a county explains part of the former-to-new design decrease in probability for PSU c, but the solution for maximum overlap is really the same as for the first example in this appendix (see the table below). The (1) after the .2 joint probability entry means that (1) if c_{old} was sampled for the stratum in the former design and (2) c_{new} is selected for the stratum in the new design, then (3) all of c_{new} overlaps with the former design.

Former design probabilities	Conditional probabilities			Joint probabilities		
	New PSU			New PSU		
	a	b	c_{new}	a	b	c_{new}
p(a)= .36	1.0	0	0	.36 (1)	0	0
p(b)= .24	0	1.0	0	0	.24 (1)	0
p(c_{old})= .4	.35	.15	.5	.14 (0)	.06 (0)	.2 (1)
	p(a)=.5	p(b)=.3	p(c_{new})=.2	p(a)=.5	p(b)=.3	p(c_{new})=.2
						Overlap sum = .36 +.24 + .2 = .8

DROPPING AND ADDING PSU'S

Suppose one PSU is dropped from a stratum and another added. Let the stratum in the former design consist of PSUs a, b, and c with probabilities of selection p(a)=.36, p(b)=.24, and p(c)=.4. In the new design, the stratum consists of PSUs b, c, and d (drop a; add d) with probabilities of selection p(b)=.3, p(c)=.2, and p(d)=.5. In the former design p(d)=.5 in its stratum. The Former Design Probabilities column in the table below gives all of the former design selection possibilities and their probabilities that sum to 1: either a, b or c was sampled in the previously existing stratum; PSU d may or may not have been sampled. The joint probability entries are constrained to sum to the probabilities of selection for both the former design (rows) and the new design PSUs (columns). The computer programs that solve the general transportation problem ensure that the maximum sum of joint overlap probabilities is achieved (.82). The following reasoning can be used to arrive at the solution.

d only Select PSU d with conditional probability 1.0, since the former design probability .18 is less than the new probability .5.

b only Select PSU b with conditional probability 1.0, since the former design probability .12 is less than the new probability .3.

c only Select PSU c with conditional probability 1.0, since the former design probability .2 does not exceed the new probability .2.

d and c Select PSU d with conditional probability 1.0. PSU c cannot be specified since its joint probability column already adds to the new probability .2.

d and b Select PSU d with conditional probability 1.0. PSU b could have been specified, but selecting PSU d provides more overlap with the former design.

none It is possible (probability .18) that none of the new stratum PSUs were sampled for the former design. Selecting PSU b with conditional probability 1.0 ensures that all the joint probability columns sum to the new design PSU probabilities. The (0) after the joint probability indicates that there is no overlap.

Former design probabilities	Conditional probabilities			Joint probabilities		
	New PSU			New PSU		
	d	b	c	d	b	c
p(d only) = .5(1-.24-.4) =.18	1.0	0	0	.18 (1)	0	0
p(b only)= .24(1-.5) =.12	0	1.0	0	0	.12 (1)	0
p(c only) = .4(1-.5) =.2	0	0	1.0	0	0	.2(1)
p(d and c)= .5(.4) =.2	1.0	0	0	.2 (1)	0	0
p(d and b)= .5(.24) =.12	1.0	0	0	.12 (1)	0	0
p(none)= (1-.5)(1-.24-.4) =.18	0	1.0	0	0	.18 (0)	0
	p(d)=.5	p(b)=.3	p(c)=.2	p(d)=.5	p(b)=.3	p(c)=.2
						overlap sum = .18 +.12 + .2 + .2 + .12 = .82

REFERENCES

Causey, B. D., Cox, L.H., and Ernst, L.K. (1985), "Applications of Transportation Theory to Statistical Problems," **Journal of the American Statistical Association,** 80, 903-909.

Ernst, L. (1986), "Maximizing the Overlap Between Surveys When Information Is Incomplete," **European Journal of Operations Research,** 27, 192-200.

Ernst, L. (1990), "Formulas for Approximating Overlap Probability When Maximizing Overlap," U.S. Census Bureau, Statistical Research Division, Technical Note Series, SRD Research Report Number Census/SRD/TN-90/01.

Gunlicks, C. (1992), "PSU Stratification and Selection Documentation: Comparison of the Overlap Program and the Expected Number of 1980 PSUs Selected Without the Overlap Program in the 1990 Redesign PSU Selection," (S-S90 DC-6), Memorandum to documentation, July 14, 1992.

Keyfitz, N. (1951), "Sampling With Probabilities Proportional to Size: Adjustment for Changes in Probabilities," **Journal of the American Statistical Association,** 46, 105-109.

Perkins, W. M. (1970), "1970 CPS Redesign: Proposed Method for Deriving Sample PSU Selection Probabilities Within 1970 NSR Strata," Memorandum to J. Waksberg, U.S. Census Bureau, August 5, 1970.

Perkins, W. M. (1971), "1970 CPS Redesign: Illustration of Computation of PSU Probabilities Within a Given 1970 Stratum," Memorandum to J. Waksberg, U.S. Census Bureau, February 19, 1971.

Raj, D. (1968), **Sampling Theory,** New York: McGraw Hill.

Roebuck, M. (1992), "1990 Redesign: PSU Selection Specifications" (Document #2.4-R-7), Memorandum to B. Walter through G. Shapiro and L. Cahoon, September 25, 1992.

Appendix B.
Sample Preparation Materials

INTRODUCTION

Despite the conversion of CPS to an automated interviewing environment, a number of paper materials are still in use. These materials include segment folders, listing sheets, maps, etc. Regional office staff and field representatives use these materials as aids in identifying and locating selected sample housing units. This appendix provides illustrations and explanations of these materials by frame (unit, area, group quarters, and permit). The information provided here should be used in conjunction with the information in Chapter 4 of this document.

UNIT FRAME MATERIALS

Segment Folders, (BC-1669 (CPS)) Illustration 1

A field representative receives a segment folder for each segment assigned for listing, updating, or interviewing. Segment folders were designed to:

1. Hold all materials needed to complete an assignment.

2. Provide instructions on listing, updating, and sampling.

The segment folder cover provides the following information about a segment:

1. Identifying information about the segment such as the regional office code, the field primary sampling unit (PSU), place name, sample designations for the segment, and basic geography.

2. Instructions for listing and updating. Part I of the segment folder contains "Field Representative Listing and Updating Instructions." This information is provided by the regional office and varies by segment type.

3. Sampling instructions. Part II on the segment folder is used only for area and group quarters segments. The regional offices use this section, "Regional Office Sampling Instructions," to record Start-with/Take-everys as described in Chapter 4.

4. Instructions for precanvassing and determining year built in area segments. Part III on the segment folder is reserved for stamps that provide the field representatives with information on whether or not to precanvass and whether or not to determine the year a structure was built.

5. Helpful information about the segment entered by the regional office or the field representative in the Remarks Section.

A segment folder for a unit segment may contain some or all of the following: Unit/Permit Listing Sheets (Form 11-3), Incomplete Address Locator Actions (Form BC-1718 (ADP)), 1990 Combined Reference File (CRF) listings, and 1990 Census Maps (county or census spotted).

A segment folder for an area segment may contain some or all of the following: Area Segment Listing Sheets (Form 11-5), Group Quarters Listing Sheets (Form 11-1), Area Segment Maps, Segment Locator Maps, County Maps, and Map Legends.

A segment folder for a group quarters segment may contain some or all of the following: Group Quarters Listing Sheets (Form 11-1), Incomplete Address Locator Actions (Form BC-1718 (ADP)), 1990 Census Maps (county or census spotted), and 1990 Combined Reference File listings.

A segment folder for a permit segment may contain some or all of the following: Unit/Permit Listing Sheets (Form 11-3), photocopies of the listed Permit Address List (PAL) (Form 11-193A), and Permit Sketch Maps (Form 11-187).

Unit/Permit Listing Sheets (Form 11-3) Illustrations 2–4

Unit/Permit Listing Sheets are provided for all multiunit addresses in unit segments. For each sample housing unit at a multiunit address, the field representative receives a preprinted, computer generated listing. This listing aids the field representative in locating the sample units and, in most cases, will eliminate the need to relist the entire multiunit address.

Listings for multiunit addresses are sorted into two categories: small and large. The type of multiunit listing is identified below the listing sheet title. The differences between these listings are identified in Table B–1.

Illustration 1. Segment Folder, BC-1669(CPS)

FORM **BC-1669(CPS)**
(11-17-92)

U.S. DEPARTMENT OF COMMERCE
BUREAU OF THE CENSUS

SEGMENT FOLDER

CPS

PSU

SEGMENT

Survey

RO	PSU	SEG #	SEGTYPE	SUR
SAMPLE	ESD	PAN/ROT	QTR	STR
ZIP	TEA	P/N	U/R	SSG

PLACE/MCD:

COUNTY:

Part I – FIELD REPRESENTATIVE LISTING AND UPDATING INSTRUCTIONS

ALL SEGMENTS
(See year built instructions in Part III, item A for area segments.)

SAMPLE →	List/update only	List/verify at time of interview	SAMPLE →	List/update only	List/verify at time of interview

Part II – REGIONAL OFFICE SAMPLING INSTRUCTIONS

AREA SEGMENTS AND GROUP
QUARTERS SEGMENTS

SAMPLE →	Start with	Take every	SAMPLE →	Start with	Take every

Part III – STAMPS

PSU

SEG

REMARKS

	DATE	
	Mo/Yr	Mo/Yr Mo/Yr Mo/Yr
Reduced		
Reinstated		
Reassigned		

Table B–1. **Listings for Multiunit Addresses**			
Listing type	Expected number of units	Listing characteristics	Illustration number
Small	2-9	Lists all expected units and could have some units with missing and/or duplicate unit designations as recorded in the 1990 census	2
Large	10 or more	Lists sample units only. On rare occasions, lists both sample and nonsample units	3
Large (Special Exception)	10 or more	No unit designations are listed and the Footnotes Section instructs the field representative to "List all units at the basic address."	4

The field representative also receives blank Unit/Permit Listing Sheets for use when an address needs relisting or when the field representative finds more units than expected causing him/her to run out of lines on the preprinted listing sheet.

Single-unit addresses in unit segments are not listed on Unit/Permit Listing Sheets.

Incomplete Address Locator Actions (Form BC-1718(ADP)) Illustration 5

The Incomplete Locator Actions form is used when the information obtained from the 1990 census is in some way incomplete (i.e., missing a house number, unit designation, etc.). The form provides the field representative with additional information that can be used to locate the incomplete address. The information on the Incomplete Address Locator Actions includes:

1. The address as it appeared in the Combined Reference File and possibly a complete address resulting from research conducted by Census Bureau staff;

2. A list of additional materials provided to the field representative to aid in locating the address; and

3. An outline of the actions the field representative should take to locate the address.

After locating the address, the field representative completes or corrects the basic address on the Incomplete Address Locator Actions form.

1990 Combined Reference File Listings

For each assigned segment that has at least one incomplete basic address, the field representative receives a computer-generated printout of addresses from the 1990 Combined Reference File. Each printout contains one or more incomplete addresses along with complete addresses in the same block.

The field representative uses the Combined Reference File listing by finding the incomplete basic address on the listing. Then the field representative uses the addresses listed directly before and after it to physically locate the incomplete address.

1990 Census Maps (County or Census Spotted)

For each assigned segment that has at least one incomplete basic address, the field representative receives a 1990 county or census-spotted map. For each incomplete basic address, the following control numbers are entered in red at the top of the map: field PSU number, sample designation, segment number, and serial number. The map spot of the incomplete address will be circled in red. The field representative uses the surrounding map spots to locate the incomplete address.

AREA FRAME MATERIALS

Segment Folder (BC-1669(CPS)) Illustration 1

See the description and illustration in the Unit Frame Materials above.

Area Segment Listing Sheets (Form 11-5) Illustration 6

Area Segment Listing Sheets are used to record information about housing units within the assigned block and to record information on the year a structure was built (when necessary). A field representative receives several Area Segment Listing Sheets for a block since the actual number of housing units in that block is unknown until the listing is complete. Only one of the listing sheets has the heading items filled in by the regional office. The heading items include: regional office code, field PSU, segment number, and survey acronym.

Group Quarters Listing Sheets (Form 11-1) Illustration 10

See the description and illustration provided in the Group Quarters Frame Materials below.

Area Segment Map Illustration 7

An Area Segment Map is a large scale map of the area segment, but it does not show roads or features outside of the segment boundaries. Within the segment, the map

Illustration 2. Unit/Permit Listing Sheet (Unit Segment)

FORM **11-3** (9-8-92)	U.S. DEPARTMENT OF COMMERCE BUREAU OF THE CENSUS	RO 24	PSU 26081	Segment 2958	CD	Type UNIT	Exp. Units 7

UNIT/PERMIT LISTING SHEET

Small Multi-Unit Sample and Non-Sample Units

	Survey CPS	ST-County MI–KENT		Com. Add.

Address	Permit office:
1435 MAIN ST	Permit #: DO: 2418 Permit day of issue: ARA: 1024 Permit year/month: Block: 203

Post Office GRAND RAPIDS	ZIP Code 49508	Permit year/month: PAL sequence/line #: PAL keyed remarks:

Line No. (1)	Unit designation (Or apartment number) (2)	Sample Desig. (3)	Serial number (4)	Remarks (Reason and date for changes) (5)	CENSUS ID
1	B1	A64	01		2846723
2	B2				2846731
3	101				2846756
4	102				2846797
5	103	A64	02		2846815
6	201	A64	03		2846856
7	202				2846879
8		A64			
9					
10					
11					
12		A64			
13		A64			
14					
15		A64			
16					
17					
18					
19		A64			
20		A64			

Multi-units	Listed and updated									
Name of complex:	FR Code									
Contact person:	FR Int.									
Title:	Mo/Year									
Telephone No.:	Total No. of units									

Footnotes	Mobile home park? ☐ Yes ☐ No

Sheet _1_ of _1_ sheets

(SEE INSTRUCTIONS ON REVERSE SIDE)

Illustration 3. Unit/Permit Listing Sheet (Multiunit Structure)

FORM **11-3** (9-8-92)	U.S. DEPARTMENT OF COMMERCE BUREAU OF THE CENSUS	RO 30	PSU 48155	Segment 2824	CD	Type UNIT	Exp. Units 18

UNIT/PERMIT LISTING SHEET

Large Multi-Unit; Sample Units Only

| Survey CPS | ST-County TX–WALLER | | Com. Add. |

Address		Permit office:	
1806 BELMONT STREET		Permit #:	DO: 2898
		Permit day of issue:	ARA: 1104
Post Office WALLACE	ZIP Code 77484	Permit year/month: PAL sequence/line #: PAL keyed remarks:	Block: 208

Line No. (1)	Unit designation (Or apartment number) (2)	Sample Desig. (3)	Serial number (4)	Remarks (Reason and date for changes) (5) CENSUS ID
1	104	A65	01	3142155
2	105	A65	02	3142156
3	106	A65	03	3142159
4	202	A66	01	3142163
5	205	A66	02	3142165
6	206	A66	03	3142166
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				

Multi-units	Listed and updated								
Name of complex:	FR Code								
Contact person:	FR Int.								
Title:	Mo/Year								
Telephone No.:	Total No. of units								

Footnotes	Mobile home park?
	☐ Yes
	☐ No

Copy distribution: **WHITE** – Regional Office   **YELLOW** – DPD/SMQCB after listing   **PINK** – DPD/SMQCB before listing

Sheet 1 of 1 sheets

(SEE INSTRUCTIONS ON REVERSE SIDE)

Illustration 4. Unit/Permit Listing Sheet (Large Special Exception)

FORM **11-3** (9-8-92)		U.S. DEPARTMENT OF COMMERCE BUREAU OF THE CENSUS	RO	PSU	Segment	CD	Type	Exp. Units
UNIT/PERMIT LISTING SHEET			30	48155	2824		UNIT	30
Large Multi-Unit; Sample Units Only			Survey CPS	ST-County TX–WALLER				Com. Add.

Address		
1806 BELMONT STREET		

Permit office:
    DO: 2898
Permit #:   ARA: 1104
Permit day of issue:
Permit year/month: Block: 208
PAL sequence/line #:
PAL keyed remarks:

Post Office	ZIP Code
WALLACE	77484

Line No. (1)	Unit designation (Or apartment number) (2)	Sample Desig. (3)	Serial number (4)	Remarks (Reason and date for changes) (5)
1				
2		A62	01	
3		A62	02	
4				
5		A65	01	
6				
7		A62	03	
8		A62	04	
9				
10		A65	02	
11		A65	03	
12		A64	01	
13		A64	02	
14		A64	03	
15		A64	04	
16				
17		A65	04	
18		A63	01	
19				
20		A63	02	

Multi-units	Listed and updated								
Name of complex:	FR Code								
Contact person:	FR Int.								
Title:	Mo/Year								
Telephone No.:	Total No. of units								

Footnotes	Mobile home park?
List All Units at the Basic Address	☐ Yes ☐ No

Sheet 1 of 3 sheets

(SEE INSTRUCTIONS ON REVERSE SIDE)

Illustration 5. Incomplete Address Locator Actions, BC-1718 (ADP)

```
FORM BC-1718 (ADP)                                    U.S. Department of Commerce
(05-05-93)                                                 Bureau of the Census
                              1990 SAMPLE REDESIGN
                       INCOMPLETE ADDRESS LOCATOR ACTIONS
```

SECTION I - IDENTIFICATION

Survey CPS	Sample A65	County 029	DO 2111	ARA 101B	Block 605A	Census ID 2501234
Frame Unit	Sample Date 199409	Tract/BNA 4011	RO 28	PSU 21029	Segment 5999	Serial # 04

SECTION II - DPD LOCATOR REVIEW/RESULTS

1. Incomplete Basic Address BEFORE DPD Review:

   Address: SPRING STREET
   Post Office: LINCOLN CITY   State: KY   ZIP: 40215

2. Address AFTER DPD Review:
 	
   Address: P.O. Box 202C Spring Street
   Post Office: Lincoln City   State: KY   ZIP: 40215
	Expected # Units
	
	_1Ø_

3. 1990 Census
   Household Name:

4. Locator Materials DPD is providing Field:

   [X]   Census Spotted Map   [X]   Printout of Addresses for Entire Block
   [ ]   County Locator Map   [ ]   Printout of Addresses for Part of Block
   [ ]   Other Type of Map   [ ]   Unit/Permit Listing Sheet for
   [ ]   Other - Specify below: Clerically Combined Multi-Unit

SECTION III   - ACTIONS THE RO/FR MUST PERFORM

To locate the sample unit:
    [X]   Use Locator Materials Enclosed
    [ ]   Use Single-Unit. Complete Address Information in II.2 Above
    [ ]   Use Multi-Unit Listing Sheet Enclosed
    [ ]   Use Multi-Unit Listing Sheet, previously sent to you for:
                    Sample:
                    Sample Date:
    [ ]   Other - Specify:

REMARKS/COMMENTS

Illustration 6. Area Segment Listing Sheet

FORM **11-5** (2-16-93)	U.S. DEPARTMENT OF COMMERCE BUREAU OF THE CENSUS	**1.** RO 22	**2.** PSU 36067	**3.** Segment Number O813	**4.** Survey CPS
AREA SEGMENT LISTING SHEET		**5.** Post Office and ZIP Code Canton City 13604			**6. NHIS** Week/Stratum

7. Listing

Line No. (a)	Street Name (b)	House No.	OR Description/Location of unit (c)	Sample Desig. (d)	Serial No. (e)	Remarks Reason & date for changes (f)	Year Built B = Before 4/1/90 A = After 4/1/90 ND = Not to be determined (g)
1			Name Extra Unit:				B A ND
2			Name Extra Unit:				B A ND
3			Name Extra Unit:				B A ND
4			Name Extra Unit:				B A ND
5			Name Extra Unit:				B A ND
6			Name Extra Unit:				B A ND
7			Name Extra Unit:				B A ND
8			Name Extra Unit:				B A ND
9			Name Extra Unit:				B A ND
10			Name Extra Unit:				B A ND
11			Name Extra Unit:				B A ND
12			Name Extra Unit:				B A ND
13			Name Extra Unit:				B A ND
14			Name Extra Unit:				B A ND
15			Name Extra Unit:				B A ND

8. Listed and Updated

a. FR code/ initials

b. Date Month/ Year

9. Footnotes

Copy distribution: **WHITE** – Segment Folder   **YELLOW** – Regional Office

Sheet _____ of _____ sheets

includes such features as highways, streets, roads, trails, walkways, railroads, bodies of water, military installations, and parks. The area segment map is used for:

1. Determining the exact location and boundaries of the area segment, and

2. Mapspotting the locations of housing units when listing the area segment.

Segment Locator Map Illustration 8

Segment Locator Maps are used to aid in the location of the area segment. The map identifies street patterns and names surrounding the exterior of the area segment. The location of the area segment is a shaded area, centered on the segment locator map.

County Map

Field representatives receive county maps for each field PSU in which they work. The map is divided into blocks or grids. Each segment locator map is one or more blocks or grids from the county map.

Map Legend Illustration 9

Map Legends are used to help in the identification of boundaries and features from the symbols and names printed on the area segment map and the segment locator map.

GROUP QUARTERS FRAME MATERIALS

Segment Folder (BC-1669(CPS)) Illustration 1

See the description and illustration in the Unit Frame Materials above.

Group Quarters Listing Sheets (Form 11-1) Illustration 10

The Group Quarters Listing Sheet is used to record the name, type and address of the group quarters, the name and telephone number of the contact person at the group quarters and to list the eligible units within the group quarters.

In group quarters segments, the field representative receives group quarters listing sheets with certain information preprinted by the computer. This information includes: group quarters name and address, group quarters type, group quarters ID number (from the 1990 census), regional office code, field PSU, segment number, etc.

In area segments, field representatives receive blank group quarters listing sheets since at the time of the initial listing, it is not known whether or not the block contains any group quarters.

Incomplete Address Locator Actions (Form BC-1718(ADP)) Illustration 3

See the description and illustration in the Unit Frame Materials above.

1990 Census Maps (County or Census Spotted)

See the description in the Unit Frame Materials above.

1990 Combined Reference File Listings

See the description in the Unit Frame Materials above.

PERMIT FRAME MATERIALS

Segment Folder (BC-1699(CPS)) Illustration 1

See the description and illustration in the Unit Frame Materials above.

Unit/Permit Listing Sheets (Form 11-3) Illustration 11

For each permit address in sample for the segment, the field representative receives a Unit/Permit Listing Sheet with heading information, sample designations, and serial numbers preprinted on it. The field representative also receives blank Unit/Permit Listing Sheets in case there are not enough lines on the preprinted listing sheet(s) to list all units at a multiunit address.

Permit Address List (Form 11-193A) Illustration 12

The permit address list (PAL) form is computer generated for each of the Building Permit Offices in sample. It is used to obtain necessary address information for individual permits issued. The PALs for Survey of Construction (SOC) building permit offices and non-SOC building permit offices are identical with one exception: For SOC building permit offices, the sample permit numbers, the date issued, and the number of housing units for each sample permit are preprinted on the PALs. The field representatives obtain the address information for these sample permits only. For

Illustration 7. Area Segment Map

Illustration 8. Segment Locator Map

Illustration 9. Map Legend

Current Population Survey

1997 CURRENT SURVEYS MAP LEGEND

SYMBOL DESCRIPTION	SYMBOL	NAME STYLE
International	☆ ☆ ☆ ☆ ☆ ☆ ☆ ☆	International
American Indian Reservation/Trust Land	X X X X X X X X	**American Indian Reservation/Trust Land**
STATE	∅ ∅ ∅ ∅ ∅ ∅ ∅ ∅	STATE
COUNTY	▦ ▦ ▦ ▦ ▦ ▦ ▦ ▦	COUNTY
Minor Civil Division [1]	⊕ ⊕ ⊕ ⊕ ⊕ ⊕ ⊕ ⊕	Minor Civil Division
Census County Division	⊕ ⊕ ⊕ ⊕ ⊕ ⊕ ⊕ ⊕	*Census County Division*
Incorporated Place	⊕ ⊕ ⊕ ⊕ ⊕ ⊕ ⊕ ⊕	**Incorporated Place**
Census Designated Place	⊕ ⊕ ⊕ ⊕ ⊕ ⊕ ⊕ ⊕	***Census Designated Place***
Highway	▬▬▬▬▬▬	Highway
Secondary Road	▬▬▬▬	Secondary Road
County Road	————	County Road
Neighborhood Road	————	Neighborhood Road
Railroad	╋━╋━╋━╋━╋	Railroad
Ferry	– – – – – – –	Ferry
Perennial Stream/Shoreline	————	*Perennial Stream/Shoreline*
Intermittent Stream/Shoreline	—— ·· —— ——	*Intermittent Stream/Shoreline*
Pipeline or Powerline	■—■—■—■—■	Pipeline or Powerline
Nonvisible Boundary	··············	Nonvisible Boundary
Airport	————	Airport
Block Number	1326	·
Fishhook	↩	
River, Lake, or Other Water		*River, Lake, or Other Water*
Military		Military
Airport		Airport
National Park		National Park
Park		Park
Subject Area		

✈	⋀	▦	⚓	🏛	✚
Airport	Campground	Cemetery	Golf Course	Gov't Bldg.	Hospital
⊨	▯	☗	▬	▲▲	▦
Hotel/Motel	Jail	Lookout Tower	Mil. Inst.	Mtn. Peak	Multi–Unit Hsg.
♣	✧	▣	⚑	⚏	▴
Nat'l. Park	Nursing Home	Orphanage	Park	Rel. Inst.	School
✇					
Shopping Ctr.					

[1] A five-spoked asterisk following a minor civil division name indicates that the minor civil division is coextensive with an incorporated place and has the same name.

Illustration 10. Group Quarters Listing Sheet

FORM **11-1** (9-8-92)	U.S. DEPARTMENT OF COMMERCE BUREAU OF THE CENSUS	(1) RO	(2) PSU	(3) Segment	(4) Survey
GROUP QUARTERS LISTING SHEET		30	48155	2981	CPS

(6) Name and Address Bob's Home for Wayward Children 10 State Street	(5) DO-ARA-Block 2204-2160-897
	(7) State-County-Tract 48-439-1555.04
Type of place:	(8) ZIP Code
(9) Post Office name Dallas, TX	77484

(10) Type code 15-N	(11b) Register is a record of	(11c) Type of register	(12) Group Quarters ID No. 1532547
(11a) Is a usable register available? ☐ Yes – *Fill b, c, d* ☐ No – *Begin listing*	_____ Living quarters _____ Persons	(11d) Location of register	(13) Military block flag

Line No. (14)	Group Quarters unit (15)	Sample Desig. (16)	Serial Number (17)	Remarks (18)
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				
24				
25				

(19) FR Code/Initials											
(20) Date (Mo/Year)											
(21) Number of GQ units											

Footnotes	(22)
	Contact person: Title: Telephone:

Copy distribution: WHITE – Regional Office segment folder   **YELLOW** – Regional Office files   **PINK** – DPD-SMQCB before listing   Sheet _1_ of _1_ sheets

(SEE INSTRUCTIONS ON REVERSE SIDE)

Illustration 11. Unit/Permit Listing Sheet

FORM **11-3** (9-8-92)		RO	PSU	Segment	CD	Type	Exp. Units
U.S. DEPARTMENT OF COMMERCE BUREAU OF THE CENSUS		32	06059	3201	2	PERMIT	12
UNIT/PERMIT LISTING SHEET		Survey	ST-County				Com. Add.
		CPS	06–059–URBAN				YES

Address

8438 CHERRY ROAD

Permit office: ANAHEIM TOWNSHIP

Permit #: 266378
Permit day of issue: 06/04
Permit year/month: 199206
PAL sequence/line #: 68067732/4
PAL keyed remarks: ANN KALE

Post Office	ZIP Code
ANAHEIM	87890

Line No. (1)	Unit designation (Or apartment number) (2)	Sample Desig. (3)	Serial number (4)	Remarks (Reason and date for changes) (5)
1		A64	01	
2		A64	02	
3		A64	03	
4		A64	04	
5		J22	01	
6		J22	02	
7		J22	03	
8		J22	04	
9				
10				
11				
12				
13		A64		
14		A64		
15		A64		
16		A64		
17		J22		
18		J22		
19		J22		
20		J22		

Multi-units	Listed and updated									
Name of complex:	FR Code									
Contact person:	FR Int.									
Title:	Mo/Year									
Telephone No.:	Total No. of units									

Footnotes

ALSO IN NCVS, J22, 6201

Mobile home park?

☐ Yes

☐ No

Copy distribution: WHITE – Regional Office   **YELLOW** – DPD/SMQCB after listing   **PINK** – DPD/SMQCB before listing

Sheet 1 of 1 sheets

(SEE INSTRUCTIONS ON REVERSE SIDE)

Illustration 12. Permit Address List

SEQUENCE NUMBER 26846281

1. RO	2. PSU	3. ID code	4. Permit month/year	5. Permit Office name	6. County	7. State	8. SOC code
23	42017	076200	05/1994	Curwensville Borough	CLEARFIELD	PA	

X ALIGNMENT

9. Field Representative name: Kim Newman

Code	10. Return date	11. Charge to project (operation code 12)	12. Expected HUs	13. Total HUs	14. Discrepancy flag
A24	07 11 693	CPS – 7242	28	28	

15. Group Quarters permits issued: ☐ Yes – List on separate 11-193A ☒ No

16. ☐ TA ☒ SAM

STREET TYPES– Please use the appropriate abbreviation in column (g). (See PAL Listing Aid for additional abbreviations.)
Avenue – AV Boulevard – BV Circle – CL Court – CT Drive – DR Lane – LA Road – RD Square – SQ Street – ST Terrace – TR

Line number (a)	Permit number (b)	Date issued Mo. Day (c)	Number of HUs (d)	House number (e)	Street name (f)	Street type (g)	Duplicate add. flag (h)	Locality (Post office) (i)	ZIP Code (j)	(k)	(l)	(m)	Sketch made (n)	Keyed remarks (o)
1	B4752	0504		1250	GREEN OAK	DRO		CURWENSVILLE	16833					LOT 5 BLK 18
2	B4753	0506		2572	HILL	ST02								LOT 2 BLK 25
3	B4755	0512		1260	GREEN OAK	DR01								LOT 7 BLK 18
4	B4758	0512		1250	GREEN OAK	DR01								LOT 6 BLK 18
5	C4804	0520	20	4550	FREDERICK	CT								LOT 12 BLK 10
6	B4759	0525		2572	HILL	ST02								REAR APTS
7	B4761	0525		110085	16TH ST BLDG 3									LOT 2-5 BLK 9
8														
9														
0														

Unkeyed remarks – For incomplete address enter name, address, and phone number of builder or owner. See manual instructions for follow-up on incomplete address.

1 Master Builders, 500 Center St, Clearfield 768-5225
2 Corley Construction Co, 425 Anderson St, Curwensville 236-8710
3 Master Builders
4 Master Builders
5 Swallow Lake
6 Two units built earlier ones
7 Aspen Hills Apts Complex
8
9
0

ATTENTION REGIONAL OFFICE
Retain original in Regional Office files

PERMIT ADDRESS LIST
PERMIT ADDRESS OPERATION

FORM 11-193A (7-1-92) U.S. DEPARTMENT OF COMMERCE BUREAU OF THE CENSUS

Sheet 1 of 1 sheet(s)

Illustration 13. Permit Sketch Map

FORM **11-187** (10-15-92) U.S. DEPARTMENT OF COMMERCE BUREAU OF THE CENSUS	**1.** RO 30	**2.** PSU 48065	**3.** Permit month/year 05/98	**4.** Sequence number 75110871

PERMIT SKETCH MAP

PERMIT ADDRESS LIST OPERATION

5. Permit Office name Panhandle Town		
6. Locality (Post Office) Panhandle		**7.** County Carson

8. State TX	**9.** ZIP Code 25616	**10.** Permit number 6501	**11.** PAL line 2

ATTENTION REGIONAL OFFICE ▶ Keep both copies in the Regional Office files.

12. U/R Rural	**13.** Field representative name Samuel Butcherd	Code C-15

Quimby St.

☒◄— 1 mile

L2, #6501

↑ N

Hill Rd.

Hwy 64

Notes

non-SOC building permit offices, the PALs do not contain any preprinted permit numbers. Field representatives list all residential permits issued by the building permit office for a specific month.

Permit Sketch Map (Form 11-187) Illustration 13

When completing a PAL form, if the address given on a building permit is incomplete, the field representative attempts to obtain a complete address. If a complete address cannot be obtained, the field representative visits the new construction site and draws a Permit Sketch Map. The map shows the location of the construction site, streets in the vicinity of the site, and the directions and distance from the construction site to the nearest town.

Appendix C.
Maintaining the Desired Sample Size

INTRODUCTION

The Current Population Survey (CPS) sample is continually updated to include housing units built after the most recent census. If the same sampling rates were used throughout the decade, the growth of the U.S. housing inventory would lead to increases in the CPS sample size and, consequently, to increases in cost. To avoid exceeding the budget, the sampling rate is periodically reduced to maintain the desired sample size. Referred to as maintenance reductions, these changes in the sampling rate are implemented in a way that retains the desired set of reliability requirements.

These maintenance reductions are different from changes to the base CPS sample size resulting from modifications in the CPS funding levels. The methodology for designing and implementing this type of sample size change is generally dictated by new requirements specified by the Bureau of Labor Statistics (BLS). For example, the sample reduction implemented in January 1996 was due to a reduction in CPS funding and new design requirements were specified (see Appendix H). As a result of the January 1996 reduction, however, any impending maintenance reduction to the 1990 design sample because of sample growth was preempted.

MAINTENANCE REDUCTIONS

Developing the Reduction Plan

The CPS sample size for the United States is projected forward for about a year using linear regression based on previous CPS monthly sample sizes. The future CPS sample size must be predicted because CPS maintenance reductions are gradually introduced over 16 months and operational lead time is needed to prevent dropped cases from being interviewed.

The growth is examined in all states and major substate areas to determine whether it is uniform or not. The states with faster growth are candidates for a possible maintenance reduction. The remaining sample must be sufficient to maintain the individual state and national reliability requirements. Generally, the sample in a state is reduced by the same proportion in all frames in all primary sampling units (PSUs) to maintain the self-weighting nature of the state design.

Implementing the Reduction Plan

Reduction groups. The CPS sample size is reduced by deleting one or more subsamples of ultimate sampling units (USUs) from both the old construction and permit frames. The original sample of USUs is partitioned into 101 subsamples called reduction groups; each is representative of the overall sample. The decision to use 101 subsamples is somewhat arbitrary. A useful attribute of the number used is it is prime to the number of rotation groups (eight) so that reductions have a uniform effect across rotations. A number larger than 101 would allow greater flexibility in pinpointing proportions of sample to reduce. However, a large number of reduction groups can lead to imbalances in sample cut distribution across PSUs since small PSUs may not have enough sample to have all reduction groups represented. That is, when there are a large number of reduction groups randomly assigned, a smaller PSU may or may not experience reductions and quite possibly it may undergo too much reduction in the PSU.

All USUs in a hit string have the same reduction group number (see Chapter 3). For the unit, area, and group quarters (GQ) frames, hit strings are sorted and then sequentially assigned a reduction group code from 1 through 101. The sort sequence is:

1. State or substate.

2. Metropolitan statistical area/nonmetropolitan statistical area status.

3. Self-representing/nonself-representing status.

4. Stratification PSU.

5. Final hit number, which defines the original order of selection.

For the permit frame, a random start is generated for each stratification PSU and permit frame hits are assigned a reduction group code from 1 through 101 following a specific, nonsequential pattern. This method of assigning reduction group code is used to improve balancing of reduction groups because of small permit sample sizes in some PSUs and the uncertainty of which PSUs will actually have permit samples over the life of the design. The sort sequence is:

1. Stratification PSU.

2. Final hit number.

The state or national sample can be reduced by deleting USUs from both the old construction and permit frames in one or more reduction groups. If there are k reduction groups in sample, the sample may be reduced by 1/k by deleting one of k reduction groups. For the first reduction applied to redesigned samples, each reduction group represents roughly 1 percent of the sample. Reduction group numbers are chosen for deletion in a specific sequence designed to maintain the nature of the systematic sample to the extent possible.

For example, suppose a state has an overall state sampling interval of 500 at the start of the 1990 design. Suppose the original selection probability of 1 in 500 is modified by deleting 5 of 101 reduction groups. The resulting overall state sampling interval (SI) is

$$SI = 500 \times \frac{101}{101 - 5} = 526.0417$$

This makes the resulting overall selection probability in the state 1 in 526.0417. In the subsequent maintenance reduction, the state has 96 reduction groups remaining. A reduction of 1 in 96 can be accomplished by deleting 1 of the remaining 96 reduction groups.

The resulting overall state sampling interval is the new basic weight for the remaining uncut sample.

Introducing the reduction. A maintenance reduction is implemented only when a new sample designation is introduced and is gradually phased in with each incoming rotation group to minimize the effect on survey estimates and reliability and to prevent sudden changes to interviewer workloads. The basic weight applied to each incoming rotation group reflects the reduction. Once this basic weight is assigned, it does not change until future sample changes are made. In all, it takes 16 months for a maintenance sample reduction and new basic weights to be fully reflected in all eight rotation groups interviewed for a particular month. During the phase-in period, rotation groups have different basic weights; consequently, the average weight over all eight rotation groups changes each month. After the phase-in period, all eight rotation groups have the same basic weight.

Maintenance Reductions and GVFs

National generalized variance parameters are not updated for maintenance reductions because the purpose of these reductions is to maintain the sample size. However, the national generalized parameters should be updated to reflect any changes in sample size. For example, the parameters were updated for the January 1996 reduction since survey design parameters changed. (See Chapter 14 for more details.)

Appendix D.
Derivation of Independent Population Controls

INTRODUCTION

Each month, for the purpose of the iterative, second-stage weighting of the Current Population Survey (CPS) data, independent projections of the eligible population are produced by the Census Bureau's Population Division. While the CPS provides the fundamental reason for the existence of these monthly projections, other survey-based programs sponsored by the Bureau of the Census, the Bureau of Labor Statistics, the National Center for Health Statistics, and other agencies use these independent projections to calibrate surveys. These projections consist of the adjusted civilian noninstitutional population of the United States, distributed by demographic characteristics in three ways: (1) age, sex, and race, (2) age, sex, and Hispanic[1] origin, and (3) state of residence, restricted to the population 16 years of age and over. They are produced in association with the Population Division's population estimates and projections programs, which provide published estimates and long-term projections of the population of the United States, the 50 states and District of Columbia, the counties, and subcounty jurisdictions.

Organization of This Appendix

The CPS population controls, like other population estimates and projections produced by the Census Bureau, are based on a demographic framework of population accounting. Under this framework, time series of population estimates and projections are anchored by decennial census enumerations, with populations for dates between two previous censuses, since the last census, or in the future derived by the estimation or projection of population change. The method by which population change is estimated depends on the defining demographic and geographic characteristics of the population. The issue of adjustment for net underenumeration in the census is handled outside of this demographic framework, and is applied to the CPS control series as a final step in its production.

This appendix seeks first to systematically present this framework, defining terminology and concepts, then to describe data sources and their application within the framework. The first subsection is operational and is devoted to the organization of the chapter and a glossary of terms.

The second and third subsections deal with two broad conceptual definitions. The second subsection distinguishes population estimates from population projections and describes how monthly population controls for surveys "fit" into this distinction. The third subsection defines the "CPS control universe," the set of inclusion and classification rules specifying the population to be projected for the purpose of weighting CPS data. The fourth subsection, "Calculation of Population Projections for the CPS Control Universe," comprises the bulk of the appendix. It provides the mathematical framework and data sources for updating total population from a census date via the demographic components of change and how each of the requisite inputs to the mathematical framework is measured. This is presented separately for the total population of the United States, its distribution by demographic characteristics, and its distribution by state of residence. The fifth subsection, "Adjustment of CPS Controls for Net Underenumeration in the 1990 Census," describes the application of data from the 1990 Post-Enumeration Survey (PES) to include an estimate of net underenumeration in the 1990 census. The final two sections are, once again, operational. A subsection, "The Monthly and Annual Revision Process for Independent Population Controls" describes the protocols for incorporating new information in the series through revision. We conclude with a brief section, "Procedural Revisions," containing an overview of various technical problems for which solutions are being sought.

Terminology Used in This Appendix

The following is an alphabetical list of terms used in this appendix, essential to understanding the derivation of census-based population controls for surveys, but not necessarily prevalent in the literature on survey methodology.

An *adjusted* (as opposed to unadjusted) population estimate or projection includes an allowance for net undercoverage (underenumeration) in a census. In this appendix (as in other Census Bureau publications), the term, when applied to series based on the 1990 census, refers to incorporation of the results of the 1990 Post-Enumeration Survey (PES).

The *population base,* (or base population) for a population estimate or projection is the population count or estimate to which some measurement or assumption of population change is added to yield the population estimate or projection.

---

[1]Hispanics may be of any race.

A *census-level* population estimate or projection is one that can be linked to a census count through an estimate or projection of population change; specifically, it does not include an adjustment for underenumeration in the census.

The *civilian population* is the portion of the resident population not in the active-duty military. Active-duty military, used in this context, refers to persons defined to be in active duty by one of the five branches of the Armed Forces, as well as persons in the National Guard or the reserves actively participating in certain training programs.

Components of population change are any subdivisions of the numerical population change over a time interval. The demographic components often cited in the literature consist of births, deaths, and net migration. In this appendix, reference is also made to changes in the institutional and active-duty Armed Forces populations, which affect the CPS control universe.

The *CPS control universe* is characterized by three attributes: (1) restriction to the civilian noninstitutional population, (2) modification of census data by age, race, and sex (MARS), and (3) adjustment for net census underenumeration. Each of the three defining concepts appears separately in this glossary.

The *Demographic Analysis Population* ("DA" population) is a distribution of population by characteristics—in the present case age, sex, and race—developed by observation of births, deaths by age, and international migration by age, over a period of time for each sex and race category. It is derived independent of census data. Thus, the number of persons x years of age of a given sex and race is determined by the number of births x years before the reference date, as well as the number of deaths and net migration of persons born the same year during the x-year interval from birth date to reference date. In this text, we also refer to the "DA/MARS population." This is a hybrid distribution based on demographic analysis for age, sex, and three race groups, but on the census for Hispanic origin detail within these groups. "MARS" refers to the fact that the census detail is based on modified age and race, as defined elsewhere in this glossary.

Emigration is the permanent departure of a person from a country of residence, in this case the United States. In the present context, it refers to the departure of persons legally resident in the United States, but not confined to legally permanent residents (immigrants). Departures of undocumented residents are excluded, because they are included in the definition of net undocumented migration (elsewhere in this glossary).

Population *estimates* are population figures that do not arise directly from a census or count, but can be determined from available data (e.g., administrative); hence, they are not based on assumptions or modeling. Population estimates discussed in this appendix stipulate an enumerated base population, coupled with estimates of population change from the date of the base population to the date of the estimate.

Immigration is the acquisition, by a foreign national, of legal permanent resident status in the United States. Hence, an immigrant may or may not reside in the United States prior to immigration. This definition is consistent with the parlance of the Immigration and Naturalization Service (INS), but differs from popular usage, which associates immigration with an actual change of residence.

The *institutional population* refers to a population universe consisting of inmates of CPS-defined institutions, such as prisons, nursing homes, juvenile detention facilities, or residential mental hospitals.

Internal consistency of a population time series occurs when the difference between any two population estimates in the series can be construed as population change between the reference dates of the two populations estimates.

International migration is generally the change of a person's residence from one country to another. In the present context, this concept includes change of residence into the United States by non-U.S. citizens and by previous residents of Puerto Rico or the U.S. outlying areas, as well as the change of residence out of the United States by persons intending to live permanently abroad, in Puerto Rico, or the outlying areas.

Legal permanent residents are persons whose right to reside in the United States is legally defined either by immigration or by U.S. citizenship. In the present context, the term generally refers to noncitizen immigrants.

Modified age, race, and sex, abbreviated MARS, describes the census population after the definition of age and race has been aligned with other administrative sources (or with OMB Directive 15, in the case of race).

The *natural increase* of a population over a time interval is the number of births minus the number of deaths during the interval.

The *Post-Enumeration Survey* is a survey of households conducted after the 1990 population census in a sample of blocks to estimate the net underenumeration in the 1990 census.

Population *projections* are population figures relying on modeled or assumed values for some or all of their components, therefore not entirely calculated from actual data. Projections discussed in this appendix stipulate a base population that may be an estimate or count, and components of population change from the reference date of the base population to the reference date of the projection.

The *reference date* of an estimate or projection is the date to which the population figure applies. The CPS control reference date, in particular, is the first day of the month in which the CPS data are collected.

The *resident population* of the United States is the population usually resident in the 50 states and District of Columbia. For the census date, this population matches the total population in census decennial publications, although

applications to the CPS series beginning in 1995 include some count resolution corrections to the 1990 census not included in original census publications.

Undocumented migration (more precisely, net undocumented migration) is the net increase of population brought about by the residential migration of persons in and out of the country who have no legal basis for residence in the United States.

A population *universe* is a set of rules or criteria defining inclusion and exclusion of persons from a population, as well as rules of classification for specified geographic or demographic characteristics within the population. Examples include the resident population universe and the civilian population universe, both of which are defined elsewhere in this glossary. Frequent reference is made to the "CPS control universe," defined separately in this glossary.

CPS Population Controls: Estimates or Projections?

Throughout this appendix, the independent population controls for the CPS will be cited as population projections. Throughout much of the scientific literature on population, demographers take care to distinguish the concepts of "estimate" and "projection"; yet, the CPS population controls lie perilously close to a somewhat ragged line of distinction. Generally, population estimates relate to past dates, and in order to be called "estimates," must be supported by a reasonably complete data series. If the estimating procedure involves the computation of population change from a base date to a reference date, the inputs to the computation must have a solid basis in data. Projections, on the other hand, allow the replacement of unavailable data with assumptions.

The reference date for CPS population controls is indeed in the past, relative to their date of production. However, the past is so recent—3 to 4 weeks prior to production—that for all intents and purposes, CPS controls are projections. No data relating to population change are available for the month prior to the reference date; very little data are available for 3 to 4 months prior to the reference date; no data for state-level geography are available past July 1 of the year prior to the current year. Hence, we refer to CPS controls as projections, despite their frequent description as estimates in the literature.

Be this as it may, CPS controls are founded primarily on a population census and on administrative data. Analysis of coverage rates has shown that even without adjustment for underenumeration in the controls, the lowest coverage rates of the CPS often coincide with those age-sex-race-Hispanic segments of the population having the lowest coverage in the census. Therefore, the contribution of independent population controls to the weighting of CPS data is highly productive. This holds, despite the use of modeling for missing data.

THE POPULATION UNIVERSE FOR CPS CONTROLS

In the concept of "population universe," as used in this appendix, are not only the rules specifying what persons are included in the population under consideration, but also the rules specifying their geographic locations and their relevant characteristics, such as age, sex, race, and Hispanic origin. In this section we consider three population universes; the resident population universe defined by the 1990 census, the resident population universe used for official population estimates, and the CPS control universe that relates directly to the calculation of CPS population controls. These three universes are distinct from one another; each one is derived from the one that precedes it; hence, the necessity of considering all three.

The primacy of the decennial census in the definition of these three universes is a consequence of its importance within the Federal statistical system. This, in turn, is a result of the extensive detail that it provides on the geographic distribution and characteristics of the U.S. population, as well as the legitimacy accorded it by the U.S. Constitution. Later in this appendix, we cite another population universe—the Demographic Analysis population—that may be technically more appropriate as a base for estimates and projections than the 1990 census in some regards, mainly in the continuity of the age distribution and the conformity of racial categories with other administrative data sources. Because the census defines the universe for estimates, the use of the DA population in the process is intended to ensure the continuity of census-related biases over time in the estimates, not to eliminate them.

The universe defined by the 1990 census is the U.S. resident population, including persons resident in the 50 states and the District of Columbia. This definition excludes residents of the Commonwealth of Puerto Rico, and residents of the outlying areas under U.S. sovereignty or jurisdiction (principally American Samoa, Guam, Virgin Islands of the United States, and Commonwealth of the Northern Mariana Islands). The definition of residence conforms to the criterion used in the census, which defines a resident of a specified area as a person "usually resident" in the area. For the most part, "usual residence" is defined subjectively by the census respondent; it is not defined by de facto presence in a particular house or dwelling, nor is it defined by any de jure (legal) basis for a respondent's presence. There are two exceptions to the subjective character of the residence definition.

1. Persons living in military barracks, prisons, and some other types of residential group-quarters facilities are generally reported by the administration of their facility, and their residence is generally the location of the facility. Naval personnel aboard ships reside at the home port of the ship, unless the ship is deployed to the overseas fleets, in which case they are not U.S. residents. Exceptions may occur in the case of Armed

Forces personnel who have an official duty location different from the location of their barracks, in which case their place of residence is their duty location.

2. Students residing in dormitories report their own residences, but are instructed on the census form to report their dormitories, rather than their parental homes, as their residences.

The universe of interest to the official estimates of population that the Census Bureau produces is the resident population of the United States, as it would be counted by the last census (1990) if this census had been held on the reference date of the estimates. However, this universe is distinct from the census universe in two regards, neither of which affects the total population.

1. The estimates differ from the census in their definition of race, with categories redefined to be consistent with other administrative-data sources. Specifically, persons enumerated in the 1990 census who gave responses to the census race question not codable to White, Black, American Indian, Eskimo, Aleut, or any of several Asian and Pacific Islander categories were assigned a race within one of these categories. Most of these persons (9.8 million nationally) were persons of Hispanic origin who gave a Hispanic origin response to the race question. Publications from the 1990 census do not incorporate this modification, while all official population estimates incorporate it.

2. Respondents' ages are modified to compensate for the aging of census respondents from the census date (April 1, 1990) to the actual date that the census form was completed. This modification was necessitated by the fact that no printed instruction on the census form required respondents to state their age as of the census date. They therefore tended to state their age at the time of completion of the form, or possibly their next anticipated birthday. This modification is not included in census publications, which tabulate respondents' age as stated. It is applied to all population estimates and projections, including controls for the CPS.

The universe underlying the CPS sample is confined to the civilian noninstitutional population. Thus, independent population controls are sought for this population, which defines the CPS control universe. The previously mentioned alternative definitions of race and age, associated with the universe for official estimates, are carried over to the CPS control universe. Four additional changes are incorporated, which distinguish the CPS control universe from the universe for official estimates.

1. The CPS control universe excludes active-duty Armed Forces personnel, including those stationed within the United States, while these are included in the resident population. "Active duty" is taken to refer to personnel reported in military strength statistics of the Departments of Defense (Army, Navy, Marines, and Air Force) and Transportation (Coast Guard), including reserve forces on 3- and 6-months' active duty for training, National Guard reserve forces on active duty for 4 months or more, and students at military academies. Reserve forces not active by this definition are included in the CPS control universe, if they meet the other inclusion criteria.

2. The CPS control universe excludes persons residing in institutions, such as nursing homes, correctional facilities, juvenile detention facilities, and long-term mental health care facilities. The resident population, on the other hand, includes the institutional population.

3. The CPS control universe, like the resident population base for population estimates, includes students residing in dormitories. Unlike the population estimates base, however, it accepts as state of residence a family home address within the United States, in preference to the address of the dormitory. This affects estimation procedures at the state level, but not at the national level.

4. An added difference between the census official estimates and the CPS population controls after January 1, 1994, is that the former do not include adjustment for undercoverage in the 1990 census based on the Post-Enumeration Survey (PES) while the latter do. Details of this adjustment are given later in this appendix.[2]

CALCULATION OF POPULATION PROJECTIONS FOR THE CPS UNIVERSE

The present section is concerned with the methodology for computing population projections for the CPS control universe used in the actual calibration of the survey. Three subsections are devoted to the calculation of (i) the total population, (ii) its distribution by age, sex, race, and Hispanic origin, and (iii) its distribution by state of residence.

Total Population

Projections of the population to the CPS control reference date (the first of each month) are determined by a base population (either estimated or enumerated) and a

---

[2]Whether the presence of an adjustment for census undercount implies an actual change in universe, as opposed to the correction of a measurement problem, is a legitimate subject for dispute. In the context of the census itself, adjustment might be viewed as a measurement issue relative to the population of a "true" universe of U.S. residents. It is treated here as a technical attribute of the universe, since unadjusted population estimates are specifically adapted to remain consistent with the census in regard to coverage. Hence "census-level" and "adjusted" define distinct population universes.

projection of population change. The latter is generally an amalgam of components measured from administrative data and components determined by projection. The "balancing equation of population change" used to produce estimates and projections of the U.S. population states the population at some reference date as the sum of the base population and the net of various components of population change. The components of change are the sources of increase or decrease in the population from the reference date of the base population to the date of the estimate. The exact specification of the balancing equation depends on the population universe, and the extent to which components of population change are disaggregated. For the resident population universe, the equation in its simplest form is given by

$$R_{t1} = R_{t0} + B - D + NM \qquad (1)$$

where:
R_{t1} = Resident population at time t1 (the reference date),
R_{t0} = Resident population at time t0 (the base date),
B = births to U.S. resident women, from time t0 to time t1,
D = deaths of U.S. residents, from time t0 to time t1,
NM = net migration to the United States (migration into the United States minus migration out of the United States) from time t0 to time t1.

It is essential that the components of change in a balancing equation represent the same population universe as the base population. If we derive the current population controls for the CPS independently, from a base population in the CPS universe, civilian noninstitutional deaths would replace resident deaths, "net migration" would be confined to civilian migrants, but would incorporate net recruits to the Armed Forces and net admissions to the institutional population. The equation would thus appear as follows:

$$CNP_{t1} = CNP_{t0} + B - CND + (NCM - NRAF - NEI) \qquad (2)$$

where:
CNP_{t1} = civilian noninstitutional population, time t1,
CNP_{t0} = civilian noninstitutional population, time t0,
B = births to the U.S. resident population, time t0 to time t1,
CND = deaths of the civilian noninstitutional population, time t0 to time t1,
NCM = net civilian migration, time t0 to time t1,
NRAF = net recruits (inductions minus discharges) to the U.S. Armed Forces from the domestic civilian population, from time t0 to time t1,
NEI = net admissions (enrollments minus discharges) to institutions, from time t0 to time t1.

In the actual estimation process, it is not necessary to directly measure the variables CND, NRAF, and NEI in the above equation. Rather, they are derived from the following equations:

$$CND = D - DRAF - DI \qquad (3)$$

where D = deaths of the resident population, DRAF = deaths of the resident Armed Forces, and DI = deaths of inmates of institutions (the institutional population)

$$NRAF = WWAF_{t1} - WWAF_{t0} + DRAF + DAFO - NRO \qquad (4)$$

where $WWAF_{t1}$ and $WWAF_{t0}$ represent the U.S. Armed Forces worldwide at times t1 and t0, DRAF represents deaths of the Armed Forces residing in the United States during the interval, DAFO represents deaths of the Armed Forces residing overseas, and NRO represents net recruits (inductions minus discharges) to the Armed Forces from the overseas civilian population

$$NEI = CIP_{t1} - CIP_{t0} + DI \qquad (5)$$

where CIP_{t1} and CIP_{t0} represent the civilian institutional population at times t1 and time t0, and DI represents deaths of inmates of institutions during the interval. The appearance of DI in both equations (3) and (5) with opposite sign, and DRAF in equations (3) and (4) with opposite sign, ensures that they will cancel each other when applied to equation (2). This fact obviates the need to measure institutional inmate deaths or deaths of the Armed Forces residing in the United States.

A further disaggregation to facilitate the description of how the components on the right side of equation (2) are measured as follows:

$$NCM = IM + CCM \qquad (6)$$

where IM = net international migration, and CCM = net movement of civilian citizens to the United States from abroad, in the time interval (t0,t1).

At this point, we restate equation (2) to incorporate equations (3), (4), (5), and (6), and simplify. The resulting equation, which follows, is the operational equation for population change in the civilian noninstitutional population universe.

$$CNP_{t1} = CNP_{t0} + B - D + (IM + CCM) - (WWAF_{t1} - WWAF_{t0} + DAFO - NRO) - (CIP_{t1} - CIP_{t0}) \qquad (7)$$

Equation (7) identifies all the components that must be separately measured or projected to update the total civilian noninstitutional population from a base date to a later reference date.

Aside from forming the procedural basis for all estimates and projections of total population (in whatever universe), balancing equations (1), (2), and (7) also define a recursive concept of internal consistency of a population series within a population universe. We consider a time series of population estimates or projections to be internally consistent if any population figure in the series can be derived

from any earlier population figure as the sum of the earlier population and the components of population change for the interval between the two reference dates.

Measuring the Components of the Balancing Equation

Balancing equations such as equation (7) requires a base population, whether estimated or enumerated, and information on the various components of population change. Because a principal objective of the population estimating procedure is to uphold the integrity of the population universe, the various inputs to the balancing equation need to agree as much as possible with respect to geographic scope, treatment of underenumeration of population or underregistration of events, and residential definition, and this is a major aspect of the way they are measured. A description of the individual inputs to equation (7), which applies to the CPS control universe, and their data sources follows.

The *census base population (CNP$_{t0}$)*. While any base population (CNP$_{t0}$, in equation (7)), whether a count or an estimate, can give rise to estimates and projections for later dates, the original base population for all unadjusted postcensal estimates and projections is the enumerated population from the last census. Although survey controls are currently adjusted for underenumeration in the 1990 census, the census population is itself not adjusted for underenumeration in the census; rather, an adjustment difference is added in a later step. After the 2000 census, it is expected that official estimates will be adjusted for undercoverage, so that the adjustment can be incorporated in the base population. However, the census base population does include minor revisions arising from count resolution corrections. As these corrections arise entirely from the tabulation of data from the enumerated population, they are unrelated to adjustment for underenumeration. As of the end of 1994, count resolution corrections incorporated in the estimates base population amounted to a gain of 8,418 persons from the originally published count.

In order to apply equation (7), which is derived from the balancing equation for the civilian noninstitutional population, the resident base population enumerated by the census must be transformed into a base population consistent with this universe. The transformation is given as follows:

$$CNP_0 = R_0 - RAF_0 - CIP_0 \qquad (8)$$

where R_0 = the enumerated resident population, RAF_0 = the Armed Forces resident in the United States on the census date, and CIP_0 = the civilian institutional population residing in the United States on the census date. This is consistent with previous notation, but with the stipulation that $t_0 = 0$, since the base date is the census date, the starting point of the series. The transformation can also be used to adapt any postcensal estimate of resident population to the CPS universe for use as a population base in equation (7). In practice, this is what occurs when the monthly CPS population controls are produced.

Adjustment for underenumeration, while essential to the CPS universe, is handled independently of the process of producing projections. It is added after the population is updated. Hence, adjustment will be discussed later in this appendix.

Births and deaths (B, D). In estimating total births and deaths of the resident population (B and D, in equation (7)), we assume the population universe for vital statistics to match the population universe for the census. If we define the vital statistics universe to be the population subject to the natural risk of giving birth or dying, and having the event recorded by vital registration systems, the assumption implies the match of this universe with the census-level resident population universe. We relax this assumption in the estimation of some characteristic detail, and this will be discussed later in the appendix.

The numbers of births and deaths of U.S. residents are supplied by the National Center for Health Statistics (NCHS). These are based on reports to NCHS from individual state and local registries; the fundamental unit of reporting is the individual birth or death certificate. For years in which reporting is considered final by NCHS, the birth and death statistics are considered final; these generally cover the period from the census until the end of the calendar year 3 years prior to the year of the CPS series. For example, the last available final birth and death statistics available in time for the 1998 CPS control series were for calendar year 1995. Final birth and death data are summarized in NCHS publications (see Ventura et al., 1997a, Anderson et al., 1997). Monthly births and deaths for calendar year(s) up to 2 years before the CPS reference dates (e.g., 1996 for 1998 controls) are based on provisional estimates by NCHS (Ventura et al., 1997b). For the year before the CPS reference date through the last month before the CPS reference date, births and deaths are projected based on the population by age according to current estimates and age-specific rates of fertility and mortality and seasonal distributions of births and deaths observed in the preceding year.

Various concerns exist relative to the consistency of the NCHS vital statistics universe with the census resident universe and the CPS control universe derived from it.

1. Births and deaths can be missed entirely by the registration system. In the accounting of the national population without adjustment for underenumeration, underregistration of vital events is tacitly assumed to match underenumeration in the census. While we have found no formal investigation of the underregistration of deaths, we surmise that it is likely to be far less than the rate of underenumeration in the census because of the requisite role of local public agencies in

the processing of bodies for cremation or burial. A test of the completeness of birth registration conducted by the Census Bureau in 1964-1968, applied to the distribution of 1992 births by race and whether born in hospital, implied an underregistration of 0.7 percent, far less than most estimates of underenumeration in recent censuses, but probably greater than the under-registration of deaths.

2. Birth and death statistics obtained from NCHS exclude those events occurring to U.S. residents while outside the United States. The resulting slight downward bias in the numbers of births and deaths would be partially compensatory.

These two sources of bias, which are assumed to be minor, affect the accounting of total population. Other, more serious comparability problems exist in the consistency of NCHS vital statistics with the CPS control universe, with respect to race, Hispanic origin, and place of residence within the United States. These will be discussed in later sections of this appendix.

International migration (IM). The objective of the current procedures to estimate international migration (IM, in equation (7)), is to transform the number of legal immigrants—for whom records of migration are available—into an estimate of the net of persons who become "usual residents" of the United States according to the census residency definition, and those who cease to be usual residents because they move out of the United States. This objective is met primarily by the availability of administrative data from Federal agencies. However, it is beset by five fundamental problems.

1. A substantial number of foreign-born persons increase the resident population each year, either by overstaying legal, nonimmigrant visas or by entering the country without inspection. Those that have not returned to their country of origin (either voluntarily or by force of law) by a CPS date are eligible for interview by the CPS and must be included in the CPS controls.

2. Because the geographic limits of the census universe do not include Puerto Rico or outlying areas under U.S. jurisdiction, persons who enter or leave the country from or to these areas must be treated as international migrants. However, migrants to or from these areas are generally U.S. citizens and need not produce any administrative record of their moves.

3. Legal residents of the United States departing for residence abroad are not required to provide any administrative record of departure. Hence, there exist no current data on emigration.

4. Some persons who enter the country legally, but temporarily (e.g., foreign students, scholars, and members of certain professions), are eligible for enumeration in the census as usual residents and should be included in population estimates. While administrative records exist for the arrival of such persons, there is no adequate source of data for their departures or their lengths of stay.

5. The census definition of usual residence is subjective and depends on the interpretation of census respondents. Immigration data generally assume a legal, rather than a subjective concept of residence (e.g., legal permanent resident, nonresident alien) that does not match the census concept. Thus, an alien could be legally resident in the United States, but perceive their usual residence to be elsewhere when responding to a census or survey.

For purposes of population accounting, migration to the United States is assumed to include persons arriving in the United States who, based on their arrival status, appear to be new census-defined residents, meaning they would report their usual residence in a census as being inside the United States. The accounting should thus include all legal permanent residents, refugees, and undocumented migrants that do not return and are not deported. It should also include arrivals (such as foreign students and foreign scholars) that generally assume steady residence in the United States for the duration of their stay and would, therefore, be enumerated in a census. Tourists and business travelers from abroad are assumed not to be U.S. residents and are, therefore, excluded from the international migration tally.

The core data source for the estimation of international migration is the Immigration and Naturalization Service (INS) public use immigrant file, which is issued yearly for immigrants (year defined by the Federal fiscal year). This file contains records for citizens of foreign countries immigrating or establishing legal permanent residence in the United States and accounts for the majority of international migration to the United States. Data from the immigrant file are summarized by the INS in its statistical yearbooks (for example, U.S. Immigration and Naturalization Service, 1997). However, the file contains no information on emigration, undocumented immigration, or nonimmigrant moves into the country. Moreover, immigrants do not always change residence at time of immigration; they may in fact already reside in the United States. INS immigrant data are, therefore, partitioned into four categories based on INS class of admission and each category is treated differently.

The first category consists of all persons classed as "new arrivals," meaning their last legal entry into the U.S. coincided with their immigration or acquisition of legal permanent resident status. These persons are simply included in the migration component for the month and year of their immigration.

The second category consists of persons already resident in the United States who are adjusting their status from nonimmigrant (e.g., temporary resident) to immigrant

and who are not refugees. They include a large component of persons who are nonimmigrant spouses or children of U.S. citizens or legal permanent residents. We cannot account for these persons at their time of arrival through current data. We could account for these persons retrospectively at the earlier year of their arrival from INS records. However, this would result in a serious downward bias for the most recent years, since we would miss similar persons, currently entering the United States, who will immigrate in future years. For this reason, we accept the number of adjustees in this category as a proxy for the number of future adjustees physically entering the country in the current year. This assumption is robust over time provided the number of persons in this category remains stable from year to year.[3]

A third category consists of persons who entered the country as refugees and are currently adjusting their status to immigrant. The assumption of year-to-year stability of the flow, made for the previous category, would generally be inadequate for refugees because they tend to enter the country in "waves" depending on emigration policies or political upheavals in foreign countries. Hence, the time series of their entry would be poorly reflected by the time series of their conversion to immigrant status. The Office of Refugee Resettlement (ORR) maintains monthly data on the arrival of refugees (as well as some special-status entrants from Haiti and Cuba) by country of citizenship. Refugees adjusting to immigrant status are thus excluded from the accounting of immigrants but included as refugees at their time of arrival based on the ORR series.

The fourth and final category consists of persons living in the country illegally—either as nonimmigrant visa over-stayers or "entered without inspection"—who adjust to legal immigrant status by proving a history of continuous residence in the United States. Since fiscal year 1989, the largest portion of these persons has been admitted under the Immigration Reform and Control Act of 1986 (IRCA), which required proof of permanent residence since 1982. Generally, these persons should qualify as census-defined residents from the date of their arrival, since they do not generally expect to return to their country of origin. Because their arrival occurred before the 1990 census and current undocumented migration is separately accounted, immigrating IRCA adjustees are excluded from the accounting of migration altogether for post-1990 estimates.

The separate accounting of net undocumented migration cannot be based on registration data because, by definition, it occurs without registration. Research conducted at the Census Bureau (Robinson, 1994) has produced an allowance of 225,000 net migration per year,

based on observation of the late 1980s, with the universe confined (appropriately) to those counted in the 1990 census. This number is a consensus estimate, based partly on a specific derivation and partly on a review of other estimates for the same period. The derivation is based on a residual method, according to which the number of foreign-born persons enumerated in 1990 by period of arrival is compared to net legal migration measured for the same period. Currently, all Census Bureau estimates and projections of the population after 1990 incorporate this annual allowance as a constant.

The remaining category of foreign-born migration to the United States, not included in the immigrant data, is the flow of legal temporary residents; persons who reside in the country long enough to consider themselves "usual residents" while they are here, but who do not have immigrant visas. These include foreign students, scholars, some business persons (those who establish residence), and some professionals who are provided a special allowance to work. INS data on nonimmigrants would provide records of admission for temporary visa holders; however, there would be no reliable data source for their departures. The stock of this category of foreign-born persons enumerated in the 1990 census was estimated at 488,000, based on various characteristics measured by the census "long form" (1 in 6 households). This migration flow is currently assumed to maintain this stock at 488,000, This tells us that net migration (arrivals minus departures) equals the estimated number of deaths occurring to the group while in the United States. The overall net migration of this group is rather trivial (less than a thousand per year); far more important is the implied distribution of migration by age, which will be discussed later in this appendix.

A major migratory movement across the U.S. frontier, not directly measurable with administrative data, is the emigration of legal permanent residents of the United States to abroad. The current projection is a constant 222,000 per year, of whom 195,000 are foreign-born (Ahmed and Robinson, 1994). Like the allowance for undocumented immigration, the method of derivation is a residual method. In this case, the comparison is between foreign-born persons enumerated in the 1980 census and foreign-born persons enumerated in the 1990 census who gave a year of arrival prior to 1980. In theory, the latter number would be smaller, with the difference attributed to either death or emigration. The number of deaths was estimated through life tables, leaving emigration as a residual. This analysis was carried out by country of birth, facilitating detailed analysis of possible biases arising from differential reporting in the two censuses. The annual estimate of 195,000 was projected forward as a constant, for purposes of producing post-1990 estimates and projections.

The remaining annual emigrant allowance of 27,000 is based on an estimate of native-born emigrants carried out during the early 1980s which employed data on U.S.-born

---

[3]In fiscal year 1995, the assumption of a stable flow of nonimmigrants adjusting to immigrant status was seriously undermined by a legal change. For the first time, persons qualifying for legal immigrant status were allowed to do so without leaving the country. As a result, a rapid increase in applications to INS resulted in an increase in the backlog of unprocessed applications. It was, therefore, necessary to adapt the accounting for the increase in applications, rather than assuming a one-for-one relationship of current adjustees to future adjustees currently entering the country.

persons enumerated in foreign censuses, as well as historical evidence from the period before 1968, when permanent departures from the United States were registered by INS.

A final class of international migration that can only be estimated roughly as a net flow is the movement of persons from Puerto Rico and the outlying areas to the United States. No definitive administrative data exist for measuring volume, direction, or balance of these migration flows. The migratory balance from all areas except Puerto Rico is assumed to be zero. For Puerto Rico, we assume an allowance of roughly 7,000 net migration (arrivals minus departures) per year. This projection is based on an imputation of net out-migration from Puerto Rico during the 1980s, based on a comparison of the island's natural increase (births minus deaths) with the population change between the 1980 and 1990 censuses.

There is considerable variation in the currency of data on international migration, depending on the source. The INS public use immigrant file is generally available during the summer following the fiscal year of its currency, so projections for control of the CPS have access to final immigration data through September, 2 years prior to the reference date of the projections (e.g., 1995 for 1997 controls). INS provides a provisional monthly series for the period from October through June for the following year, that is, through the middle of the calendar year before the CPS reference date. From that point until the CPS reference date itself, the immigrant series is projected. The Office of Refugee Resettlement data on refugees follows roughly the same timetable, except the preliminary series is ongoing and is generally current through 4 to 5 months before the CPS reference date. Net undocumented immigration, emigration of legal residents, and net migration from Puerto Rico are projected as constants from the last decennial census, as no current data are available. The determination of the base series for the projection requires considerable research, so the introduction of data for a new decade generally does not occur until about the middle of the following decade.

Net migration of Federally affiliated civilian U.S. citizens (CCM). While the approach to estimating international migration is to measure the migration flows for various types of migration, no data exist on flows for civilian U.S. citizens between the United States and abroad (except with respect to Puerto Rico, which is treated here as international); this is the variable CCM in equation (7). Because U.S. citizens are generally not required to report changes of address to or from other countries, there is no registration system from which to obtain data. Fortunately, there are data sources for measuring the current number (or stock) of some classes of civilian U.S. citizens residing overseas, including persons affiliated with the Federal government, and civilian dependents of Department of Defense employees, both military and civilian. The estimation procedure used for the movement of U.S. citizens thus

relies on an analysis of the change in the stock of overseas U.S. citizens over time. We assume the net migration of non-Federally affiliated U.S. citizens to be zero.

Imputation of migration from stock data on overseas population is by the following formula:

$$CCM = -\ (OCP_{t1} - OCP_{t0} - \ OB + OD) \qquad (9)$$

where OCP_{t1} and OCP_{t0} represent the overseas civilian citizen population (the portion that can be estimated) at the beginning of two consecutive months; OB is the number of births, and OD the number of deaths of this population during the month. This equation derives from the balancing equation of population change and states the net migration to the overseas civilian population to be its total change minus its natural increase.

Federally affiliated civilian U.S. citizens residing overseas for whom we have data, or the basis for an estimate, are composed of three categories of persons; dependents of Department of Defense (DOD) personnel, civilian Federal employees, and dependents of non-DOD civilian Federal employees. The number of civilian dependents of DOD military and civilian personnel overseas are published quarterly by DOD; the number of Federal employees—total and DOD civilian—located overseas at the beginning of each month is published by the Office of Personnel Management (OPM). We estimate the number of dependents of non-DOD civilian Federal employees by assuming the ratio of dependents to employees for non-DOD personnel to match the ratio for civilian DOD personnel. Quarterly data (e.g., DOD dependents) are converted to monthly by linear interpolation.

Estimates of the numbers of births and deaths of this population (OB and OD, in equation (9)) depend almost entirely on a single statistic, published quarterly by DOD: the number of births occurring in United States military hospitals overseas. This number (apportioned to months by linear interpolation) is adopted as an estimate of overseas births, on the reasoning that most births to Federally affiliated citizens or their dependents would occur in military hospitals. We estimate deaths by assuming deaths of persons 1 year of age or older to be nil; we estimate infant deaths by applying a life table mortality rate to the number of births. The likely underestimate of both births and deaths would tend to compensate each other. While this method of estimating the natural increase of overseas civilian citizens is very approximate, the numbers involved are very small, so the error is unlikely to have a serious effect on estimates of civilian citizen migration.

Data from DOD and OPM used to estimate civilian citizen migration are generally available for reference dates until 6 to 9 months prior to the CPS reference date. For the months not covered by data, the estimates rely on projected levels of the components of the overseas population and overseas births.

Net recruits to the Armed Forces from the civilian population (NRAF). The net recruits to the worldwide U.S. Armed Forces from the U.S. resident civilian population is given by the expression

$$(WWAF_{t1} - WWAF_{t0} + DRAF + DAFO - NRO)$$

in equation (4). The first two terms represent the change in the number of Armed Forces personnel worldwide. The third and fourth represent deaths of the Armed Forces in the U.S. and overseas, respectively. The fifth term represents net recruits to the Armed Forces from the overseas civilian population. While this procedure is indirect, it allows us to rely on data sources that are consistent with our estimates of the Armed Forces population on the base date.

Most of the information required to estimate the components of this expression is supplied directly by the Department of Defense, generally through a date 1 month prior to the CPS control reference date; the last month is projected, for various detailed subcomponents of military strength. The military personnel strength of the worldwide Armed Forces, by branch of service, is supplied by personnel offices in the Army, Navy, Marines and Air Force, and the Defense Manpower Data Center supplies total strength figures for the Coast Guard. Participants in various reserve forces training programs (all reserve forces on 3- and 6-month active duty for training, National Guard reserve forces on active duty for 4 months or more, and students at military academies) are treated as active-duty military for purpose of estimating this component, and all other applications related to the CPS control universe, although the Department of Defense would not consider them to be in active duty.

The last three components of net recruits to the Armed Forces from the U.S. civilian population, deaths of the resident Armed Forces (DRAF), deaths of the Armed Forces overseas (DAFO) and net recruits to the Armed Forces from overseas (NRO) are usually very small and require indirect inference. Four of the five branches of the Armed Forces (those in DOD) supply monthly statistics on the number of deaths within each service. Normally, deaths are apportioned to the domestic and overseas component of the Armed Forces relative to the number of the domestic and overseas military personnel. If a major fatal incident occurs for which an account of the number of deaths is available, these are assigned to domestic or overseas, as appropriate, before application of the pro rata assignment. Lastly, the number of net recruits to the Armed Forces from the overseas population (NRO) is computed annually, for years from July 1 to July 1, as the difference between successive numbers of Armed Forces personnel giving a "home of record" outside the 50 states and District of Columbia. To complete the monthly series, the ratio of persons with home of record outside the U.S. to the worldwide military is interpolated linearly, or extrapolated as a constant from the last July 1 to the CPS reference date. These monthly ratios are applied to monthly worldwide Armed Forces strengths; successive differences of the resulting estimates of persons with home of record outside the U.S. yield the series for net recruits to the Armed Forces from overseas.

Change in the institutional population ($CIP_{t1} - CIP_{t0}$). The change in the civilian population residing in institutions is measured by a report of selected group quarters facilities carried out by the Census Bureau, in conjunction with state governments, through the Federal State Cooperative Program for Population Estimates (FSCPE). Information is collected on the type of facility and the number of inhabitants, and these data are tabulated by institutional and noninstitutional. The change in the institutional population from the census date to the reference date of the population estimates, measured by the group quarters report, is used to update the institutional population enumerated by the census. Certain facilities, such as military stockades and resident hospitals on military bases, are excluded from the institutional segment, as most of their inhabitants would be military personnel; hence, the resulting institutional population is assumed to be civilian.

The last available group quarters data refer to July 1, 2 years prior to the reference year of the CPS series (July 1, 1995, for 1997 CPS controls). Institutional population from this date forward must be projected. Census-based institutional participation rates are computed for the civilian population by type of institution, age, sex, race, and Hispanic origin. These are used to produce estimates for the group quarters report dates, which are then proportionally adjusted to sum to the estimated totals from the report. Participation rates are recomputed, and the resulting rates, when applied to the monthly estimate series for the civilian population by characteristic, produce projections of the civilian institutional population. These estimates produce the last two terms on the right side of equation (7).

Population by Age, Sex, Race, and Hispanic Origin

The CPS second-stage weighting process requires the independent population controls to be disaggregated by age, sex, race, and Hispanic origin. The weighting process, as currently constituted, requires cross-categories of age group by sex and by race, and age group by sex and by Hispanic origin, with the number of age groups varying by race and Hispanic origin. Three categories of race (White, Black, and all Other[4]), and two classes of ethnic origin (Hispanic, non-Hispanic) are required, with no cross-classifications of race and Hispanic origin. Beginning in 1993, the population projection program adopted the full cross-classification of age by sex and by race and by Hispanic origin into its

_____
[4]Other includes American Indian, Eskimo, Aleut, Asian, and Pacific Islander.

monthly series, with 101 single-year age categories (single years from 0 to 99, and 100 and over), two sex categories (male, female), four race categories (White; Black; American Indian, Eskimo, and Aleut; and Asian and Pacific Islander), and two Hispanic origin categories (not Hispanic, Hispanic).[5] The resulting matrix has 1,616 cells (101 x 2 x 4 x 2), which are then aggregated to the distributions used as controls for the CPS.

In discussing the distribution of the projected population by characteristic, we will stipulate the existence of a base population and components of change, each having a distribution by age, sex, race, and Hispanic origin. In the case of the components of change, "age" is understood to be age at last birthday as of the estimate date. The present section on the method of updating the base population with components of change is followed by a section on how the base population and component distributions are measured.

Update of the Population by Sex, Race, and Hispanic Origin

The full cross-classification of all variables except age amounts to 16 cells, defined by variables not naturally changing over time (2 values of sex by 4 of race by 2 of Hispanic origin). The logic for projecting the population of each of these cells follows the same logic as the projection of the total population and involves the same balancing equations. The procedure, therefore, requires a distribution by sex, race, and Hispanic origin to be available for each component on the right side of equation (7). Similarly, the derivation of the civilian noninstitutional base population from the census resident population follows equation (8).

Update of the Age Distribution: The Inflation-Deflation Method

Having produced population series for 16 cells of sex, race, and Hispanic origin, it remains to distribute each cell to 101 age groups. Age differs fundamentally from other demographic characteristics because it changes over time. The balancing equations described in earlier sections rely on the premise that the defining characteristics of the population being estimated remain fixed, hence the procedure must be adapted to allow for the fact that age does not meet this criterion. We present the method of estimating the age distribution through the equation for the resident population, analogous to equation (1) for the total population, for reasons discussed at the end of this section. The mathematical logic producing the civilian noninstitutional population (CNP) by age from the resident population, the

active-duty resident Armed Forces, and the civilian institutional population is precisely the same for any age group as for the total population, so we omit its discussion. Of course, it is necessary to know the age distribution of the resident Armed Forces population and the civilian institutionalized population in order to derive the CNP by age from the resident population by age.

Under ideal circumstances, a very standard demographic procedure—the method of cohort components—would be employed to update the population age distribution. The cohort component method yields the population of a given birth cohort (persons born in a given year) from the same cohort in an earlier year, incremented or decremented by the appropriate components of change. To take a simple case, we assume the time interval to be 1 year, from the beginning of year t to the beginning of year t+1, and the age distribution being updated is a distribution of the resident population by single year of age. We can derive the number of persons age x (for most values of x) by the equation

$$R_{x,t+1} = R_{x-1,t} - D_{x,t} + NM_{x,t} \qquad (10)$$

where:

$R_{x,t+1}$ = resident population aged x (last birthday), at time t+1
$R_{x-1,t}$ = resident population aged x-1, at time t
$D_{x,t}$ = deaths during the time interval from t to t+1, of persons who would have been age x at time t+1
$NM_{x,t}$ = net migration during the time interval from t to t+1, of persons who would be age x at time t+1

For the special case where x=0, the equation (10) becomes

$$R_{0,t+1} = B_t - D_{0,t} + NM_{0,t} \qquad (11)$$

with variable definitions analogous to equation (10), except that B_t equals live births during the time interval from t to t+1. For the special case where the population estimated comprises an "open" age category, like 100 and over, the equation becomes

$$R_{X,t+1} = R_{x-1,t} + R_{X,t} - D_{X,t} + NM_{X,t} \qquad (10)$$

where $R_{X,t+1}$ and $R_{X,t}$ designate persons age x and older, at time t+1 and time t, respectively, and $D_{X,t}$ and $NM_{X,t}$ designate deaths and net migration, respectively, from t to t+1 of persons who would be age x and older at time t+1. Here, $R_{x-1,t}$ is defined as in equation (10).

A fundamental property of this method is its tracking of a birth cohort, or persons born in a given year, to a later date through aging; hence it optimizes the comparability of the population of birth cohorts over time. But, in producing a time series of projections or estimates of a population age distribution, the objective is rather to maintain consistency from month-to-month in the size of age groups, or the number of persons in a specific, unchanging age range. As

---

[5]Throughout this appendix, "American Indian" refers to the aggregate of American Indian, Eskimo, and Aleut; "API" refers to Asian and Pacific Islander.

a result of this, the cohort component method of updating age distributions is highly sensitive to discontinuities from age group to age group in the base population. Unfortunately, two attributes of the census base population render the assumption of continuity across age groups untenable. First, underenumeration in the census is highly age-specific. This problem is most egregious for age categories of Black males in the age range from 25 to 40, which represent a "trough" in census coverage. It is lowest near the middle of the age range, as is evident from the age pattern of census-based sex ratios within the range. Consequently, consecutive years of age are inconsistent with respect to the proportion of persons missed by the census. The second troublesome attribute of the census base population is the presence of distinct patterns of age reporting preference, as manifested by a general tendency to underreport birth years ending in one and overreport birth years ending in zero. Because the census date falls in the first half of a decade year, this results in an exaggeration of the number of persons with age ending in nine, and an understatement of the number with age ending in eight. Were the cohort component method applied to this distribution without adaptation, these perturbations in the age distribution would progress up the age distribution with the passage of time, impairing the consistency of age groups in the time series.

The solution to this problem, first adopted in the 1970s, is an adaptation of the cohort-component method that seeks to prevent the aging of spurious elements of the base population age structure. The resulting method is known as "inflation-deflation." In principle, the method depends on the existence of an alternative base population distribution by age, sex, race, and origin, for the census date, which is free of age-specific underenumeration and inconsistent age reports, but as consistent as possible with the census in every other aspect of the population universe. Under ideal circumstances, the alternative base distribution would be the population that would have been enumerated by the census, if the census had been free of undercount and age misreporting, although this requirement is neither realistic nor essential. It is essential, however, that single-year-of-age categories be consistent with one another, in the sense that they share the same biases. The correctness of the overall level of the alternative base population (absence of biases in the total population of all ages) is unimportant. Once such a base population distribution is available, it can be updated, recursively, from the census date to the estimates reference date using the cohort component logic of equations (10), (11), and (12) without concern for bias arising from discontinuities in the population by age.

To update a census-level population from one date to another, "inflation-deflation factors" are computed as the ratio of each single-year age group in the census base population to the same age group in the alternative base population. If the alternative population represents a higher degree of coverage than the census base population, the factors can be expected to have values somewhat less than one. If the alternative population is free of age misreporting, then overreported ages in the census may have factors greater than one, while underreported ages will have factors less than one. If, as in the present application, the calculations are done within categories of sex, race, and Hispanic origin, the factors can take on unnatural values if there is relative bias between the census and the alternative population in the reporting of these variables. Such is the case with the current application for "all other races," the racial category combining American Indian, Eskimo, Aleut, Asian, and Pacific Islanders. This group tends to be more heavily reported in the census than in other administrative data sources.

The census-level population at the base date is "inflated," through multiplication by the reciprocal of the factors, to be consistent with the alternative population base. The resulting distribution is updated via cohort component logic (equations (10) through (12)), and the factors are multiplied by the same age groups (not the same birth cohorts) on the reference date, thereby deflating the estimates back to "census level." Because the inflation-deflation factors are held constant with respect to age rather than birth cohort, those aspects of the base age distribution that should not increase in age from year-to-year are embodied in the inflation-deflation factors, and those aspects that should indeed "age" over time are embodied in the alternative base distribution. A mathematical representation of this procedure is given by the following equation, analogous to equation (10) for the cohort component method

$$\acute{R}_{x,t+1} = \frac{R_{x,0}}{R_{x,0}^A}\left(\frac{R_{x-1,0}^A}{R_{x-1,0}} R_{x-1,t} - D_{x,t}^A + NM_{x,t}^A + NM_{x,t}^A\right) \quad (13)$$

where $\acute{R}_{x,t+1}$ is the census-level estimate of population age x at time t+1, $R_{x-1,t}$ is the same for age x-1 at time t, $R_{x,0}^A$ is the resident population age x in the alternative distribution at the census date, $R_{x-1,0}^A$ is the same for age x-1, and $D_{x,t}^A$ and $NM_{x,t}^A$ are deaths and net migration for the interval beginning at time t, respectively, consistent with the alternative population, age x at time t+1. Note that $R_{x,0} / R_{x,0}^A$ is the inflation-deflation factor for age x. A similar adaptation can be applied to equations (11) and (12) to obtain expressions for estimates for age 0 and the open age category, respectively.

The inflation-deflation procedure does not preserve exact additivity of the age groups to the external sex-race-origin total R_t. It is therefore necessary, in order to preserve the balancing equation for the population totals, to introduce a final proportional adjustment. This can be expressed by the formula

$$R_{x,t} = \acute{R}_{x,t} \frac{R_t}{\sum_{y=0}^{100} \acute{R}_{y,t}} \quad (14)$$

which effectively ensures the additivity of the resident population estimates $R_{x,t}$ by age (x) to the external total R_t.

The alternative base population for the census date used in the Bureau's population estimates program, including the survey control projections, is known as the Demographic Analysis (DA) resident population. Originally devised to measure underenumeration in the census, this population is developed, up to age 65, from an historical series of births, adjusted for underregistration. This series is updated to a population on April 1, 1990, by sex and race using cumulative data on deaths and net migration from date of birth to the census date. For ages 65 and over, the population is based on medicare enrollees, as this population is considered more complete than the census enumeration with respect to coverage and age reporting. This population is assumed to possess internal consistency with respect to age superior to that of the census because births and deaths, which form the mainstay of the estimates of age categories, are based on administrative data series rather than subjective reporting of age from an underenumerated population. Some adaptation of the administrative series was necessary to standardize the registration universe for the time series of births and deaths, which changed, as various states were added to the central registration network. For a description of this method, see Fay, Passel and Robinson (1988).

One might reasonably ask why the results of the Post-Enumeration Survey (PES), used to adjust the CPS controls for underenumeration, were not adopted as the alternative population. The reason is that the PES age distribution was based on application of adjustment factors to the census population defined only for large age categories. This method does not address census-based inconsistencies among single years of age within the categories; moreover, it introduces serious new inconsistencies between ages close to the edges of the categories. Hence, its use as an alternative base for inflation-deflation would be inappropriate. This fact does not impair the appropriateness of PES-based adjustment as a means of addressing undercoverage in the completed survey controls, since the results may be superior in their treatment of other dimensions of the population distribution, such as state of residence, race, origin, or large aggregates of age.

Having noted the relative appropriateness of the DA base population for the inflated series, it is not without technical limitations.

1. Because of its dependency on historical vital statistics data, it can only be generated for three categories of race—White, Black, and all other. The distribution of the third category to American Indian and API assumes the distribution of the census base for each age-sex group. Therefore, spurious disturbances in the census age distributions that differ for these two groups remain uncorrected in the DA base population.

2. For the same reason, the two categories of Hispanic origin cannot be generated directly in the DA base. The

practical solution was to assume that the proportion Hispanic in each age, sex, and race category matched the 1990 census base population. Because of this assumption, no distinctly Hispanic properties of the age discontinuity in the census population could be reflected in the DA base employed in the estimates.

3. The reliance by the DA procedure on medicare enrollee data for the population 65 years of age and over, while defensible as a basis for adjusting the elderly population, leaves open the possibility of inconsistency in the age distribution around age 65.

4. The DA procedure assumes internally consistent registration of deaths and an accounting of net migration in the historical series, although adaptation had to be made for the changing number of states reporting registered births and deaths to the Federal government. Deviations from this assumption could affect the internal consistency of the age distribution, especially for the older ages under 65.

5. Because the DA population is based primarily on vital registration data, race reporting does not always match race reporting in the census. In particular, births and deaths reported to NCHS are less likely to be coded American Indian and Asian and Pacific Islander than respondents in the census. While this has no direct effect on estimated distributions of population by race (inflation-deflation only affects age within sex-race totals), the population defining the "ageable" attributes of the base age distribution for these groups is smaller (roughly 15 percent for most ages) than the base used to compute sex-race-origin-specific population totals.

6. The inflation-deflation method, as currently implemented, assumes current deaths and all the components of migration to match in the DA universe and the census base universe by sex, race, and origin. This amounts to saying that $D_{x,t}^A$ and $NM_{x,t}^A$ in equation (13) sum to D and NM in equation (1), respectively. In the case of births, we assume births adjusted for underregistration to be consistent with the DA population, and births without adjustment to be consistent with the census-level population. In principle, the application of components of change to a DA-consistent population in equation (13) roughly parallels the derivation of the DA population.[6] However, the components used to update the census-level population in equations (1) or (7)

---

[6]One deviation from this rule occurs in the treatment of births by race. The DA population defines race of child by race of father; the postcensal update currently assigns race of child by race of mother. This practice will likely be modified in the near future. A second deviation occurs in the accounting of undocumented immigration and the emigration of legal residents. As of 1995, the postcensal estimates had incorporated revised assumptions, which had not been reflected in the DA base population; moreover, the revised assumptions were conspicuously "census-level," rather than "DA-level" assumptions.

should theoretically exclude the birth, death, or migration of persons who would not be enumerated, if the census-level character of the estimates universe is to be maintained.

Finally, we note the DA population can be estimated only for the resident population universe, so the process of estimating age distributions must be carried out on the resident population. The accounting of total population, on the other hand, is for the civilian noninstitutional universe. The resident universe must then be adapted to the civilian noninstitutional population by subtracting resident Armed Forces and civilian institutional population estimates and projections by age, sex, and race, from the resident population, as of the estimate or projection reference date. Fortunately, our sources of information for the institutional and Armed Forces populations yield age distributions for any date for which a total population is available; hence, this adaptation is uncomplicated.

Distribution of Census Base and Components of Change by Age, Sex, Race, and Hispanic Origin

Our discussion until now has addressed the method of estimating population detail in the CPS control universe, and has assumed the existence of various data inputs. The current section is concerned with the origin of the inputs.

Modification of the census race and age distributions. The distribution of sex, race, and Hispanic origin in the census base population is determined by the census, but with some adaptation. The 1990 census included questions on sex, race, and Hispanic origin, as was the case in 1980. The race question elicited responses of White, Black, American Indian, Eskimo, Aleut, and ten Asian and Pacific Islander categories, including "other API." It also allowed respondents to write-in a race category for "other race." The Hispanic origin question elicited responses of "No (not Spanish/Hispanic)," and a series of "Yes" answers, including Mexican, Puerto Rican, Cuban, and other Spanish/Hispanic, with the last category offering a possibility of write-in. While the initial census edit process interpreted the write-in responses to the race question, it left a substantial number of responses not directly codable to any White, Black, American Indian, Eskimo, Aleut, or Asian and Pacific Islander group. In 1980 this "other or not specified" category comprised 6.7 million persons; in 1990, it comprised 9.8 million persons. The overwhelming majority, in both cases, consisted of persons who gave a positive response to the Hispanic origin question, and/or a Hispanic origin response to the race question. However, the Hispanic component of the "other or not specified" race category comprised somewhat less than half of the total Hispanic population (based on the Hispanic origin response), with most of the balance classified as White.

The existence of a residual race category with these characteristics was inconsistent with other data systems essential to the estimation process, namely vital statistics

and Armed Forces strength statistics. It also stood at variance with Office of Management and Budget Directive 15, a 1977 ruling intended to standardize race and ethnicity reporting across Federal data systems. This directive specified an exhaustive distribution of race that could either be cross-classified with Hispanic origin, into a 4-by-2 matrix, or combined into a single-dimensional variable with five categories, including American Indian, Eskimo, and Aleut (one category), Asian and Pacific Islander (one category), non-Hispanic White, non-Hispanic Black, and Hispanic origin. It was thus necessary to adapt the census distribution in such a way as to eliminate the residual race category. The Bureau opted to follow the 4-by-2 cross-classification in 1990. Prior to 1990, the American Indian and Asian and Pacific Islander categories were combined into a single category in all Census Bureau population estimates, and the race distribution included persons of Hispanic origin separately, with race unspecified. In 1990-based estimates, these two racial categories were separated, and the procedure was expanded to estimate the full race-Hispanic cross-classification, although at present this disaggregation is recombined for the purpose of weighting the CPS.

The modification of the 1990 race distribution occurred at the level of individual microdata records. Persons of unspecified race were assigned to a racial category using a pool of "race donors." This pool was derived from persons with a specified race response, and the identical response to the Hispanic origin question (e.g., Mexican, Puerto Rican, Cuban, other Hispanic, or not Hispanic). Thus, a respondent's race (if initially unspecified) was, in expectation, determined by the racial distribution of persons with the same Hispanic origin response who resided in the vicinity of the respondent.

The distribution by age from the 1990 census also required modification. The 1990 census form asked respondents to identify their ages and their years of birth. While the form provided explicit direction to include only those persons who were alive and in the household on the census date, there was no parallel instruction to state the ages of household members as age on April 1, 1990. It was apparent that many respondents reported age at time of completion of the form, time of interview by an enumerator, or time of their next anticipated birthday. Any of these could occur several months after the April 1 reference date. As a result, age was biased upward, a fact most noticeable through gross understatement of the under 1-year-of-age category. While this distribution was published in most 1990 census publications and releases, it was considered inadequate for postcensal estimates and survey controls. The resulting modification was based on reported year of birth. Age was respecified by year of birth, with allocation to first quarter (persons aged 1990 minus year of birth) and last three quarters (aged 1989 minus year of birth) based on a historical series of birth by month derived from birth registration data. This methodology is detailed in U.S. Bureau of the Census (1991).

As was the case with race, the recoding of age occurred directly on the individual records, after preliminary edits to eliminate reports of year of birth or age inconsistent with other variables (such as household relationship). The tabulations required for the base distribution for estimates could thus be obtained by simple frequency distributions from the individual-record file. There was also a trivial modification of the distribution by sex that came about because the assignment of a very small percentage of census respondents of unknown sex was linked to respondent's age. The modification of the age distribution will be discussed under "Distribution of Population Projections by Age." The resulting distribution is known to the Census Bureau estimates literature as "MARS," an acronym for "modified age, race, and sex."

We note parenthetically that the Demographic Analysis (DA) population, used as the alternative base for the inflation-deflation method, did not require modifying with respect to age and race because it was derived primarily from vital statistics and immigration. However, the historical distribution of international migrants by race was based on the MARS distribution by race within country of birth. More importantly, the need to expand the race detail of the original DA population required introducing the MARS distribution to the split between American Indians and API, as well as any indirect effect of MARS on the distribution by Hispanic origin within race.

Distribution of births and deaths by sex, race, and Hispanic origin. The principal source of information on sex and race for births and deaths is the coding of sex and race on birth and death certificates (race of mother, in the case of births). These results are coded by NCHS on detail files of individual birth and death records. Hispanic origin is also coded on most vital records, but a substantial number of events did not receive a code, in some cases because they occurred in states that do not code Hispanic origin on birth and death certificates. For births, the unknown category was small enough to be distributed to Hispanic and not Hispanic in proportion to the MARS distribution of persons aged under 1 year. For deaths, the number of unknowns was sufficiently large (in some years, exceeding the number of deaths known to be Hispanic) to discourage the use of the NCHS distribution. Life tables were applied to a projected Hispanic origin population by age and sex; the resulting distribution was aggregated to produce totals for Hispanic and not Hispanic by sex. While the resulting distributions of births and deaths resemble closely the MARS distribution of the base population, a few discrepancies remained. Some are of sufficient concern to prompt adaptations of the distributions.

1. NCHS final data through 1992 included, for both births and deaths, a small category of "other races," not coded to one of the categories in the MARS (OMB) distribution. The practice to date has been to include these in the Asian and Pacific Islander category,

although a closer examination of the characteristics of these events in 1994 suggested some should be coded elsewhere in the distribution, principally Hispanic origin, with race other than API.

2. The reporting of American Indian and Asian and Pacific Islander categories in the census, hence also in the MARS base population, has tended to exceed the reporting of these categories in birth and death data. This fact is symptomized by unrealistically low fertility and mortality rates when MARS-consistent population estimates are used as denominators. This fact has not been addressed in estimates and projections to date, but is currently under review for possible future adaptation.

3. Persons of Hispanic origin and race other than White are substantially more numerous in the MARS base population than in the vital statistics universe. This can best be explained by the procedure used to code race in the base population. The initial edit of write-in responses to the census race question was done independently of the Hispanic origin response. Consequently, a considerable number of write-in responses allocated to non-White racial categories were of Hispanic origin. While the subsequent allocation of the "other (unspecified) race" category in the MARS procedure did indeed reflect Hispanic origin responses, it was possible for "race donors" selected for this procedure to have been persons of previously assigned race. Because the census (MARS) base population defines the population universe for estimates and projections, the non-White race categories of the Hispanic origin population have been estimated using a race-independent assumption for Hispanic age-specific fertility and mortality rates, applied to MARS-consistent projected populations. The NCHS race totals for Hispanic and non-Hispanic combined have been considered sufficiently close to consistency with MARS to warrant their adoption for the population estimates; hence, the non-Hispanic component of each racial group is determined by subtraction of Hispanic from the total. Hence, this adaptation will have no direct effect on second stage controls to the CPS, as long as the control procedure does not depend on the cross-classification of race with Hispanic origin.

A further issue that has not been fully addressed is the possible effect on the race distribution of the practice of identifying the race of a child by the race of its mother. Research is currently in progress to define race and Hispanic origin of child by a method that recognizes the race and origin of both parents, using census results on the reported race of children of biracial parentage.

Deaths by age. The distribution of deaths by age presents yet another need for adaptation of the population universe. However, its use for census-based estimates requires the

application of projected rates to the oldest age groups, rather than a simple decrement of the population by deaths in an age group. Because a cohort-component method (or in the present case, inflation-deflation) is required to estimate the age distribution of the population, the procedure is extremely sensitive to the reporting of age on death certificates and in the Demographic Analysis base population among elderly persons. We recall that medicare enrollees, not vital statistics, form the basis for the DA age distribution of the elderly. Because deaths of the oldest age groups may represent a substantial portion of the living population at the beginning of an age interval, relatively small biases caused by differential age reporting can produce large cumulative biases in surviving populations over a few years' time. This, in fact, occurred for estimates produced during the 1980s. It was the practice then to decrement the population at all ages using data on numeric deaths. Because there were not enough deaths to match the extreme elderly population, the population in the extreme elderly groups grew rapidly in the estimates. This was most noted in the population over 100 years of age; while this is of no direct concern to the CPS, which does not disaggregate age above 85, the situation provoked legitimate concern that an overestimate of the population 85 and over was also occurring, albeit of lesser proportion. At the opposite extreme, the application of numeric deaths for the oldest age groups could produce negative populations, if the number of deaths in a cohort exceeded the living population at the beginning of the interval.

This problem was solved in 1992 by developing a schedule of age-specific death rates using life tables for racial and Hispanic origin categories. When life table death rates are applied to a population, the fact that proportions of the living population in each age group must die each year ensures the timely demise of the oldest birth cohorts. Of course, the life tables are themselves reliant on the assumed compatibility of the numerators and denominators of the death rates. However, the fact that the population estimates underlying the rates had incorporated earlier death rates was sufficient to limit any cumulative error from differential age reporting. The current practice is to incorporate numeric deaths from NCHS through 69, apply life table-based death rates to ages 70 and over by single year of age, then proportionately adjust deaths to ages 70 and over to sum to NCHS-based totals by sex and race. Thus, the only cumulative bias entering the series affecting the extreme elderly would arise from differential reporting of age by under or over 70 years, which is deemed to be of minor concern.

International migration by age, sex, race, and Hispanic origin. Legal international migration to the U.S. is based on actual records of immigration from INS or refugee arrivals from ORR, so the statistics include direct reporting of sex and country of birth or citizenship as well as age. Race and Hispanic origin are not reported. The procedure to obtain the race and Hispanic origin variables depends on a distribution from the 1990 census sample edited detail file (1 in 6 households or "long form" sample) of foreign-born persons arrived since the beginning of 1985, by sex, race (MARS categories) Hispanic origin, and country of birth. This distribution is used to distribute male and female immigrants and refugees by race and Hispanic origin. This procedure is very robust for most country-of-birth categories, since migrants from most foreign countries are heavily concentrated in a single race-origin category (the notable exception being Canada).

For the remaining components of international migration, indirect methods are required to produce the distribution by age, sex, race, and origin.

1. For annual emigration of legal residents (222,000 total), the foreign-born and native-born were distributed separately. For emigration of the foreign-born, the age-sex distribution is a byproduct of the determination of the annual allowance of 195,000 persons. Because emigration is computed as a residual of the foreign-born from two censuses (with adjustment for age-specific mortality), their age and sex distribution can rest on the censuses themselves. This logic also yields country of birth, so race and origin are imputed in the same manner as for legal immigrants and refugees, previously described. The 27,000 annual allowance of native-born emigrants is assigned a distribution by all characteristics matching the native-born population from the 1990 census.

2. For net undocumented migration, we rely primarily on a distribution supplied by INS, giving the distribution by age, sex, country of birth, and period of arrival, for persons who legalized their residence under provisions of the Immigration Reform and Control Act (IRCA). Age is back-dated from date of legalization to date of arrival, since undocumented migrants are assumed to be residents of the United States from the time they arrive. An allowance is made, based on a life table application, for persons who died before they could be legalized. Once again, country of birth was used to assign race and Hispanic origin based on data from the 1990 census on the race and Hispanic origin of foreign-born migrants by country of birth. This assumption is subject to two foreseeable biases; (1) it takes no account of undocumented migrants who never legalized their status, and who may have different characteristics than those who legalized; and (2) it takes no account of persons who entered and departed prior to legalizing, who would (if nothing else) tend to make the migration flow more youthful since such persons are older upon departure than upon entry.

3. The distribution by age and sex of the net migration flow from Puerto Rico arises from the method of imputation. The method of cohort survival was used to estimate the net migration out of Puerto Rico between the 1980 and 1990 census dates. This method yields

an age distribution as a byproduct. Race and Hispanic origin were not asked in the Puerto Rican census questionnaire and could not be considered in the census survival method. The distribution within each age-sex category is imputed from the one-way flow of persons from Puerto Rico to the United States from 1985 to 1990, based on the 1990 census of the 50 states and the District of Columbia.

4. As previously noted, we assume the overall net migration of legal temporary residents other than refugees to be constant and of a magnitude and distribution necessary to maintain a constant stock of legal temporary residents in the United States. The consideration of this component in the accounting system is of virtually no importance to population totals, but is quite important to the age distribution. Were it not considered, the effect would be to age temporary residents enumerated in the census through the distribution; whereas, in reality they are more likely replaced by other temporary residents of similar age. This stock distribution of 488,000 persons enumerated in the 1990 census is derived by identifying characteristics, measured by the census "long form" (20 percent sample) data, resembling those that qualify for various nonimmigrant visa categories. The largest such category is foreign students. We estimate an annual distribution of net migration by computing the difference of two distributions of population stock. The first is simply the distribution of the 488,000 persons from the 1990 census. The second is the distribution of the same population after the effects of becoming 1 year older, including losses to mortality, estimated from a life table. Thus, the migration distribution effectively negates the effects of cumulative aging of these persons in the cohort-component and inflation-deflation methods.

Migration of Armed Forces and civilian citizens by age, sex, race, and Hispanic origin. Estimation of demographic detail for the remaining components of change requires estimation of the detail of the civilian citizen population residing overseas, as well as the Armed Forces residing overseas and Armed Forces residing in the United States. The first two are necessary to assign demographic detail to the net migration of civilian citizens; the third is required to assign detail to the effects of Armed Forces recruitment on the civilian population.

Distributions of the Armed Forces by branch of service, age, sex, race/origin, and location inside or outside the United States are provided by the Department of Defense, Defense Manpower Data Center. The location-and-service-specific totals closely resemble those provided by the individual branches of the services for all services except the Navy. For the Navy, it is necessary to adapt the location distribution of persons residing on board ships to conform to census definitions, which is accomplished through a special tabulation (also provided by Defense Manpower

Data Center) of persons assigned to sea duty. These are prorated to overseas and U.S. residence, based on the distribution of the total population afloat by physical location, supplied by the Navy.

In order to incorporate the resulting Armed Forces distributions in estimates, the race-origin distribution must also be adapted. The Armed Forces "race-ethnic" categories supplied by the Defense Manpower Data Center treat race and Hispanic origin as a single variable, with the Hispanic component of Black and White included under "Hispanic," and the Hispanic components of American Indian and API assumed to be nonexistent. There also remains a residual "other race" category. The method of converting this distribution to consistency with MARS employs, for each age-sex category, the 1990 census MARS distribution of the total population (military and civilian) to supply all MARS information missing in the Armed Forces race-ethnic categories. Hispanic origin is thus prorated to White and Black according to the total MARS population of each age-sex category in 1990; American Indian and API are similarly prorated to Hispanic and non-Hispanic. As a final step, the small residual category is distributed as a simple prorata of the resulting Armed Forces distribution, for each age-sex group. Fortunately, this adaptation is very robust; it requires imputing the distributions of only very small race and origin categories.

The overseas population of Armed Forces dependents is distributed by age and sex, based on an overseas census conducted in connection with the 1970 census, of U.S. military dependents. While this source obviously makes no claim to currency, it reflects an age distribution uniquely weighted in favor of young adult females and children. Race and origin are based on the distribution of overseas Armed Forces, with cross-categories of age-sex with race-origin determined by the marginals. Detail for civilian Federal employees overseas has been provided by the Office of Personnel Management for decennial dates, and assumed constant; their dependents are assumed to have the same distribution as Armed Forces dependents.

Having determined these distributions for populations, the imputation of migration of civilian citizens previously described (equation (9)) can be carried out specifically for all characteristics. The only variation occurs in the case of age, where it is necessary to carry out the imputation by birth cohort, rather than by age group.

The two very small components of deaths of the Armed Forces overseas and net recruits to the Armed Forces from the overseas population, previously described, are also assigned the same level of demographic detail. Deaths of the overseas Armed Forces are assigned the same demographic characteristics as the overseas Armed Forces. Net recruits from overseas are assigned race and origin, based on an aggregate of overseas censuses, in which Puerto Rico dominates numerically. The age and sex distribution follows the worldwide Armed Forces.

The civilian institutional population by age, sex, race, and origin. The fundamental source for the distribution of the civilian institutional population by age, sex, race, and origin is a 1990 census/MARS tabulation of institutional population by age, sex, race, origin, and type of institution. The last variable has four values: nursing home, correctional facility, juvenile facility, and a small residual. From this table, participation rates are computed for the civilian population. These rates are applied in each year to the current estimated distribution of the civilian population. As previously observed, the institutional population total is updated annually, until July 1 of the year 2 years prior to the CPS reference date, by the results of a special report on selected group quarters facilities. The report provides an empirical update of the group quarters populations by type, including these four institutional types. The demographic detail can then be proportionately adjusted to sum to the type-specific totals. For dates beyond the last group quarters survey, the detailed participation rates are recomputed for the last available date, and these are simply applied to the estimated or projected civilian population distribution.

Population Controls for States

The second-stage weighting procedure for the CPS requires a distribution of the national civilian noninstitutional population ages 16 and over by state. This distribution is determined by a linear extrapolation through two population estimates for each state and the District of Columbia. The reference dates are July 1 of the years that are 2 years and 1 year prior to the reference date for the CPS population controls. The extrapolated state distribution is forced to sum to the national total population ages 16 and over by proportional adjustment. For example, the state distribution for June 1, 1996, was determined by extrapolation along a straight line determined by two data points (July 1, 1994, and July 1, 1995) for each state. The resulting distribution for June 1, 1996, was then proportionately adjusted to the national total for ages 16 and over, computed for the same date. This procedure does not allow for any difference among states in the seasonality of population change, since all seasonal variation is attributable to the proportional adjustment to the national population.

The procedure for state estimates (e.g., through July 1 of the year prior to the CPS control reference date), which form the basis for the extrapolation, differs from the national-level procedure in a number of ways. The primary reason for the differences is the importance of interstate migration to state estimates and the need to impute it by indirect means. Like the national procedure, the state-level procedure depends on fundamental demographic principles embodied in the balancing equation for population change; however, they are not applied to the total resident or civilian noninstitutional population but to the household population under 65 years of age. Estimates of the group quarters population under 65 and the population 65 and over employ a different logic, based (in both cases) on independent sources of data for population size, benchmarked to the results of the last census. Furthermore, while national-level population estimates are produced as of the first of each month, state estimates are produced at 1-year intervals, with July 1 reference dates.

The Population 65 Years of Age and Over

The base population by state (which can be either a census enumeration or an earlier estimate) is divided into three large categories; the population 65 years of age and over, the group quarters population under 65, and the household (nongroup quarters) population under 65. The first two categories are estimated directly by stock-based methodologies; that is, the population change is estimated as the change in independent estimates of population for the categories. In the case of the population ages 65 and over, the principal data source is a serial account of medicare enrollees, available by county of residence. Because there is a direct incentive for eligible persons to enroll in the medicare program, the coverage rate for this source is very high. To adapt this estimate to 1990 census-level, the change in the medicare population from census date to estimate date is added to the 1990 census enumeration of the population 65 and over.

The Group Quarters Population Under 65 Years of Age

The group quarters population is estimated annually from the same special report on selected group quarters facilities, conducted in cooperation with the Federal and State Cooperative Program for Population Estimates (FSCPE), used to update national institutional population estimates. Because this survey is conducted at the level of the actual group quarters facility, it provides detail at any level of geography. Its results are aggregated to seven group quarters types: the four institutional types previously discussed, and three noninstitutional types: college dormitories, military barracks, and a small noninstitutional group quarters residual. Each type is disaggregated to the population over and under 65 years of age, based on 1990 census distributions of age by type. The change in the aggregate group quarters population under 65 forms the basis for the update of the group quarters population from the census to the estimate date.

The Household Population Under 65 Years of Age

The procedure for estimating the nongroup quarters population under 65 can be analogized to the national-level procedures for the total population, with the following major differences.

1. Because the population is restricted to persons under 65 years of age, the number of persons advancing to age 65 must be subtracted, along with deaths of persons under 65. This component is composed of the 64-year-old population at the beginning of each year, and is measured by projecting census-based ratios of the 64-year-old to the 65-year-old population by the national-level change, along with proration of other components of change to age 64.

2. All national-level components of population change must be distributed to state-level geography, with separation of each state total into the number of events to persons under 65 and persons 65 and over. Births and deaths are available from NCHS by state of residence and age (for deaths). Similar data can be obtained—often on a more timely basis—from FSCPE state representatives. These data, after appropriate review, are adopted in place of NCHS data for those states. For legal immigration (including refugees), ZIP Code of intended residence is coded by INS on the immigrant public use file; this is converted to county and state geography by a program known as ZIPCRS, developed cooperatively by the Census Bureau and the U.S. Postal Service (Sater, 1994). Net undocumented migration is distributed by state on the basis of data from INS on the geographic distribution of undocumented residents legalizing their status under the Immigration Reform and Control Act (IRCA). The emigration of legal residents is distributed on the basis of the foreign-born population enumerated in the 1990 census.

3. The population updates include a component of change for net internal migration. This is obtained through the computation for each county of rates of out-migration based on the number of exemptions under 65 years of age claimed on year-to-year matched pairs of IRS tax returns. Because the IRS codes tax returns by ZIP Code, the previously mentioned ZIPCRS program is used to identify county and state of residence on the matched returns. The matching of returns allows the identification for any pair of states or counties of the number of "stayers" who remain within the state or county of origin and the number of "movers" who change address between the two. This identification is based on the number of exemptions on the returns and the existence or nonexistence of a change of address. Dividing the number of movers for each pair by the sum of movers and stayers for the state of origin yields a matrix of out-migration rates from each state to each of the remaining states. The validity of these interstate migration rates depends positively on the level of coverage (the proportion of the population included on two consecutive returns). It is negatively affected by the difference between tax filers and tax nonfilers with respect to migration behavior, since nonfilers are assumed to migrate at the same rate as filers.

The Exclusion of the Population Under 16 Years of Age, Armed Forces, and Inmates of Civilian Institutions

The next step in the state-level process is the estimation of the age-sex distribution, which allows exclusion of the population under 16. The major input to this procedure is a nationwide inquiry to state governments and some private and parochial school authorities for data on school enrollment. The number of school-aged children (exact age 6.5 to 14.5) for each annual estimate date is projected, without migration, from the census date by the method of cohort survival. This method begins with the number of persons enumerated in the last census destined to be of school age on the estimate date, known as the school-age cohort. For reference dates more than 6.5 years from the census, the enumerated cohort must be augmented with a count of registered births occurring between the census date and the estimate date minus 6.5 years. For example, if the estimate reference date is July 1, 1995, the enumerated cohort will consist of persons age 1.25 to 9.25, since the census of April 1, 1990, is 5.25 years prior to the estimate date. On the other hand, if the estimate reference date is July 1, 1998, the cohort will consist of the census population aged under 6.25 years, plus registered births occurring from April 1, 1990, through December 31, 1991. The number of deaths of this cohort of children from the census date to the estimate date (estimated from vital registration data) is then subtracted to yield a projection of the school-age population on the estimate reference date.

Subtracting the resulting projection from actual estimates of school-age population, determined by updating the census school-age population via school enrollments, yields estimates of cumulative net migration rates for school-age cohorts for each state. These estimates are used to estimate migration rates for each age group up to age 18, using historical observation of the relationship of the migration of each age group to the migration of the school-age population. Populations by age are proportionately adjusted to sum to the national total, yielding estimates of the population under age 16. These are subtracted from the total to yield the population ages 16 and over, required to produce the population of the CPS control universe.

Once the resident population 16 years of age and older has been estimated for each state, the population must be restricted to the civilian noninstitutional universe. Armed Forces residing in each state are excluded, based on reports of location of duty assignment from the branches of the Armed Forces (including the homeport and projected deployment status of ships, for the Navy). The group quarters report is used to exclude the civilian institutional population. The group quarters data and some of the Armed Forces data (especially information on deployment of naval vessels) must be projected from the last available date to the later annual estimate dates.

Dormitory Adjustment

The final step in the production of the CPS base series for states is the adjustment of the universe from the census universe to the CPS control universe with respect to the geographic location of college students living in dormitories. The decennial census form specifically directs students living in dormitories to report the address of their dormitories, while CPS interviews identify the address of the family home. A dormitory adjustment for each state is defined as the number of students with family home address in the state residing in dormitories (in any state) minus the number of students residing in dormitories in the state. The latter estimate is a product of the previously mentioned group quarters report, restricted to college dormitories. The former is the same group quarters dormitory estimate, but adjusted by a ratio of students by family resident state to students by college enrollment state, which is computed from data from the National Center for Education Statistics (NCES). Adding the dormitory adjustment (which may be negative or positive) to the census-level civilian noninstitutional population ages 16 and over yields the population ages 16 and over for a universe consistent with the CPS in every regard except adjustment for underenumeration.

ADJUSTMENT OF CPS CONTROLS FOR NET UNDERENUMERATION IN THE 1990 CENSUS

Beginning in 1995, the CPS controls were adjusted for net undercoverage in the 1990 census. This adjustment was based on the results of the Post-Enumeration Survey (PES) carried out in the months following the census. From this, census coverage ratios were obtained for various large "poststrata" or cross-categories of a few variables determined to be most related to underenumeration. The present section addresses briefly the methodology of this survey, the application of coverage ratios for poststrata to obtain numerical levels of underenumeration in the CPS universe, and the application of the resulting numerical adjustment to the CPS population controls.

Important to the rationale behind the adjustment of CPS controls is the origin of the decision to adjust them for underenumeration, given that population estimates published by the Census Bureau are not. The Post-Enumeration Survey was, in its design, conceived as a method of providing an adjustment for net underenumeration in the 1990 census. In the year following the 1990 census date, estimates of net underenumeration were produced. In a decision effective July 15, 1991, Secretary Robert Mosbacher announced a decision not to adjust the census for undercount or overcount (Federal Register, 1991). Cited in the decision was a large amount of research pointing to improvement in the estimates of the national population resulting from incorporation of the Post-Enumeration Survey, coupled with a lessening of the accuracy of estimates for some states and metropolitan areas. The decision also called for further research into the possibility of incorporating PES results in the population estimates programs. Research conducted over the following 18 months under the auspices of the Committee on Adjustment of Postcensal Estimates (CAPE), a committee appointed by the Director of the Census Bureau, provided a description of the PES methodology, and a detailed analysis of the effect of adjustment on national, state, and county estimates. This research is summarized in a report, which forms the basis for all technical information regarding the PES discussed here (U.S. Census Bureau, 1992). After extensive public hearings, including testimony by members of Congress, Barbara Bryant, then Director of the Census Bureau, decided, effective December 30, 1992, not to incorporate PES results in the population estimates programs, but offered to calibrate Federally sponsored surveys, conducted by the Census Bureau, to adjusted population data (Federal Register, 1993). The decision cited the finding by CAPE that estimates of large geographic aggregates were improved by inclusion of PES estimates, while the same finding could not be determined for some states and substate areas. In the course of 1993, the Bureau of Labor Statistics opted to include the 1990 census, with adjustment based on the Post-Enumeration Survey, in the controls for the CPS, beginning January 1994.

We stress that the intent of the PES from its inception was to serve as a device for adjusting the census, not for the calibration of surveys or population estimates. Its application to population controls for the CPS and other surveys was based on an analysis of its results. Essential to its usefulness is the primacy of national-level detail in the evaluation of survey controls, coupled with the apparent superior performance of the PES for large geographic aggregates.

THE POST-ENUMERATION SURVEY AND DUAL-SYSTEM ESTIMATION

The PES consisted of a reinterview, after the 1990 census, of all housing units within each of a sample of small geographic units. The geographic unit chosen was the census block, a small polygon of land surrounded by visible features, frequently four-sided city blocks bounded by streets. The sampling universe of all such blocks was divided into 101 strata, based on certain variables seen to be related to the propensity to underenumeration. These consisted of geography, city size, racial and ethnic composition, and tenure of housing units (owned versus rented). The strata were sampled; the entire sample consisted of more than 5,000 blocks.

The reinterview and matching procedures consisted of an interview with all households within each of the sampled blocks followed by a comparison of the resulting PES observations with original census enumeration records of individual persons. Care was taken to ensure the independence of the reinterview process from the original enumeration, through use of different enumerators, as well as

independent local administration of the activity. Through the comparison of census and PES records of individual persons, matches were identified, as were nonmatches (persons enumerated in one survey but not in the other), and erroneous enumerations (persons who either should not have been enumerated, or should not have been enumerated in that block). Where necessary, follow-up interviews were conducted of census and PES households to determine the status of each individual with respect to his/her match between census and PES or his/her possible erroneous enumeration.

This matching procedure yielded a determination for each block of the number of persons correctly enumerated by the census only, the number of persons enumerated by the PES only, and the number enumerated by both. The number of persons missed by both census and survey could then be imputed by observing the relationship—within the PES sample—of the number missed by the census to the number enumerated, and generalizing this relationship to those missed by the PES sample. The resulting estimate was called a "dual-system estimate," because of its reliance on two independent enumeration systems. In most cases, the dual-system estimate was higher than the original enumeration, implying a net positive undercount in the census. For some blocks, the dual-system estimate was lower than the census count, meaning the number of incorrect enumerations exceeded the number of persons imputed by the matching process, resulting in a net overcount (negative undercount).

Having established dual-system estimates of the sampled blocks, a new stratification—-this time of individual persons (rather than blocks)—was defined; the resulting strata were dubbed "poststrata," as they were used for retrospective analysis of the survey results (as opposed to being part of the sampling design). These poststrata formed the basic unit for which census coverage ratios were computed. They were defined on data for characteristics considered relevant to the likelihood of not being enumerated in the census, including large age group, sex, race (including residence on a reservation, if American Indian), Hispanic origin, housing tenure (living in an owned versus rented dwelling), large city, urban, or rural residence, and geographic region of residence. The choice of categories of these variables was intended to balance two conflicting objectives. The first was to define strata that were homogeneous with respect to the likelihood of enumeration in the census. The second was to ensure that each stratum was large enough to avoid spuriousness in the estimation of census coverage. As a result, the categories were asymmetric with respect to most of the variables: the choice of categories of one variable depended on the value of another. Seven categories of age and sex (four age categories by two sex categories, but with males and females combined for the youngest category) were crossed with 51 categories of the other variables to produce a total of 357 poststrata.

The percentages of undercount or overcount were computed for each poststratum by dividing the difference between the dual-system estimate of population and the census enumeration for the sampled population by the census enumeration. Applying these ratios to the poststrata, this time defined for the entire noninstitutional population of the United States enumerated in the 1990 census, yielded estimates of undercoverage. Among variables tabulated in the CPS, there was considerable variation in the adjustment factors by race, since race was an important source of heterogeneity among poststrata. Moreover, as a result of the small number of age categories distinguished in the poststrata, adjustment factors tended to vary sharply between single year age groups close to the limits of neighboring age categories.

Application of Coverage Ratios to the MARS Population

Undercount ratios were applied to the 1990 MARS distribution by age, sex, race, and origin. Because the poststrata included, in their definition, one racial category ("Other") that did not exist in the MARS distribution, undercount ratios for census race-origin categories were imputed to the corresponding MARS categories. These undercount ratios were implemented in the MARS microdata file, which was then tabulated to produce resident population by age, sex, race, Hispanic origin, and state. "Difference matrices" were obtained by subtracting the unadjusted MARS from the adjusted MARS distribution. This was done for both the national age-sex-race-origin distribution, and the 51-state distribution of the civilian noninstitutional population ages 16 years and over.

An important assumption in the definition of the difference matrix was the equality of the adjustment for resident population and civilian noninstitutional population. This amounted to an assumption of zero adjustment difference for the active-duty military and civilian institutional populations. The assumption was natural in the case of the institutional population, because this population was expressly excluded from the PES. In the case of the Armed Forces, the population stock is estimated, for both the nation and for states, from data external to the census; hence, there was no basis on which to estimate PES-consistent undercount for the Armed Forces population.

Once derived, the differences have been added to all distributions of population used for independent controls for the CPS produced since January 1, 1995. While it would have been technically more satisfying to apply undercount ratios, as defined by the PES, directly to the CPS population controls, this would have required projections of the poststrata for CPS control reference dates, for which no methodology has been developed.

This adjustment of the CPS universe is the final step in the process of computing independent population controls for the second-stage weighting of the Current Population Survey.

A question that frequently arises regarding adjustment for underenumeration is how it relates to inflation-deflation, as previously discussed. We need to inflate estimated populations to an ''adjusted level'' in order to apply cohort-component logic to the estimation of the age distribution. Why then is it necessary to deflate to census level, then inflate again to adjust for underenumeration? The answer to this lies in the relationship of the purpose of the inflation-deflation method to the method of adjustment. In the case of inflation-deflation, the sole objective is to provide an age distribution that is unaffected by differential undercount and age reporting bias, from single year to single year of age throughout the distribution. This is necessary because cohort component logic advances the population 1 year of age for each 1-year time interval. The Demographic Analysis population meets this objective well, as it is developed, independently of the census, from series of vital and migratory events that are designed to be continuous, based on time series of administrative data. However, the definition of poststrata for dual-system estimates in the PES is based on the application of uniform adjustment factors to large age categories, with large differences at the limits of the categories. This would not serve the needs of inflation-deflation, even though the PES has been assumed to be a superior means of adjusting the census for underenumeration (at least for subnational geographic units).

THE MONTHLY AND ANNUAL REVISION PROCESS FOR INDEPENDENT POPULATION CONTROLS

Each month a projection of the civilian noninstitutional population of the United States is produced for the CPS control reference date. Each projection is derived from the population at some base date and the change in the population from the base date to the reference date of the projection. In every month except January, the base date (the date after which new data on population change can be incorporated in the series) is 2 months prior to the CPS reference date. In January, the entire monthly series back to the last census is revised, meaning the base date is the date of the last census (currently April 1, 1990). Early in the decade (in the current decade, January 1, 1994), the population from the most recent decennial census is introduced for the first time as a base for the January revision.

As a consequence of the policy of ongoing revision of a month back, for monthly intervals from January 1 to December 1, the monthly series of population figures produced each month for 1 month prior to the CPS reference date is internally consistent for the year of reference dates from December 1 to November 1—meaning the month-to-month change is determined by the measurement of population change. For actual CPS reference dates from January 1 to December 1, the series is not strictly consistent; there is a small amount of ''slippage'' in the month-to-month intervals, which is the difference between:

1. The population change during the 1 month immediately preceding the CPS reference date, as measured a month after the CPS control figure is produced, and

2. The population change in the month preceding the CPS date, associated with the actual production of the CPS controls.

This slippage is maintained for the purpose of preventing cumulative error, since the measurement of population change in the last month at the time of measurement (2) is based entirely on projection (no data are available to measure the change); whereas, some key administrative data (e.g., preliminary estimates of births and deaths) are generally available for the month before the last (1). The slippage rarely exceeds 20,000 persons for the population ages 16 and over, compared to a monthly change in this population that varies seasonally from roughly 150,000 to 260,000 persons, based on calendar year 1994. Because the numerically largest source of slippage is usually the projection of births, most of the slippage is usually confined to the population under 1 year of age, hence, of no concern to labor force applications of the survey.

Monthly population controls for states, by contrast, normally contribute no slippage in the series from January to December, because they are projections from July 1 of the previous year. Generally, any slippage in the state series is a consequence of their forced summation to the national total for persons ages 16 and over. The annual revision of the population estimates to restart the national series each January affects states as well as the Nation, as it involves not only revisions to the national-level estimates of population change, but also the distribution of these components to states, and an additional year of internal migration data.

For the revision of the entire series each January, preliminary estimates of the cumulative slippage for the national population by demographic characteristic (not the states) are produced in November and reviewed for possible effect on the population control series. Annual slippage (the spurious component of the difference between December 1 and January 1 in CPS production figures) is generally confined to less than 250,000 in a year, although larger discrepancies can occur if the revision to the series entails cumulative revision of an allowance for an unmeasured component (such as undocumented immigration), the introduction of a new census, or a redefinition of the universe (e.g., with respect to adjustment for undercoverage).

PROCEDURAL REVISIONS

The process of producing estimates and projections, like any research activity, is in a constant state of flux. Various technical issues are addressed with each cycle of estimates for possible implementation in the population estimating procedures, which carry over to the CPS controls. These procedural revisions tend to be concentrated early in the decade, because of the introduction of data from a

new decennial census, which can be associated with new policies regarding population universe or estimating methods. However, they may occur at any time, depending either on new information obtained regarding the components of population change, or the availability of resources to complete methodological research.

REFERENCES

Ahmed, B. and Robinson, J. G. (1994), **Estimates of Emigration of the Foreign-Born Population: 1980-1990**, U.S. Census Bureau, Population Division Technical Working Paper No. 9, December, 1994.

Anderson, R. N., Kochanek, K. D., Murphy, S. L. (1997), "Report of Final Mortality Statistics, 1995," **Monthly Vital Statistics Report,** National Center for Health Statistics, Vol. 45, No. 11, Supplement 2.

Fay, R. E., Passel, J. S., and Robinson, J. G. (1988), **The Coverage of the Population in the 1980 Census**, 1980 Census of Population and Housing, U.S. Census Bureau, Vol. PHC–E–4, Washington, DC: U.S. Government Printing Office.

Federal Register, Vol. 56, No. 140, Monday, July 22, 1991.

Federal Register, Vol, 58, No. 1, Monday, January 4, 1993.

Robinson, J. G. (1994), "Clarification and Documentation of Estimates of Emigration and Undocumented Immigration," unpublished U.S. Census Bureau memorandum.

Sater, D. K. (1994), **Geographic Coding of Administrative Records— Current Research in the ZIP/SECTOR-to-County Coding Process**, U.S. Census Bureau, Population Division Technical Working Paper No. 7, June, 1994.

U.S. Census Bureau (1991), **Age, Sex, Race, and Hispanic Origin Information From the 1990 Census: A Comparison of Census Results With Results Where Age and Race Have Been Modified,** 1990 Census of Population and Housing, Publication CPH-L-74.

U.S. Census Bureau (1992), **Assessment of Adjusted Versus Unadjusted 1990 Census Base for Use in Intercensal Estimates: Report of the Committee on Adjustment of Postcensal Estimates**, unpublished, August 7, 1992.

U.S. Immigration and Naturalization Service (1997), **Statistical Yearbook of the Immigration and Naturalization Service,** 1996, Washington, DC: U.S. Government Printing Office.

Ventura, S. J., Martin, J. A., Curtin, S. C., Mathews, T.J. (1997a), "Report of Final Natality Statistics, 1995," **Monthly Vital Statistics Report**, National Center for Health Statistics, Vol. 45, No. 11, supplement.

Ventura, S. J., Peters, K. D., Martin, J. A., and Maurer, J. D. (1997b), "Births and Deaths: United States, 1996," **Monthly Vital Statistics Report**, National Center for Health Statistics, Vol. 46, No. 1.

SUMMARY LIST OF SOURCES FOR CPS POPULATION CONTROLS

The following is a summary list of agencies providing data used to calculate independent population controls for the CPS. Under each agency are listed the major data inputs obtained.

1. The U.S. Census Bureau (Department of Commerce)

   The 1990 census population by age, sex, race, Hispanic origin, state of residence, and household or type of group quarters residence

   Sample data on distribution of the foreign-born population by sex, race, Hispanic origin, and country of birth

   The population of April 1, 1990, estimated by the Method of Demographic Analysis

   Decennial census data for Puerto Rico

2. National Center for Health Statistics (Department of Health and Human Services)

   Live births by age of mother, sex, race, and Hispanic origin

   Deaths by age, sex, race, and Hispanic origin

3. The U.S. Department of Defense

   Active-duty Armed Forces personnel by branch of service

   Personnel enrolled in various active-duty training programs

   Distribution of active-duty Armed Forces Personnel by age, sex, race, Hispanic origin, and duty location (by state and outside the United States)

   Deaths to active-duty Armed Forces personnel

   Dependents of Armed Forces personnel overseas

   Births occurring in overseas military hospitals

4. The Immigration and Naturalization Service (Department of Justice)

   Individual records of persons immigrating to the United States, month and year of immigration, including age, sex, country of birth, and ZIP Code of intended residence, and year of arrival (if different from year of immigration)

5. Office of Refugee Resettlement (Department of Health and Human Services)

   Refugee arrivals by month, age, sex, and country of citizenship

6. Department of State

   Supplementary information on refugee arrivals

7. Office of Personnel Management

   Civilian Federal employees overseas in overseas assignments

Appendix E.
State Model-Based Labor Force Estimation

INTRODUCTION

Small samples in each state and the District of Columbia result in unacceptably high variation in the monthly CPS composite estimates of state employment and unemployment. The table below gives the sample sizes, the standard errors, and coefficient of variation (CVs) for unemployment and employment assuming an unemployment rate of 6 percent for the states and the Nation as a whole. These numbers are based on the current design which was in effect January 1996 through July 2001.[1]

Table E–1. **Reliability of CPS Estimators Under the Current Design**

	Unemployment rate		Employment-to-population		Number of house-holds in sample
	CV	STD	CV	STD	
Nation	1.90	0.11	0.25	0.16	50,000
States	15.70	0.94	1.98	1.25	546-4055

In an effort to produce less variable labor force estimates, BLS introduced time series models to "borrow strength" over time. The models are based on a signal-plus-noise approach to small area estimation where the monthly CPS estimates are treated as stochastically varying labor force values obscured by sampling error (Tiller, 1992). Given a model for the true labor force values and sampling error variance-covariance information, signal-extraction techniques are used to estimate the true labor force values. This approach was first suggested by Scott and Smith (1974) and has been more fully developed by Bell and Hillmer (1990), Binder and Dick (1990) and Pfefferman (1992).

TIME SERIES COMPONENT MODELING

For each state, separate models for the CPS employment-to-population ratio and the unemployment rate are developed. In the signal-plus-noise model, the CPS labor force estimate at time t, denoted by $y(t)$, is represented as the sum of two independent processes

$$y(t) = \theta(t) + e(t) \qquad (1)$$

where $\theta(t)$ represents the systematic part of the true labor force value and $e(t)$ is the sampling error. The basic objective is to produce an estimator for $\theta(t)$ that is less variable than the CPS composite estimator. The signal component, $\theta(t)$, represents the variation in the sample values due to the stochastic behavior of the population. The sampling error component, $e(t)$, consists of variation arising from sampling only a portion of the population.

The Signal Component

In the state labor force models, the signal component is represented by an unobserved variance component model with explanatory variables (Harvey, 1989)

$$\theta(t) = \theta'(t) + \eta(t)$$

where

$$\theta'(t) = X(t)\beta(t) + T(t) + S(t) \qquad (2)$$

$X(t)$ = vector of known explanatory variables
$\beta(t)$ = random coefficient vector
$T(t)$ = trend component
$S(t)$ = periodic or seasonal component
$\eta(t)$ = residual noise.

The stochastic properties of each of these components are determined by one or more normally distributed, mutually independent white noise disturbances

$$\upsilon_j(t) \sim NID(0, \sigma_{\upsilon_j}^2) \qquad (3)$$

where j indexes the individual components.

Regression component. The regression component consists of explanatory variables, $X(t)$, with time-varying coefficients, $\beta(t)$

$$M(t) = X(t)\beta(t). \qquad (4)$$

This component accounts for variation in the CPS estimates that can be explained by a set of observable economic variables developed from auxiliary data sources that are independent of the CPS sampling error. A common core of state-specific explanatory variables has been developed: unemployment insurance claims from the Federal-State Unemployment Insurance System (Blaustein, 1979),

---

[1]They do not reflect the increased reliability after the sample expansion in July 2001 due to the State Children's Health Insurance Program. (See Appendix J.)

a nonagricultural payroll employment estimate from the BLS Current Employment Statistics (CES) program (US DOL, 1997), and intercensal population estimates developed by the Census Bureau. The explanatory variable included in the employment model is the CES estimate, adjusted for strikes, expressed as a percent of the state's population. For the unemployment rate model the explanatory variable is the claims rate defined as the ratio of the worker claims for unemployment insurance (UI) benefits to CES employment.

The regression coefficients are modeled as a random walk process

$$\beta(t) = \beta(t-1) + v_\beta(t), \; v_\beta(t) \sim NID(0,\sigma_\beta^2). \tag{5}$$

The effect of $v_\beta(t)$ is to allow the coefficients to evolve slowly over time. A zero variance, $\sigma_\beta^2 = 0$, results in a fixed regression coefficient.

Time series components. The explanatory variables in the models account for an important part of the variation in the CPS estimates but, because of definitional and other differences, leave significant amounts of the nonsampling error variation in the CPS estimates unexplained (Tiller, 1989). Therefore, stochastic trend and seasonal components are added to the model to adjust the explanatory variables for differences from the CPS measure of state employment and unemployment. Since the explanatory variables account for a portion of the trend and seasonal variation in the CPS, these stochastic time series components are residuals; they do not account for all the trend and seasonal variation in the CPS.

Residual trend component. The trend component is represented by a local approximation to a linear trend with a random level, T(t), and slope, R(t)

$$T(t) = T(t-1) + R(t-1) + v_T^*(t)$$

and

$$R(t) = R(t-1) + v_R(t)$$

where

$$v_T^*(t) = \sum_{k=1}^{m} \delta_k \xi_k(t) + v_T(t)$$

and

$$\xi_k(t) = \begin{cases} 1 \text{ if } t = t_k \\ 0 \text{ if } t \neq t_k. \end{cases} \tag{6}$$

External shocks which cause permanent shifts in the level of the trend are specified in the disturbance term associated with the trend level, $v_T(t)$. The coefficient, δ_k, represents the impact of the shock at time t_k.

The disturbance terms $v_T(t)$ and $v_R(t)$ are assumed to be mutually uncorrelated white noise random variates with zero means and constant variances

$$v_T(t) \sim NID(0,\sigma_T^2) \tag{7}$$

$$v_R(t) \sim NID(0,\sigma_R^2)$$

$$E\left[v_T'(t)v_T^*(t)\right] = 0.$$

The larger the variances, the greater the stochastic movements in the trend. The effect of $v_T(t)$ is to allow the level of the trend to shift up and down, while $v_R(t)$ allows the slope to change. This two-parameter trend model allows for a variety of patterns. If the variances of the two disturbances are both zero, then this component reduces to a fixed linear trend. A random walk results when the level variance is positive and the slope is identically zero and, when explanatory variables are included in the model, results in an intercept which varies over time.

Residual seasonal component. This component is the sum of up to six trigonometric terms associated with the 12-month frequency and its five harmonics

$$S(t) = \sum_{j=1}^{6} S_j(t). \tag{8}$$

Each frequency component is represented by a pair of stochastic variables

$$S_j(t) = \cos(\omega_j)S_j(t-1) + \sin(\omega_j)S_j^*(t-1) + v_{s_j}(t) \tag{9}$$

$$S_j^*(t) = -\sin(\omega_j)S_j(t-1) + \cos(\omega_j)S_j^*(t-1) + v_{sj}^*(t)$$

$$\omega_j = \frac{2\pi j}{12}$$

where v_{s_j} and $v_{s_j}^*$ are zero mean white noise processes which are uncorrelated with each other, and have a common variance σ_s^2. The white noise disturbances are assumed to have a common variance, so that the change in the seasonal pattern depends upon a single parameter. If the common variance is zero then the seasonal pattern is fixed over time. The expected values of the seasonal effects add to zero over a 12-month period.

Residual noise component. The residual noise component consists of random variation unaccounted for by other components plus unusually large transitory fluctuations or outliers

$$\eta(t) = I(t) + O(t). \tag{10}$$

Irregular component, I*(t).* In some cases after estimating the signal and sampling error components, we find that the signal still contains significant irregular movements, not attributable to sampling error, which tend to average to zero over a short period of time. In this case an additional irregular component is added to the model specification to

further smooth the estimates of the signal. This component is specified as consisting of a single white noise disturbance with a zero mean and constant variance

$$I(t) = \upsilon_I(t), \ \upsilon_I(t) \sim N(0, \sigma_I^2). \tag{11}$$

If the variance for this component is zero, then the irregular is identically zero and can be dropped from the model.

Additive outlier component, O(t). An additive outlier represents a one-period transitory shift in the level of the observed series

$$O(t) = \sum_j \lambda_j \zeta_j(t)$$

where

$$\zeta_j(t) = \begin{cases} 1 \text{ if } t = j \\ 0 \text{ otherwise.} \end{cases} \tag{12}$$

The coefficient λ_j is the change in the level of the series at time j. Time series are typically influenced by exogenous disturbances that affect specific observations. These outliers may occur because an unusually nonrepresentative sample of households was selected or because some real, nonrepeatable event occurred in the population. Irrespective of its origin, an outlier of this type affects only a single observation at a time and is unrelated to the time series model of the signal. Because an outlier represents a sudden, large, temporary change in the level of the CPS, the model may attempt to initially adjust to this break as if it were permanent. Accordingly, it is important to identify such outliers and then discount them when estimating the signal.

While the general model of the signal, just described, is very flexible, not all of the components discussed above are necessarily needed. The seasonal component may have less than 6 frequency components, depending upon the seasonal nature of the series being modeled. The regressor variables may be able to explain a substantial amount of variation in the observed series with fixed coefficients. On average, the CES explains 11 percent of the month-to-month variation in the CPS employment-to-population ratio (CPSEP) while the UI data account for about 15 percent of the monthly variation in the CPS unemployment rate (CPSRT). If good regressor variables are not available, the signal may be represented by just the time series component; this results in a univariate analysis.

Sampling error component. Sampling error is defined as the difference between the population value, $\theta(t)$, and the survey estimate

$$e(t) = y(t) - \theta(t) \tag{13}$$

where e(t) has the following properties:[2]

$$E[e(t)] = 0 \tag{14}$$

$$Var[e(t)] = \sigma_{e(t)}^2 \tag{15}$$

$$\rho_e(l) = \frac{E\{e(t)e(t-l)\}}{\sigma_{e(t)}^2} \tag{16}$$

The CPS estimates contain measurement error which is very difficult to quantify. We ignore this source of error by treating our target variable as the set of values the CPS would produce with a complete census of the population. The sampling error is, therefore, assumed to have a zero expectation. While the variances and, hence, the covariances change over time, the autocorrelations are treated as stationary (see below.)

When modeling the sampling error, it is important to account for sample design features which are likely to have a major effect on the error structure e(t). The autocorrelation structure of the CPS sampling error depends strongly upon the CPS rotating panel design and population characteristics (see Chapter 3). Another important characteristic of the state CPS estimator is its changing reliability over time. That is, the absolute size of the sampling error is not fixed but changes because of redesigns, sample size changes, and variation in labor force levels. Prior to 1985, some state estimates were produced from a national design. Special sample supplementations under the old national design also had an effect on the reliability of selected state samples in the late 1970s and early 1980s. A state-based design was phased in during 1984/85 along with improved procedures for noninterviews, ratio adjustments, and compositing. This redesign had a major effect on the reliability of many of the state CPS estimates. Even with a fixed design and sample size, the variance of the sampling error component changes because it is also a function of the size of the labor force characteristics of the population being measured. Since the CPS variance changes across time and the autocorrelation structure is stable over time, we express e(t) in multiplicative form as

$$e(t) = \gamma(t)e^*(t). \tag{17}$$

This expression allows us to capture the autocorrelated and heteroscedastic structure of e(t). The variance inflation factor, $\gamma(t)$, accounts for heteroscedasticity in the CPS and is defined by

$$\gamma(t) = \frac{\sigma_{e(t)}}{\sigma_{e^*}} \tag{18}$$

where $\sigma_{e(t)}$ is the GVF estimate of the standard error for the CPS estimate and σ_{e^*} is the standard error of the autoregressive moving average (ARMA) process, $e^*(t)$.

---

[2]In the context of time series analysis, we view the variances and covariances of the CPS estimators as properties of the sampling error component.

The CPS autocorrelation structure is captured through $e^*(t)$ which is modeled as an ARMA process

$$e^*(t) = \phi^{-1}(L)\theta(L)\upsilon_{e^*}(t) \qquad (19)$$

where L is the lag operator, that is, $L^k(X_t) = X_{t-k}$ so that

$$\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p$$

and

$$\theta(L) = 1 - \theta_1 L - \theta_2 L^2 - \ldots - \theta_q L^q. \qquad (20)$$

The parameters, $\phi_1, \phi_2, \ldots, \phi_p$, and $\theta_1, \theta_2, \ldots, \theta_q$ are estimated from the sampling error lag correlations (Dempster and Hwang, 1990) through the autocorrelation function

$$\rho_{e^*}(l) = \frac{\theta(L)\theta(L^{-1})}{\phi(L)\phi(L^{-1})} \qquad (21)$$

where we set $\rho_{e^*}(-l) = \rho_{e^*}(l)$. The coefficients θ and ϕ are then used to compute the impulse response weights $\{g_k\}$ through the generating function

$$g(L) = \phi^{-1}(L)\theta(L). \qquad (22)$$

The variance of the ARMA process $e^*(t)$ is computed as

$$\sigma_{e^*}^2 = \sum_{k=0}^{\infty} g_k^2. \qquad (23)$$

We can then compute the variance inflation $\gamma(t)$ factor using CPS standard error estimates which are obtained from generalized variance function (see Chapter 13) for $\sigma_{e(t)}$.

ESTIMATION

The parameters of the noise component are derived directly from design-based variance-covariance information. The state CPS variance estimates are obtained through the method of generalized variance functions (see Chapter 13). State level autocorrelations of the sampling error are based on research conducted by Dempster and Hwang (1990) that used a variance component model to compute autocorrelations for the sampling error. After the unobserved signal and noise components are put into the state-space form, the unknown parameters of the variance components of the signal are estimated by maximum likelihood using the Kalman filter (Harvey, 1989). Given these parameter values, the filter calculates the expected value of the signal and the noise components at each point of time conditional on the observed data up to the given time point. As more data become available, previous estimates are updated by a process called smoothing (Maybeck, 1979). For more details, see Tiller (1989).

DIAGNOSTIC TESTING

A model should adequately represent the main features of movements in the CPS. An analysis of the model's prediction errors is the primary tool for assessing goodness of fit. This is an indirect test of the model. The actual model error is the difference between the true value of the signal and the model's estimate of that value. Since we do not observe the true values, but only the CPS, which contains sampling error, we cannot compute the actual model error. The overall model, however, provides an estimate of the signal and sampling error, which sum to an estimate of the CPS. We may, therefore, use the model to predict new CPS observations. If the model's prediction errors are larger than expected or do not average to zero, then this suggests that the signal and/or noise components may be misspecified. In this way, an examination of the model errors in predicting the CPS provides an overall test of its consistency with the CPS data.

The prediction errors are computed as the difference between the current values of the CPS and the predictions of the CPS made from the model, based on data prior to the current period. Since these errors represent movements not explained by the model, they should not contain any systematic information about the behavior of the signal or noise component of the CPS. Specifically, the prediction errors, when standardized, should approximate a randomly distributed normal variate with zero mean (unbiased) and constant variance. The models are subjected to a battery of diagnostic tests to check the prediction errors for departure from these properties.

MONTHLY PROCESSING

State agency staff prepare their official monthly estimates using software developed by BLS that implements the KF. The state model-based labor force estimates are generally released 2 weeks after the release of the national labor force estimates. The KF algorithm is particularly well suited for the preparation of current estimates as they become available each month. Since it is a recursive data processing algorithm, it does not require all previous data to be kept in storage and reprocessed every time a new sample observation becomes available. All that is required is an estimate of the state vector and its covariance matrix for the previous month. The computer interface used by the states to make, review, and transmit their model estimates is a system of interactive programs called STARS (State Time Series Analysis and Review System). The software is interactive, querying users for their UI and CPS data and then combining these data with CPS estimates to produce model-based estimates.

END-OF-THE-YEAR PROCESSING

At the end of the year, a number of revisions are made to the model estimates. The model estimates are first re-estimated to incorporate new population estimates obtained

from the Census Bureau and revisions to state-supplied data. The revised estimates are then smoothed and benchmarked to CPS annual averages. Finally, the benchmarked series is then seasonally adjusted using the X-11 ARIMA seasonal adjustment procedure.

Re-Estimation and Benchmarking

After the Census Bureau provides new population controls to BLS, the state CPS labor force estimates are adjusted to these controls. Similarly, state agencies revise their CES and UI claims data as more information about the series becomes available throughout the year. Using the revised state CPS and input data, the KF produces a revised estimate for the current year. The revised estimates are smoothed, using a fixed interval smoothing algorithm, to incorporate data accumulated during the current year.

A benchmarking process follows the smoothing of the forward filter estimates. Since the CPS state samples were designed to produce reliable annual average estimates, the annual average of the smoothed estimates is forced to equal the CPS annual average. The method used during the benchmarking procedure is called the Denton method. It forces the average of the monthly smoothed estimates to equal the CPS state annual averages while minimizing distortions in the month-to-month changes of the smoothed estimates.

Seasonal Adjustment

The model estimates are seasonally adjusted using the X-11 ARIMA seasonal adjustment procedure. Seasonal factors for the first half of the current year, January through June, are based on the historical series. Factors for July through December are calculated from a data series comprising the historical series and the forward filter estimates for the first half of the current year. The models are designed to suppress sampling error but not to decompose the series into seasonal and nonseasonal variation. This preprocessed series can then be adequately decomposed by the X-11 filters (Tiller, 1996).

RESULTS

Using the basic model structure described above, employment and unemployment models were developed for all of

State Unemployment Rate: 1976-99

Percent of labor force

Signal
CPS

the states. Auxiliary variables, available by state, were used in the regression components: Current Employment Statistics survey employment (CESEM), worker claims for unemployment insurance benefits (UI), and population estimated by the Census Bureau. The general form for each model is

$$CPSEP(t) = \alpha(t)CESEP(t) + Trend(t) + Seasonal(t) + Noise(t) \tag{24}$$

$$CPSRT(t) = \delta(t)CLRST(t) + Trend(t) + Seasonal(t) + Noise(t)$$

where:

CESEP = 100(CESEM/POP)
CLRST = 100(UI/CESEM)
POP = noninstitutional civilian 16+ population

The basic signal component consists of a regression component with a time varying coefficient, trend level and slope, and six seasonal frequencies, irregular variation and outliers. The basic noise component consists of sampling error. Once estimated, these models were subjected to diagnostic testing. In a well-specified model, the standardized one-step-ahead prediction errors should behave approximately as white noise, that is, be uncorrelated with a zero mean and fixed variance. To be acceptable, the final model was required to show no serious departures from the white noise properties. Once satisfactory results were obtained, further decisions were based on goodness of fit measures and Akaike's Information Criterion (Harvey,1989) and on subject matter knowledge.

Often, one or more of these components could be simplified. For the signal, the trend slope could often be dropped and the number of seasonal frequencies reduced. In many cases, the estimated variance of the irregular was close to zero, allowing it to be dropped from the model.

A graph of the CPS unemployment rate and the signal from a state unemployment rate model is on page E–5. The signal is considerably smoother than the CPS. Elimination of the sampling error from the CPS by signal extraction removed about 65 percent of the monthly variation in the series.

REFERENCES

Bell, W.R. and Hillmer, S.C. (1990), "The Time Series Approach to Estimation for Repeated Surveys," **Survey Methodology**, 16, 195-215.

Binder, D.A. and Dick, J.P. (1989),"Modeling and Estimation for Repeated Surveys," **Survey Methodology,** 15, 29-45.

Blaustein, S.J. (1979), paper prepared for the National Commission on Employment and Unemployment Statistics.

Dempster, A.P., and Hwang, J.S. (1990), "A Sampling Error Model of Statewide Labor Force Estimates From the CPS," Paper prepared for the U.S. Bureau of Labor Statistics.

Evans, T., Tiller, R., and Zimmerman, T. (1993), "Time Series Models for State Labor Force Series," **Proceedings of the Survey Research Methods Section**, American Statistical Association.

Harvey, A.C. (1989), **Forecasting, Structural Time Series Models and the Kalman Filter,** Cambridge: Cambridge University Press.

Maybeck, P.S. (1979), **Stochastic Models, Estimation, and Control** (Vol. 2), Orlando: Academic Press.

Pfeffermann, D. (1992), "Estimation and Seasonal Adjustment of Population Means Using Data From Repeated Surveys," **Journal of Business and Economic Statistics,** 9, 163-175.

Scott, A.J., and Smith, T.M.F. (1974), "Analysis of Repeated Surveys Using Time Series Methods," **Journal of the American Statistical Association,** 69, 674-678.

Scott, A.J., Smith, T.M.F., and Jones, R.G. (1977), "The Application of Time Series Methods to the Analysis of Repeated Surveys," **International Statistical Review**, 45, 13- 28.

Thisted, R.A. (1988), **Elements of Statistical Computing: Numerical Computation,** New York: Chapman and Hall.

Tiller, R. (1989), "A Kalman Filter Approach to Labor Force Estimation Using Survey Data," **Proceedings of the Survey Research Methods Section,** American Statistical Association, pp. 16-25.

Tiller, R. (1992), "An Application of Time Series Methods to Labor Force Estimation Using CPS Data," **Journal Of Official Statistics,** 8, 149-166.

U.S. Department of Labor, Bureau of Labor Statistics (1997), **BLS Handbook of Methods,** Bulletin 2490.

Zimmerman, T., Tiller, R., and Evans, T. (1994), "State Unemployment Rate Time Series Models," **Proceedings of the Survey Research Methods Section,** American Statistical Association.

Appendix F.
Organization and Training of the Data Collection Staff

INTRODUCTION

The data collection staff for all Census Bureau programs is directed through 12 regional offices (ROs) and 3 telephone centers (computer assisted telephone interviewing - CATI). The ROs collect data in two ways; CAPI (computer assisted personal interviewing) and PAPI (paper-and-pencil interviewing). The 12 ROs report to the Chief of the Field Division whose headquarters is located in Washington, DC. The three CATI facility managers report to the Chief of the National Processing Center (NPC).

ORGANIZATION OF REGIONAL OFFICES/CATI FACILITIES

The staffs of the ROs and CATI facilities carry out the Census Bureau's field data collection programs, both sample surveys and censuses. Currently, the ROs supervise about 6,000 part-time and intermittent field representatives (FRs) who work on continuing current programs and one-time surveys. Approximately 2,100 of these FRs work on the Current Population Survey (CPS). When a census is being taken, the field staff increases greatly.

The location of the ROs and the boundaries of their responsibilities are displayed in Figure F–1. RO areas were originally defined to evenly distribute the office workloads for all programs. Table F–1 shows the average number of CPS units assigned for interview per month in each RO.

A regional director is in charge of each RO. Program coordinators report to the director through an assistant regional director. The CPS is the responsibility of the demographic program coordinator who has one or two CPS program supervisors on staff. The program supervisor has a staff of one or two office clerks working essentially full time. Most of the clerks are full-time civil servants who work in the RO. The typical RO employs about 100 to 250 FRs who are assigned to the CPS. Most FRs also work on other surveys. The RO usually has 15 to 18 senior field representatives (SFRs) who act as team leaders to FRs. Each team leader is assigned 6 to 10 FRs. The primary function of the team leader is to assist the program supervisors with training and supervising the field interviewing staff. In addition, the SFRs conduct response follow-up with eligible households. Like other FRs, the SFR is a part-time or intermittent employee who works out of his or her home. Despite the geographic dispersion of the sample areas, there is a considerable amount of personal contact between the supervisory staff and the FRs. This is accomplished mainly through the training programs and various aspects of the quality control program. For some of the outlying PSUs, it is necessary to use the telephone and written communication to keep in continual touch with all FRs. With the introduction of CAPI, the ROs also communicate with the FRs using e-mail. Assigning new functions, such as team leaders, also improves communications between the ROs and the interviewing staff. In addition to communications relating to the work content, there is a regular system for reporting progress and costs.

The CATI centers are staffed with one facility manager who directs the work of two to three supervisory survey statisticians. Each supervisory survey statistician is in charge of about 15 supervisors and between 100-200 interviewers.

A substantial portion of the budget for field activities is allocated to monitoring and improving the quality of the FRs' work. This includes FRs group training, monthly home studies, personal observation, and reinterview. Approximately 25 percent of the CPS budget (including travel for training) was allocated to quality enhancement. The remaining 75 percent of the budget went to FR and SFR salaries, all other travel, clerical work in the ROs, recruitment, and the supervision of these activities.

TRAINING FIELD REPRESENTATIVES

Approximately 20 to 25 percent of the CPS FRs leave the staff each year. As a result, the recruitment and training of new FRs is a continuing task in each RO. To be selected as a CPS FR, a candidate must pass the Field Employee Selection Aid test on reading, arithmetic, and map reading. The FR is required to live in the Primary Sampling Unit (PSU) in which the work is to be performed and have a residence telephone and in most situations, an automobile. As a part-time or intermittent employee, the FR works 40 hours or less per week or month. In most cases, new FRs are paid at the GS-3 level and are eligible for payment at the GS-4 scale after 1 year of fully successful or better work. FRs are paid mileage for the use of their own cars while interviewing and for commuting to classroom training sites. They also receive pay for completing their home study training packages.

FIELD REPRESENTATIVE TRAINING PROCEDURES

Initial training for new field representatives. Each FR, when appointed, undergoes an initial training program prior to starting his/her assignment. The initial training program

Figure F–1. Map

2000 Decennial and Survey Boundaries

Table F–1. **Average Monthly Workload by Regional Office: 2001**

Regional office	Base workload	CATI workload	Total workload	Percent
Total	66,576	5,640	72,216	100.00
Boston	8,914	615	9,529	13.20
New York	2,887	361	3,248	4.50
Philadelphia	5,915	786	6,701	9.28
Detroit	5,006	286	5,292	7.33
Chicago	4,430	525	4,955	6.86
Kansas City	6,566	450	7,016	9.71
Seattle	5,147	529	5,676	7.86
Charlotte	5,310	366	5,677	7.86
Atlanta	4,853	376	5,230	7.24
Dallas	4,094	320	4,414	6.11
Denver	9,681	815	10,495	14.53
Los Angeles	3,773	212	3,984	5.52

consists of up to 20 hours of pre-classroom home study, 3.5 to 4.5 days of classroom training (dependent upon the trainee's interview experience) conducted by the program supervisor or coordinator, and as a minimum, an on-the-job field observation by the program supervisor or SFR during the FRs first 2 days of interviewing. The classroom training includes comprehensive instruction on the completion of the survey using the laptop computer. In classroom training, special emphasis is placed on the labor force concepts to ensure that the new FRs fully grasp these concepts before conducting interviews. In addition, a large part of the classroom training is devoted to practice interviews that reinforce the correct interpretation and classification of the respondents' answers.

Each FR completes a home study exercise before the second month's assignment and, during the second month's interview assignment, is observed for at least 1 full day by the program supervisor or the SFR who gives supplementary training, as needed. The FR also completes a home study exercise and a final review test prior to the third month's assignment.

Training for all field representatives. As part of each monthly assignment, FRs are required to complete a home study exercise which usually consists of questions concerning labor force concepts and survey coverage procedures. Once a year, the FRs are gathered in groups of about 12 to 15 for 1 or 2 days of refresher training. These sessions are usually conducted by program supervisors with the aid of SFRs. These group sessions cover regular CPS and supplemental survey procedures.

Training for interviewers at the CATI centers. Candidates selected to be CATI interviewers receive 3 days of training on using the computer. This initial training does not cover subject matter material. While in training, new CATI interviewers are monitored for a minimum of 5 percent of their time on the computer. This is compared to 2.5 percent monitoring time for experienced staff. In addition, once a CATI interviewer has been assigned to conduct interviews on the CPS, she/he receives an additional 3 1/2 days of classroom training.

FIELD REPRESENTATIVE PERFORMANCE GUIDELINES

Performance guidelines have been developed for CPS CAPI FRs for response/nonresponse and production. Response/nonresponse rate guidelines have been developed to ensure the quality of the data collected. Production guidelines have been developed to assist in holding costs within budget and to maintain an acceptable level of efficiency in the program. Both sets of guidelines are intended to help supervisors analyze activities of individual FRs and to assist supervisors in identifying FRs who need to improve performance.

Each CPS supervisor is responsible for developing each employee to his/her fullest potential. Employee development can only be accomplished by providing meaningful feedback on a continuous basis. By acknowledging strong points and highlighting areas for improvement, the CPS supervisor can monitor an employee's progress and take appropriate steps to improve weak areas.

FR performance is measured by a combination of the following: response rates, production rates, supplement response, reinterview results, observation results, submitting accurate payrolls on time, meeting deadlines, reporting to team leaders, and attending training sessions.

The most useful tool to help supervisors' evaluate the FRs performance is the CPS 11-39 form. These reports are generated monthly and are produced using data from the ROSCO and CARMN systems.

This report provides the supervisor with information on:

Workload and number of interviews
Response rate and adjective rating
Noninterview results (counts)
Production rate, adjective rating, and mileage
Observation and reinterview results
Meeting transmission goals
Rate of personal visit telephone interviews
Supplement response rate
Refusal/Don't Know counts and rates
Industry and Occupation (I&O) entries Jeffersonville could not code.

EVALUATING FIELD REPRESENTATIVE PERFORMANCE

Census Bureau headquarters, located in Suitland, Maryland provides guidelines to the ROs for developing performance standards for FRs response and production rates. The ROs have the option of using the guidelines, modifying them, or establishing a completely different set of standards for their FRs. If the RO establishes their own standards, the RO must notify the FRs of the standards.

Maintaining high response rates is of primary importance to the Census Bureau. The response rate is defined as the proportion of all sample households eligible for interview that are actually interviewed. It is calculated by dividing the total number of interviewed households by the sum of interviewed households and the number of refusals, those that are temporarily absent, noncontacts, and noninterviewed households for other reasons. (All of these noninterviews are referred to as Type A noninterviews.) Type A cases do not include vacant units, those that are used for nonresidential purposes, or other addresses that are not eligible for interview.

Production guidelines. The production guidelines used in the CPS CAPI program are designed to measure the efficiency of individual FRs and the RO field functions. Efficiency is measured by total minutes per case which includes interview time and travel time. It is calculated by dividing total time reported on payroll documents by total workload. The standard acceptable minutes per case rate for FRs varies with the characteristics of the PSU. When looking at an FRs production, a program supervisor must consider extenuating circumstances, such as:

- Unusual weather conditions such as floods, hurricanes, or blizzards.

- Extreme distances between sample units, or assignment covers multiple PSUs.

- Large number of inherited or confirmed refusals.

- Working part of another FRs assignment.

- Inordinate number of temporarily absent cases.

- High percentage of Type B/C noninterviews that decrease the base or nonresponse rate.

- Other substantial changes in normal assignment conditions.

Supplement response rate. The supplement response rate is another measure that CPS program supervisors must use in measuring the performance of their FR staff.

Transmittal rates. The ROSCO system allows the supervisor to monitor transmittal rates of each CPS FR. A daily receipts report is printed each day showing the progress of each case on CPS.

Observation of field work. Field observation is one of the methods used by the supervisor to check and improve performance of the FR staff. It provides a uniform method for assessing the FRs attitudes toward the job, use of the computer and evaluating the extent to which FRs apply CPS concepts and procedures during actual work situations. There are three types of observations:

1. Initial observations.

2. General performance review.

3. Special needs.

Initial observations are an extension of the initial classroom training for new hires and provides on-the-job training for FRs new to the survey. They also allow the survey supervisor to assess the extent to which a new CPS CAPI FR grasps the concepts covered in initial training and, therefore, are an integral part of the initial training given to all FRs. A 2-day initial observation (N1) is scheduled during the FRs first CPS CAPI assignment. A second 1-day initial observation (N2) is scheduled during the FRs second CPS CAPI assignment. A third 1-day initial observation (N3) is scheduled during the FRs fourth through sixth CPS CAPI assignment.

General performance review observations are conducted at least annually and allow the supervisor to provide continuing developmental feedback to all CPS CAPI FRs. Each CPS CAPI FR is regularly observed at least once a year.

Special-needs observations are made when there is evidence of a FR having problems or poor performance. The need for a special-needs observation is usually detected by other checks on the FR's work. For example, special-needs observations are conducted if a FR has a high Type A noninterview rate, a high minutes per case rate, a failure on reinterview, an unsatisfactory on a previous observation, a request for help, or for other reasons related to the FR's performance.

An observer accompanies the FR for a minimum of 6 hours during an actual work assignment. The observer takes note of the FR's performance including how the interview is conducted and how the computer is used. The observer stresses good interviewing techniques: asking questions as worded and in the order presented on the CAPI screen, adhering to instructions on the instrument and in the manuals, knowing how to probe, recording answers correctly and in adequate detail, developing and maintaining good rapport with the respondent conducive to an exchange of information, avoiding questions or probes that suggest a desired answer to the respondent, and determining the most appropriate time and place for the interview.

The observer reviews the FR's household performance and discusses the FR's strong and weak points with an emphasis on correcting habits that interfere with the collection of reliable statistics. In addition, the FR is encouraged to ask the observer to clarify survey procedures not

fully understood and to seek the observer's advice on solving other problems encountered.

Unsatisfactory performance. When the performance of a FR is at the unsatisfactory level over any period (usually 90 days), he/she may be placed in a trial period for 30 to 90 days. Depending on the circumstances, the FR will be issued a letter stating that he/she is being placed in a Performance Opportunity Period (POP) or a Performance Improvement Period (PIP). These administrative actions warn the FR that his/her work is substandard, makes specific suggestions on ways to improve performance, alerts the FR to actions that will be taken by the survey supervisor to assist the FR to improve his/her performance, and notifies the FR that he/she is subject to separation if the work does not show improvement in the allotted time.

Appendix G.
Reinterview: Design and Methodology

INTRODUCTION

A continuing program of reinterviews on subsamples of Current Population Survey (CPS) households is carried out every month. Reinterview involves a second interview where all the labor force questions are repeated. The reinterview program is one of our major tools for limiting the occurrence of nonsampling error and is a critical part of the CPS program. The CPS reinterview program has been in place since 1954. Reinterviewing for CPS serves two main purposes: as a quality control (QC) tool to monitor the work of the field representatives (FRs) and to evaluate data quality via the measurement of response error (RE).

Prior to the automation of CPS in January 1994, the reinterview consisted of one sample selected in two stages. The FRs were primary sampling units and the households within the FRs assignments were secondary sampling units. In 75 percent of the reinterview sample differences between original and reinterview responses were reconciled, and the results were used both to monitor the FRs and to estimate response bias; that is, the accuracy of the original survey responses. In the remaining 25 percent of the sample the differences were not reconciled and were used only to estimate simple response variance; that is, the consistency in response between the original interview and reinterview. Because the one sample approach did not provide a monthly reinterview sample that was fully representative of the original survey sample for estimating response error, the decision was made to separate the RE and QC reinterview samples beginning with the introduction of the automated system in January 1994.

As a QC tool, reinterviewing is used to deter and detect falsification. As such, it provides a means of limiting nonsampling error as described in Chapter 15. The RE reinterview currently measures simple response variance or reliability.

The measurement of simple response variance in the reinterview assumes an independent replication of the interview. However, this assumption does not always hold, since the respondent may remember his or her interview response and repeat it in the reinterview (conditioning). Also the fact that the reinterview is done by telephone for personal visit interviews may violate this assumption. In terms of the measurement of RE, errors or variations in response may affect the accuracy and reliability of the results of the survey. Responses from interviews and reinterviews are compared and differences identified and analyzed. (See Chapter 16 for a more detailed description of response variance.)

Sample cases for QC and RE reinterviews are selected by different methods and have somewhat different field procedures. To minimize response burden, a household is only reinterviewed one time (or not at all) during its life in sample. This rule applies to both the RE and QC samples. Any household contacted for reinterview is ineligible for selection during its remaining months in sample. An analysis of respondent participation in later months of the CPS showed that the reinterview had no significant effect on the respondent's willingness to respond (Bushery, Dewey, and Weller, 1995). Sample selection for reinterview is done immediately after the monthly assignments are certified (see Chapter 4).

RESPONSE ERROR SAMPLE

The regular RE sample is selected first. It is a systematic random sample across all households eligible for interview each month. It includes households assigned to both the telephone centers and the regional offices. Only households which can be reached by telephone and for which a completed or partial interview is obtained are eligible for RE reinterview. This restriction introduces a small bias into the RE reinterview results because households without "good" telephone numbers are made ineligible for the RE reinterview. About 1 percent of CPS households are assigned for RE reinterview each month.

QUALITY CONTROL SAMPLE

The QC sample is selected next. The QC sample has the FRs in the field as its first stage of selection. The QC sample does not include interviewers at the telephone centers. It is felt that the monitoring operation at the telephone centers sufficiently serves the QC purpose. The QC sample uses a 15-month cycle. The FRs are randomly assigned to 15 different groups. Both the frequency of selection and the number of households within assignments are based upon the length of tenure of the FRs. Through a falsification study it was determined that a relationship exists between tenure and both frequency of falsification and the percentage of assignment falsified (Waite, 1993 and 1997). Experienced FRs (those with at least 5 years of service) were found less likely to falsify. Also, experienced FRs who did falsify were more circumspect, falsifying fewer cases within their assignments than inexperienced FRs (those with under 5 years of service).

Because inexperienced FRs are more likely to falsify, more of them are selected for reinterview each month: three groups of inexperienced FRs to two groups of experienced FRs. On the other hand, since inexperienced FRs falsify a greater percentage of cases within their assignments, fewer of their cases are needed to detect falsification. For inexperienced FRs, five households are selected for reinterview. For experienced FRs, eight households are selected. The selection system is set up so that an FR is in reinterview at least once, but no more than four times within a 15-month cycle.

A sample of households assigned to the telephone centers is selected for QC reinterview, but these households become eligible for reinterview only if recycled to the field and assigned to FRs already selected for reinterview. Recycled cases are included in reinterview because recycles are more difficult to interview and may be more subject to falsification.

All cases, except noninterviews for occupied households, are eligible for QC reinterview: completed and partial interviews and all other types of noninterviews (vacant, demolished, etc.). FRs are evaluated on their rate of noninterviews for occupied households. Therefore, FRs have no incentive to misclassify a case as a noninterview for an occupied household.

Approximately 2 percent of CPS households are assigned for QC reinterview each month.

REINTERVIEW PROCEDURES

QC reinterviews are conducted out of the regional offices by telephone, if possible, but in some cases by personal visit.[1] They are conducted on a flow basis extending through the week following the interview week. They are conducted mostly by senior FRs and sometimes by program supervisors. For QC reinterviews, the reinterviewers are instructed to try to reinterview the original household respondent, but are allowed to conduct the reinterview with another eligible household respondent.

The QC reinterviews are computer assisted and are a brief check to verify the original interview outcome. The reinterview asks questions to determine if the FR conducted the original interview and followed interviewing procedures. The labor force questions are not asked.[2] The QC reinterview instrument also has a special set of outcome codes to indicate whether or not any noninterview misclassifications occurred or any falsification is suspected.

RE reinterviews are conducted only by telephone from both the telephone centers and out of the regional offices (ROs). Cases interviewed at the telephone centers are reinterviewed from the telephone centers, and cases interviewed in the field are reinterviewed in the field. At the telephone centers reinterviews are conducted by the telephone center interviewing staff on a flow basis during interview week. In the field they are conducted on a flow basis and by the same staff as the QC reinterviews. The reinterviewers are instructed to try to reinterview the original household respondent, but are allowed to conduct the reinterview with another knowledgeable adult household member.[3] All RE reinterviews are computer assisted and consist of the full set of labor force questions for all eligible household members. The RE reinterview instrument dependently verifies household membership and asks the industry and occupation questions exactly as they are asked in the original interview.[4] Currently, no reconciliation is conducted.[5]

SUMMARY

Periodic reports on the QC reinterview program are issued showing the number of FRs determined to have falsified data. Since FRs are made aware of the QC reinterview program and are also informed of the results following reinterview, falsification is not a major problem in the CPS. Only about 0.5 percent of CPS FRs are found to falsify data. These FRs either resign or are terminated.

From the RE reinterview, results are also issued on a periodic basis. They contain response variance results for the basic labor force categories and for certain selected other questions.

REFERENCES

Bushery, J. M., Dewey, J. A., and Weller, G. (1995), "Reinterview's Effect on Survey Response," **Proceedings of the 1995 Annual Research Conference,** U.S. Census Bureau, pp. 475-485.

U.S. Census Bureau (1996), **Current Population Survey Office Manual,** Form CPS-256, Washington, DC: Government Printing Office, ch. 6.

U.S. Census Bureau (1996), **Current Population Survey SFR Manual,** Form CPS-251, Washington, DC: Government Printing Office, ch. 4.

U.S. Census Bureau (1993), "Falsification by Field Representatives 1982-1992," memorandum from Preston Jay Waite to Paula Schneider, May 10, 1993.

U.S. Census Bureau (1997), "Falsification Study Results for 1990-1997," memorandum from Preston Jay Waite to Richard L. Bitzer, May 8, 1997.

---

[1]Because the mail component was ineffective in detecting falsification, it was stopped as of March 1998.

[2]Prior to October 2001, QC reinterviews asked the full set of labor force questions for all eligible household members.

[3]Beginning in February 1998, the RE reinterview respondent rule was made the same as the QC reinterview respondent rule.

[4]Prior to October 2001, RE reinterviews asked the industry and occupation questions independently.

[5]Reconciled reinterview, which was used to estimate reponse bias and provide feedback for FR performance, was discontinued in January 1994.

Appendix H.
Sample Design Changes of the Current Population Survey: January 1996

(See Appendix J for sample design changes in July 2001.)

INTRODUCTION

As of January 1996, the 1990 Current Population Survey (CPS)[1] sample changed because of a funding reduction. The budget made it necessary to reduce the national sample size from roughly 56,000 eligible housing units to 50,000 eligible housing units and from 792 sample areas to 754 sample areas. The U.S. Census Bureau and the Bureau of Labor Statistics (BLS) decided to achieve the budget reduction by eliminating the oversampling in CPS in seven states and two substate areas that made it possible to produce reliable monthly estimates of unemployment and employment in these areas. In effect, this decision produced the least possible damage to the precision of national estimates with a sample reduction of this size. It was a confirmation of the high priority attached to national statistics and data by demographic and social subgroups as compared to geographic detail. The four largest states had not required any oversampling for the production of monthly estimates, so no reductions were made in their sample sizes. The sample was reduced in seven states: Illinois, Massachusetts, Michigan, New Jersey, North Carolina, Ohio, and Pennsylvania; and two substate areas: Los Angeles and New York City.

Survey Requirements

The original survey requirements stated in Chapter 3 remain unchanged except for the reliability requirements. The new reliability requirements are:

1. The coefficient of variation (CV) on the monthly unemployment level for the Nation, given a 6 percent unemployment rate, is 1.9 percent.

2. The CV on the annual average unemployment level given a 6 percent unemployment rate for each of the 50 states and the District of Columbia is 8 percent or lower.

   These requirements effectively eliminated substate area reliability requirements and monthly CV requirements for 11 large states: California, Florida, Illinois, Massachusetts, Michigan, New Jersey, New York, North Carolina, Ohio, Pennsylvania, and Texas. The 39 other states and the District of Columbia were not affected by this sample reduction. These states kept their original designs which meet an 8 percent CV requirement on the annual average estimates of unemployment, assuming a 6 percent unemployment rate.

Strategy for Reducing the Sample

The sample reduction plan was designed to meet both the national and state reliability requirements. The sample sizes in all states and the District of Columbia, other than the 11 largest, were already at the levels necessary to meet the national reliability requirements, so changes in their sizes were not required. The sample sizes in 7 of the 11 largest states and in New York City and Los Angeles were greater than necessary to meet the new requirements, and the sample sizes in these areas were reduced. The sample sizes in the four most populous states — Florida, Texas, and the parts of California and New York outside Los Angeles and New York City — were also greater than the precision requirements necessitated, but reductions in their sample sizes would have created problems in achieving the national goals. The sample sizes in these four areas used in the redesign, therefore, were retained.

For the redesigned sample, an attempt was made to allocate sample proportionally to the state populations in the remaining large states and substate areas: Illinois, Massachusetts, Michigan, New Jersey, North Carolina, Ohio, Pennsylvania, Los Angeles, and New York City. Such an approach would yield the same sampling rate in all of these areas. The use of a consistent sampling rate would tend to minimize the increase in national variance resulting from the reduction. However, it was necessary to allow for a fair amount of variation in the sampling rates among these areas in order to meet the reliability requirement for annual average data for each of the 50 states (see Table H–2). This design results in variations in the precision among the individual states resulting in increased expected annual average CVs on estimates of unemployment ranging from 2.9 to 5.7 percent. The proportion of sample cut because of the reduction and the expected CVs before and after the reduction is shown in Table H–1. All 11 large states have better expected reliability than the 39 other states and the District of Columbia.

---

[1]The CPS produced by the redesign after the 1990 census is termed the 1990 CPS.

Table H–1. **The Proportion of Sample Cut and Expected CVs Before and After the January 1996 Reduction**

State	Proportion of sample cut	Expected CV (percent)			
		Before reduction		After reduction	
		Monthly	Annual	Monthly	Annual
United States.	0.107	1.8	0.9	1.9	* 0.9
California	0.024	6.2	2.9	6.3	2.9
Los Angeles	0.059	9.2	4.1	9.5	4.3
Remainder of state	–	7.9	3.7	7.9	3.7
Florida	–	8.2	3.7	8.2	3.7
Illinois	0.130	7.9	3.7	8.8	4.2
Massachusetts	0.478	7.9	3.5	11.1	4.2
Michigan	0.320	7.8	3.6	9.7	4.5
New Jersey	0.348	8.0	3.6	9.8	4.4
New York	0.195	6.6	3.0	7.1	3.2
New York City	0.348	9.1	4.1	11.5	5.2
Remainder of state	–	8.9	4.1	8.9	4.1
North Carolina	0.458	8.0	3.9	11.1	5.7
Ohio	0.240	7.9	3.6	9.4	4.4
Pennsylvania	0.154	8.0	3.6	8.8	4.1
Texas	0.000	7.7	3.6	7.7	3.6

  * The national CV increased after the reduction but was within the rounding error.

CHANGES TO THE CPS SAMPLE DESIGN

Implementing the January 1996 sample reduction involved modifying the original 1990 CPS state designs described in Chapter 3. Both the first-stage and second-stage sample designs were changed to realize major cost savings. This differed from the maintenance reductions described in Appendix C which affected only the second-stage design. In addition, unlike the maintenance reductions which are implemented over a period of 16 months, this reduction was fully implemented in January 1996.

Many characteristics of the original design were retained. Some of the major characteristics are:

- State-based design.

- PSU definitions.

- Self-representing PSU requirements.

- One PSU per stratum selected to be in sample.

- Systematic sample of ultimate sampling units consisting of clusters of four housing units.

- Rotation group and reduction group assignments.

First-Stage of the Sample Design

New first-stage designs were developed by restratifying the original 1990 design strata. First, the 1990 design strata that were required to be self-representing (SR) remain SR for the new designs[2]. For each state, any nonrequired SR strata and all of the nonself-representing (NSR) strata were grouped into super-strata of two strata

_____
  [2] See Chapter 3 "Stratification of Primary Sampling Units" for the areas which are required to be SR.

(and one group of three if there was an odd number of strata in the state.) The super-strata were formed using a linear programming technique to minimize the between-stratum variation on the 1990 census levels of unemployment and civilian labor force. Then, one stratum was selected from the super-strata with probability proportional to the 1990 population. The sample PSU from this selected stratum remained in sample while the sample PSU from the nonselected stratum was dropped.

The first-stage probability of selection for the PSUs in the new design is the product of the probability of selecting the stratum from the super-stratum and the probability of selecting the PSU from the stratum.

$$P_{fs} = P(\text{stratum|super-stratum}) \times P(\text{PSU | stratum})$$

$$= \frac{N_{stratum}}{N_{super\text{-}stratum}} \times \frac{N_{PSU}}{N_{stratum}} \quad\quad (H.1)$$

$$= \frac{N_{PSU}}{N_{super\text{-}stratum}}$$

where

P_{fs} is the first-stage unconditional probability of selection of the PSU from the super-stratum,

$P(\text{stratum|super-stratum})$ is the probability of selection of the stratum from the super-stratum,

$P(\text{PSU|stratum})$ is the probability of selection of the PSU from the stratum,

$N_{PSU},$ $N_{stratum},$ and $N_{super\text{-}stratum}$ are the corresponding 1990 populations.

The restratifications were developed for five states: Illinois, Michigan, North Carolina, Ohio, and Pennsylvania. Two states, Massachusetts and New Jersey, and two substate areas, Los Angeles and New York City, are entirely SR and for the most part required to be SR so no

PSUs were eliminated. The reduced CPS national design, as of January 1996, contains 754 stratification PSUs in sample of which 428 are SR and 326 are NSR. Table H–2 contains the number of strata by state for the nine states in which reductions were made.

Table H–2. **January 1996 Reduction: Number and Estimated Population of Strata for 754-PSU Design for the Nine Reduced States**

State	Self-representing (SR)		Nonself-representing (NSR)		State sampling interval
	Strata	Estimated population[1]	Strata	Estimated population[1]	
California	21	20,866,697	5	1,405,500	2,745
Los Angeles	1	6,672,035	0	0	1,779
Remainder of state	20	14,194,662	5	1,405,500	3,132
Illinois	11	6,806,874	6	1,821,548	2,227
Massachusetts	31	4,719,188	0	0	1,853
Michigan	12	5,643,817	5	1,345,074	2,098
New Jersey	11	6,023,359	0	0	1,841
New York	13	12,485,029	7	1,414,548	1,970
New York City	1	5,721,495	0	0	1,774
Remainder of state	12	6,763,534	7	1,414,548	2,093
North Carolina	11	2,840,376	13	2,221,874	1,986
Ohio	13	6,238,600	6	1,958,138	2,343
Pennsylvania	13	7,584,129	6	1,628,780	2,149

[1]Projected 1995 estimates of civilian noninstitutional population ages 16 and over.

Note: This results in an overall national sampling interval of 2,255 based on the 1995 population distribution.

Table H–3. **Reduction Groups Dropped From SR Areas**

State	Number of reduction groups dropped
California	–
Los Angeles	5
Remainder of state	0
Illinois	3
Massachusetts	49
Michigan	25
New Jersey	34
New York	–
New York City	35
Remainder of state	0
North Carolina	41
Ohio	13
Pennsylvania	6

Second-Stage of the Sample Design

New second-stage designs were developed for SR PSUs but not NSR PSUs. The NSR PSUs remaining in sample were originally designed to have a manageable workload for a single field representative. Cutting sample housing units for NSR PSUs would make workloads inefficient for these PSUs.

Within SR PSUs, the within-PSU sampling intervals were adjusted and subsamples of ultimate sampling units (USU) were dropped from sample. The subsample of USUs to drop was identified by reduction groups (see Appendix C). The number of reduction groups to drop from the SR PSUs in each state was a function of the state sampling interval required to achieve the national reliability given the new first-stage designs. The reduction groups were deleted in a specified order so that the nature of the systematic sample was retained to the extent possible. Within each state, the number of reduction groups dropped was uniform for all SR PSUs. Table H–3 gives the number of reduction groups dropped from each SR PSU by state.

Calculation of Sampling Intervals

The design changes do not maintain the self-weighting nature of the original designs. The stratum sampling intervals differ for the SR and NSR portions of the sample. In other words, the probability of selecting a housing unit varies by stratum. The stratum sampling interval is calculated as

$$SI_{stratum} = \frac{SI_{PSU}}{P_{fs}} \times \frac{101}{101 - RGD} \qquad (H.2)$$

where
$SI_{stratum}$ is the stratum sampling interval,
SI_{PSU} is the original within-PSU sampling interval,
P_{fs} is the first-stage probability of selection as defined for formula (H.1), and
RGD is the integer number of reduction groups dropped $(0 \le RGD \le 101)$.

Each NSR PSU has a first-stage probability of selection calculated in formula (H.1), and the number of reduction

groups dropped equals zero. For NSR strata, formula (H.2) reduces to

$$SI_{NSR\ stratum} = \frac{SI_{PSU}}{P_{fs}} \qquad (H.3)$$

SR strata have a first-stage probability of selection equal to one, and the number of reduction groups dropped is found in Table H–3. For SR strata, formula (H.2) reduces to

$$SI_{SR\ stratum} = SI_{PSU} \times \frac{101}{101 - RGD} \qquad (H.4)$$

The state sampling interval is a weighted average of the stratum sampling intervals using 1990 census stratum populations

$$SI = \frac{\sum_{h=1}^{H} N_h\ SI_h}{\sum_{h=1}^{H} N_h} \qquad (H.5)$$

where
SI is the state sampling interval,
N_h is the 1990 census population of the hth stratum in the state,
SI_h is the stratum sampling interval for the hth stratum in the state, and
H is the total number of SR and NSR strata in the state.

Table H–2 gives the resulting state sampling intervals. Since Massachusetts and New Jersey contain only SR strata, they continue to have a self-weighting design.

REVISIONS TO CPS ESTIMATION

Three revisions were made to CPS estimation because of the January 1996 sample reduction, but the general estimation methodology was not changed. The revisions affected only the seven states and two substate areas affected by the reduction. The stages of weighting affected were the basic weight, adjustment for nonresponse, and the first-stage ratio adjustment (see Chapter 10).

Basic Weight Changes

The new basic weights for the remaining sample are the stratum sampling intervals given in formula (H.2). These stratum sampling intervals are the inverse of the probability of selecting a housing unit from the stratum.

Adjustment for Nonresponse Changes

In New Jersey and North Carolina, two noninterview clusters within each state were combined due to small sample sizes. For January 1996 and subsequent months, there are 125 noninterview clusters for the Nation.

First-Stage Ratio Adjustment Changes

The first-stage factors for CPS were recomputed in the five states where PSUs were dropped. The first-stage ratio adjustment factors were revised to account for the new first-stage designs in the following states: Illinois, Michigan, North Carolina, Ohio, and Pennsylvania. The factors did not change for California, Massachusetts, New Jersey, and New York. Table H–4 contains the revised CPS first-stage factors.

Table H–4. **State First-Stage Factors for CPS January 1996 Reduction**

[Beginning January 1996]

State	CPS	
	Black	Non-Black
Illinois	0.96670	0.99477
Michigan	1.00000	** 1.00000
North Carolina	1.04952	1.00641
Ohio	0.92339	0.99953
Pennsylvania	1.17540	0.99157

** Race cells were collapsed.

Monthly State and Substate Estimate Change

After the January 1996 sample reduction, the 11 large states and 2 substate areas had insufficient sample to produce reliable monthly estimates of employment and unemployment. Now these estimates are produced using the same "signal-plus-noise" time series model used for the 39 small states and the District of Columbia discussed in Chapter 10.

ESTIMATION OF VARIANCES

Changing the CPS design because of the January 1996 sample cut required changes to the CPS variance estimation procedures, but the variance estimation methodology did not change. These revisions affected only the seven states and two substate areas where the reduction occurred. A consequence of these changes is that consistent estimates of covariances between December 1995 (and previous months) and January 1996 (and subsequent months) cannot be produced. See Chapter 14 for a description of the CPS variance estimation methodology.

Included in this appendix are estimates of variance properties of the CPS design following the January 1996 reduction. These estimates replace those in Chapter 14 as measures of the current CPS design variance properties[3].

---

[3]The chapters of this document are written as of the conclusion of the implementation of the 1990 CPS design, through December, 1995. Many of the appendixes are written to provide updates that began in January, 1996.

Changes to the Variance Estimation Procedure

Three changes were made to the variance estimation procedure affecting replicate sample composition: new standard error computation units (SECUs) were defined (pairs of USUs), the NSR strata were combined into new pseudostrata, and associated replicate factors were calculated using new Hadamard matrix row assignments. New SECUs were defined in all seven states and two substate areas because the second-stage reduction in SR PSUs resulted in many SECUs with only one USU. NSR strata were combined into new pseudostrata only in the five states with new first-stage designs. These changes resulted in a break in consistent replicate sample definitions between December 1995 and January 1996.

CPS Variance Estimates - 754 PSU Design

Monthly levels. The estimated variance properties on monthly levels of select labor force characteristics from the CPS design following the January 1996 reduction are provided in Tables H–5 to H–8. These tables replace the corresponding Tables 14–2 to 14–5 in Chapter 14 as indicators of the variance properties of the current CPS design. (See previous footnote and see Chapter 14 for description of table headings.) The variance estimates in Tables H–5 to H–8 are averages over 12 months (January to December 1996), while Tables 14–2 to 14–5 are averages over 6 months, making the Appendix H tables more stable than the Chapter 14 tables.

Overall, the reduction was expected to increase CVs by about 4 percent, based on the ratio of the square root of the national sampling rates before and after the reduction. The estimated CVs in Table H–5 are larger than those estimated prior to the reduction (Table 14–2) for all characteristics but Hispanic[4] origin. The reliability of Hispanic origin characteristics is also adversely affected by the reduction. However, the variance on the Hispanic origin CVs is too large to measure this effect.

The effect of the reduction on the contribution of the within- and between-PSU variances to the total variance

---

[4]Hispanics may be of any race.

(Table H–5) is even more difficult to assess because of the variance on the variance estimates. In general, no noticeable effect on the variance estimates is found in the contribution from the first- and second-stages of selection.

Tables H–6 and H–7 show that the sample reduction did not appear to affect the variance properties related to estimation. The new estimated design effects in Table H–8 appear to have increased slightly.

Month-to-month change in levels. The estimated variance properties on month-to-month change in levels of select labor force characteristics from the CPS design following the January 1996 reduction are in Tables H–9 and H–10. These variance estimates are averages of 11 month-to-month change estimates.

Table H–9 indicates how the two stages of sampling affect the variance of the composited estimates of month-to-month change. The last two columns of the table show the percentage of the total variance due to sampling housing units within PSUs (within-PSU variance) and the variance due to sampling a subset of NSR PSUs (between-PSU variance). For all characteristics in the table, the largest component of the total variance is due to within-PSU sampling. The Hispanic origin characteristics and teenage unemployed have smaller within-PSU components than most of the other characteristics. Total and White employed in agriculture also have relatively small within-PSU variance components. Negative numbers in the last column of the table imply negative estimates of the between-PSU variance. Because this component is relatively small and estimated by subtraction, some negative estimates are expected. Since variances cannot be negative, the variance on the estimated variance components is what is actually reflected in these negative estimates.

The reduction in the variance of month-to-month change estimates due to the compositing is shown in Table H–10. The variance factor is the variance of the composited estimate of month-to-month change divided by the first- and second-stage combined (FSC) estimate of month-to-month change. The factors are less than one for all but one of the characteristics shown in the table, indicating that compositing reduces the variance of most month-to-month change estimates.

Table H–5. **Components of Variance for FSC Monthly Estimates**

[Monthly average: January - December 1996]

Civilian noninstitutional population, ages 16 and older	FSC[1]estimate (x 10^6)	Standard error (x 10^5)	Coefficient of variation (%)	Percent of total variance	
				Within	Between
Unemployed	7.31	1.48	2.02	97.4	2.6
White	5.34	1.26	2.36	95.2	4.8
Black	1.61	0.71	4.40	102.8	−2.8
Hispanic origin[2]	1.14	0.62	5.45	94.1	5.9
Teenage, 16-19	1.32	0.59	4.51	97.1	2.9
Employed - agriculture	3.45	1.61	4.68	70.7	29.3
White	3.29	1.56	4.78	69.9	30.1
Black	0.10	0.19	19.82	95.3	4.7
Hispanic origin[2]	0.61	0.73	12.08	91.0	9.0
Teenage, 16-19	0.26	0.27	11.03	104.1	−4.1
Employed - nonagriculture	123.37	3.55	0.29	90.2	9.8
White	104.62	3.13	0.30	91.6	8.4
Black	13.46	1.34	1.00	100.2	−0.2
Hispanic origin[2]	11.06	1.43	1.30	92.1	7.9
Teenage, 16-19	6.26	0.99	1.58	103.5	−3.5
Civilian labor force	134.13	3.18	0.24	93.9	6.1
White	113.25	2.73	0.24	96.1	3.9
Black	15.17	1.26	0.83	95.4	4.6
Hispanic origin[2]	12.80	1.14	0.89	98.0	2.0
Teenage, 16-19	7.84	1.00	1.30	105.0	−5.0
Not in labor force	66.46	3.18	0.48	93.9	6.1
White	55.07	2.73	0.50	96.1	3.9
Black	8.43	1.26	1.50	95.4	4.6
Hispanic origin[2]	6.41	1.14	1.79	98.0	2.0
Teenage, 16-19	7.09	1.04	1.48	104.5	−4.5

[1]FSC = first- and second-stage combined.
[2]Hispanics may be of any race.

Table H–6. Effects of Weighting Stages on Monthly Relative Variance Factors

[Monthly averages: January - December 1996]

Civilian noninstitutional population, ages 16 and older	Relative variance of FSC estimate of level $(\times 10^{-4})$	Relative variance factor[1]			
		Unbiased estimator	Unbiased estimator with—		
			NI[2]	NI & FS[3]	NI & SS[4]
Unemployed	4.109	1.03	1.03	1.10	1.00
White	5.594	1.05	1.04	1.10	1.00
Black	19.392	1.16	1.16	1.20	0.99
Hispanic origin[5]	30.261	1.15	1.14	1.21	1.00
Teenage, 16-19	19.959	1.13	1.13	1.16	1.00
Employed - agriculture	21.763	1.00	0.98	1.10	0.93
White	22.695	1.00	0.98	1.09	0.93
Black	385.214	1.09	1.08	1.06	1.02
Hispanic origin[5]	144.816	1.05	1.03	1.20	0.95
Teenage, 16-19	106.648	1.06	1.05	1.07	0.99
Employed - nonagriculture	0.083	4.41	4.20	5.48	0.96
White	0.090	5.61	5.46	6.48	0.95
Black	0.998	5.88	5.76	5.54	1.01
Hispanic origin[5]	1.674	3.33	3.34	3.60	0.97
Teenage, 16-19	2.499	1.97	1.96	2.01	1.00
Civilian labor force	0.056	6.23	5.88	8.24	0.99
White	0.058	8.38	8.10	10.20	0.99
Black	0.693	7.82	7.66	7.45	1.01
Hispanic origin[5]	0.800	6.65	6.61	7.34	1.00
Teenage, 16-19	1.653	2.47	2.46	2.57	1.00
Not in labor force	0.230	2.88	2.74	3.74	0.99
White	0.247	3.24	3.12	3.98	0.99
Black	2.245	3.98	3.88	3.41	1.00
Hispanic origin[5]	3.193	2.57	2.52	2.76	1.00
Teenage, 16-19	2.154	2.06	2.03	2.26	1.00

[1]Relative Variance Factor is the ratio of the relative variance of the specified level to the relative variance of FSC level.
[2]NI = Noninterview
[3]FS = First Stage
[4]SS = Second Stage
[5]Hispanics may be of any race.

Table H–7. Effect of Compositing on Monthly Variance and Relative Variance Factors

[Monthly averages: January - December 1996]

Civilian noninstitutional population, ages 16 and older	Variance of composited estimate of level $(\times 10^9)$	Variance factor[1]	Relative variance factor[2]
Unemployed	21.063	0.96	0.98
White	15.487	0.97	0.99
Black	4.712	0.93	0.96
Hispanic origin[3]	3.887	0.99	1.00
Teenage, 16-19	3.429	0.99	1.01
Employed - agriculture	23.478	0.91	0.91
White	22.151	0.90	0.91
Black	0.325	0.85	0.87
Hispanic origin[3]	4.764	0.89	0.89
Teenage, 16-19	0.645	0.88	0.89
Employed - nonagriculture	109.120	0.86	0.86
White	85.476	0.87	0.87
Black	15.741	0.87	0.87
Hispanic origin[3]	18.009	0.88	0.88
Teenage, 16-19	8.787	0.90	0.90
Civilian labor force	86.616	0.85	0.86
White	64.105	0.86	0.86
Black	14.260	0.89	0.90
Hispanic origin[3]	11.474	0.87	0.88
Teenage, 16-19	9.163	0.90	0.91
Not in labor force	86.616	0.85	0.85
White	64.105	0.86	0.85
Black	14.260	0.89	0.89
Hispanic origin[3]	11.474	0.87	0.87
Teenage, 16-19	9.696	0.89	0.89

[1]Variance Factor is the ratio of the variance of the composited level to the variance of the FSC level.
[2]Relative Variance Factor is the ratio of the relative variance of the composited level to the relative variance of the FSC level.
[3]Hispanics may be of any race.

Note: Composite formula constants used: K = 0.4 and A = 0.2

Table H–8. **Design Effects for Total Monthly Variance**

[Monthly averages: January - December 1996]

Civilian noninstitutional population, ages 16 and older	Design effect		Within-PSU variance
	Total variance		
	Not composited	Composited	Not composited
Unemployed	1.382	1.339	1.346
White	1.362	1.331	1.296
Black	1.398	1.323	1.437
Hispanic origin[1]	1.539	1.531	1.448
Teenage, 16-19	1.173	1.171	1.139
Employed - agriculture	3.390	3.076	2.396
White	3.362	3.048	2.350
Black	1.702	1.465	1.622
Hispanic origin[1]	3.916	3.482	3.564
Teenage, 16-19	1.243	1.097	1.294
Employed - nonagriculture	1.181	1.018	1.065
White	0.870	0.757	0.797
Black	0.639	0.557	0.640
Hispanic origin[1]	0.869	0.766	0.800
Teenage, 16-19	0.716	0.645	0.742
Civilian labor force	1.014	0.863	0.952
White	0.674	0.576	0.648
Black	0.504	0.452	0.481
Hispanic origin[1]	0.485	0.425	0.476
Teenage, 16-19	0.598	0.542	0.628
Not in labor force	1.014	0.863	0.952
White	0.832	0.710	0.799
Black	0.876	0.780	0.835
Hispanic origin[1]	0.938	0.816	0.919
Teenage, 16-19	0.702	0.625	0.734

[1]Hispanics may be of any race.

Table H–9. **Components of Variance for Composited Month-to-Month Change Estimates**

[Change averages: January - December 1996]

Civilian noninstitutional population, ages 16 and older	Estimate of change[1] (x10^6)	Standard error of change[2] (x10^5)	Percent of total variance[3]	
			Within	Between
Unemployed	0.29	1.59	98.0	2.0
White	0.23	1.30	102.7	−2.7
Black	0.08	0.80	100.3	−0.3
Hispanic origin[4]	0.06	0.67	96.9	3.1
Teenage, 16-19	0.11	0.73	93.7	6.3
Employed - agriculture	0.14	0.82	92.6	7.4
White	0.13	0.79	92.6	7.4
Black	0.01	0.16	103.7	−3.7
Hispanic origin[4]	0.04	0.42	88.8	11.2
Teenage, 16-19	0.05	0.24	101.5	−1.5
Employed - nonagriculture	0.60	2.40	102.5	−2.5
White	0.47	2.11	105.3	−5.3
Black	0.12	1.00	100.7	−0.7
Hispanic origin[4]	0.10	0.97	94.8	5.2
Teenage, 16-19	0.35	0.91	100.6	−0.6
Civilian labor force	0.77	2.30	106.0	−6.0
White	0.60	2.00	106.0	−6.0
Black	0.17	0.98	98.7	1.3
Hispanic origin[4]	0.12	0.89	97.5	2.5
Teenage, 16-19	0.46	0.99	101.9	−1.9
Not in labor force	0.76	2.30	106.0	−6.0
White	0.59	2.00	106.0	−6.0
Black	0.18	0.98	98.7	1.3
Hispanic origin[4]	0.10	0.89	97.5	2.5
Teenage, 16-19	0.45	1.00	100.9	−0.9

[1]The arithmetic mean of the absolute values of the 11 month-to-month changes.
[2]The square root of the arithmetic mean of the variances of the 11 month-to-month changes.
[3]The percent of the estimated total variance attributed to the two stages of sampling.
[4]Hispanics may be of any race.

Table H–10. **Effect of Compositing on Month-to-Month Change Variance Factors**

[Change averages: January - December 1996]

Civilian noninstitutional population, ages 16 and older	Variance of composited estimate of change[1] (x 10^9)	Variance factor[2]
Unemployed	25.124	0.90
White	16.800	0.88
Black	6.409	0.94
Hispanic origin[3]	4.484	0.90
Teenage, 16-19	5.322	0.96
Employed - agriculture	6.717	0.77
White	6.244	0.77
Black	0.249	0.80
Hispanic origin[3]	1.796	0.67
Teenage, 16-19	0.556	0.82
Employed - nonagriculture	57.684	0.70
White	44.718	0.68
Black	10.007	0.76
Hispanic origin[3]	9.397	0.70
Teenage, 16-19	8.228	1.02
Civilian labor force	52.997	0.70
White	39.834	0.68
Black	9.582	0.87
Hispanic origin[3]	7.919	0.70
Teenage, 16-19	9.782	0.98
Not in labor force	52.997	0.70
White	39.834	0.68
Black	9.582	0.87
Hispanic origin[3]	7.919	0.70
Teenage, 16-19	9.907	0.94

[1]The arithmetic mean of the 11 variance estimates.
[2]Variance Factor is the ratio of the variance of the composited estimate to the variance of the FSC estimate.
[3]Hispanics may be of any race.

Appendix I.
New Compositing Procedure

INTRODUCTION

Effective with the release of January 1998 data, BLS implemented a new composite estimation method for the Current Population Survey. The new technique provides increased operational simplicity for microdata users and allows optimization of compositing coefficients for different labor force categories.

In the CPS estimation procedure, described in Chapter 10, a separate weight for each person in the CPS sample is computed and ratio adjusted through a sequence of weighting steps. These adjustments are followed by a composite estimation step that improves the accuracy of current estimates by incorporating information gathered in previous months, taking advantage of the fact that 75 percent of sample households are common in each pair of consecutive months. Under the old procedure, composite estimation was performed at the macro level. The composite estimator for each tabulated cell was a function of aggregated weights for sample persons contributing to that cell in current and prior months. The different months of data were combined together using compositing coefficients. Thus, microdata users needed several months of CPS data to compute composite estimates. To ensure consistency, the same coefficients had to be used for all estimates. The values of the coefficients selected were much closer to optimal for unemployment than for employment or labor force totals.

The new composite weighting method involves two steps: (1) the computation of composite estimates for the main labor force categories, classified by important demographic characteristics and (2) the adjustment of the microdata weights, through a series of ratio adjustments, to agree with these composite estimates, thus incorporating the effect of composite estimation into the microdata weights. Under this procedure, the sum of the composite weights of all sample persons in a particular labor force category equals the composite estimate of the level for that category. To produce a composite estimate for a particular month, a data user may simply access the microdata file for that month and compute a weighted sum. The new composite weighting approach also improves the accuracy of labor force estimates by using different compositing coefficients for different labor force categories. The weighting adjustment method assures additivity while allowing this variation in compositing coefficients.

COMPOSITE ESTIMATION IN THE CPS

As described in Chapter 10, eight panels or rotation groups, approximately equal in size, make up each monthly CPS sample. Due to the 4-8-4 rotation pattern, six of these panels (three-quarters of the sample) continue in sample the following month and one-half of the households in a given month's sample will be back in the sample for the same calendar month 1 year later. The sample overlap improves estimates of change over time. Through composite estimation, the positive correlation among CPS estimators for different months is increased. This increase in correlation improves the accuracy of monthly labor force estimates.

The CPS AK composite estimator for a labor force total (e.g., the number of persons unemployed) in month t is given by

$$Y'_t = (1-K)\hat{Y}_t + K(Y'_{t-1}+\triangle_t) + A\hat{\beta}_t$$

where:

$$\hat{Y}_t = \sum_{i=1}^{8} x_{t,i}$$

$$\triangle_t = \frac{4}{3} \sum_{i \in S} (x_{t,i} - x_{t-1,i-1}) \text{ and}$$

$$\hat{\beta}_t = \sum_{i \notin S} x_{t,i} - \frac{1}{3} \sum_{i \in S} x_{t,i}$$

i   = 1,2,...,8 month in sample

$x_{t,i}$ = sum of weights after second-stage ratio adjustment of respondents in month t, and month-in-sample i with characteristic of interest

S   = {2,3,4,6,7,8} sample continuing from previous month

K   = 0.4 for unemployed
      0.7 for employed

A   = 0.3 for unemployed
      0.4 for employed

The values given above for the constant coefficients A and K are close to optimal with respect to variance for month-to-month change estimates of unemployment level and employment level. The coefficient K determines the weight, in the weighted average, of each of two estimators

for the current month: (1) the current month's ratio estimator $\hat{Y}_t$ and (2) the sum of the previous month's composite estimator Y'_{t-1} and an estimator $\triangle_t$ of the change since the previous month. The estimate of change is based on data from sample households in the six panels common to months t and t-1. The coefficient A determines the weight of $\hat{\beta}_t$, an adjustment term that reduces both the variance of the composite estimator and the bias associated with time in sample. (See Breau and Ernst 1983, Bailar 1975.)

Before January 1998, a single pair of values for K and A was used to produce all CPS composite estimates. Optimal values of the coefficients, however, depend on the correlation structure of the characteristic to be estimated. Research has shown, for example, higher values of K and A result in more reliable estimates for employment levels because the ratio estimators for employment are more strongly correlated across time than those for unemployment. The new composite weighting approach allows use of different compositing coefficients, thus improving the accuracy of labor force estimates, while ensuring the additivity of estimates. For a more detailed description of the selection of compositing parameters, see Lent et al. (1997).

COMPUTING COMPOSITE WEIGHTS

Composite weights are produced only for sample persons age 16 or older. As described in Chapter 10, the CPS estimation process begins with the computation of a "baseweight" for each adult in the survey. The baseweight–the inverse of the probability of selection–is adjusted for non-response, and two successive stages of ratio adjustments to population controls are applied. The second-stage raking procedure, performed independently for each of the eight sample rotation groups, ensures that sample weights add to independent population controls for states, as well as for age/sex/ethnicity groups and age/sex/race groups, specified at the national level.

The post-January 1998 method of computing composite weights for the CPS imitates the second-stage ratio adjustment. Sample person weights are raked to force their sums to equal control totals. Composite labor force estimates are used as controls in place of independent population estimates. The composite raking process is performed separately within each of the three major labor force categories: employed, unemployed, and those not in the labor force.

Adjustment of microdata weights to the composite estimates for each labor force category proceeds as follows. For simplicity, we describe the method for estimating the number of people unemployed (UE); analogous procedures are used to estimate the number of people employed and the number not in the labor force. Data from all eight rotation groups are combined for the purpose of computing composite weights.

1. For each state and the District of Columbia (51 cells) j, the direct (optimal) composite estimate of UE, comp(UE$_j$), is computed as described above. Similarly, direct composite estimates of UE are computed for 9 National age/sex/ethnicity cells (8 Hispanic[1] age/sex cells and 1 non-Hispanic cell) and 66 National age/sex/race cells (38 White age/sex cells, 24 Black age/sex cells, and 4 other age/sex cells). These computations use cell definitions specified in the second-stage ratio adjustment (Tables 10–2 and 10–3), excluding cells containing only persons aged 15 or under. Coefficients K=0.4 and A=0.3 are used for all UE estimates in all categories (state, age/sex/ethnicity, and age/sex/race).

2. Sample records are classified by state. Within each state j, a simple estimate of UE, simp(UE$_j$), is computed by adding the weights of all unemployed sample persons in the state.

3. Within each state j, the weight of each unemployed sample person in the state is multiplied by the following ratio: comp(UE$_j$) / simp(UE$_j$).

4. Sample records are cross-classified by age, sex, and ethnicity. Within each cross-classification cell, a simple estimate of UE is computed by adding the weights (as adjusted in step 3) of all unemployed sample persons in the cell.

5. Weights are adjusted within each age/sex/ethnicity cell in a manner analogous to step 3.

6. Steps 4 and 5 are repeated for age/sex/race cells.

7. Steps 2-6 are repeated five more times for a total of six iterations.

An analogous procedure is done for estimating the number of people employed using coefficients K=0.7 and A=0.4.

For not in labor force (NILF) the same raking steps are performed, but the controls are obtained as the residuals from the population controls and the direct composite estimates for employed (E) and unemployed (UE). The formula is NILF = Population – (E + UE). During computation of composite weights for persons who are unemployed, some further collapsing of cells is needed where cells contain insufficient sample. Any such collapsing of cells for UE leads to a similar collapsing for NILF.

A NOTE ABOUT FINAL WEIGHTS

Prior to the introduction of composite weights the weight after the second-stage ratio adjustment was called the final weight, as stated in Chapter 10. This is no longer true since the weight after compositing is now the final weight. In addition, since data from all eight rotation groups are combined for the purpose of computing composite weights, summations of final weights within each panel do not match independent population controls. However, summations of final weights for the entire sample still match these independent population controls.

---

[1]Hispanics may be of any race.

REFERENCES

Breau, P., and L. Ernst, (1983). ''Alternative Estimators to the Current Composite Estimator.'' **Proceedings of the Section on Survey Research Methods,** American Statistical Association, pp. 397-402.

Bailar, B., (1975). ''The Effects of Rotation Group Bias on Estimates From Panel Surveys.'' **Journal of the American Statistical Association,** 70, pp. 23-30.

Lent, J., S. Miller, P. Cantwell, and M. Duff, (1999). ''Effect of Composite Weights on Some Estimates From the Current Population Survey.'' **Journal of Official Statistics,** to appear.

Appendix J.
Changes to the Current Population Survey Sample in July 2001

INTRODUCTION

In 1999, Congress allocated $10 million annually to the Census Bureau to "make appropriate adjustments to the annual Current Population Survey . . . in order to produce statistically reliable annual state data on the number of low-income children who do not have health insurance coverage, so that real changes in the uninsured rates of children can reasonably be detected." (Public Law 106-113)

The health insurance coverage estimates are derived from the Annual Demographic Supplement (ADS) which is discussed in Chapter 11. Briefly, the ADS is a series of questions asked each March in addition to the regular CPS questions. Consequently, the ADS is also known as the "March Supplement."

To achieve the objectives of the new law, the Census Bureau added sample to the ADS in three different ways:

- General sample increase in selected states

- "Split-path" assignments

- Month-in-sample (MIS) 9 assignments

These changes are collectively known as the State Children's Health Insurance Program (SCHIP) sample expansion. The procedures used to implement the SCHIP sample expansion were chosen in order to minimize the effect on the basic CPS. The general sample increase will be explained first. Later sections of this appendix provide details on the split-path and MIS 9 assignments.

GENERAL SAMPLE INCREASE IN SELECTED STATES

The first part of the SCHIP plan expanded basic monthly CPS sample in selected states, using retired[1] sample from CPS. Sample was identified using CPS sample designation, rotation group codes and all four frames. Expanding the monthly CPS was necessary, rather than simply interviewing many more cases in March, because of the difficulty in managing a large spike in the sample size for a single month in terms of data quality and staffing.

A new sample designation code beginning with "B" was assigned to cases included in the general sample increase to distinguish SCHIP cases from regular CPS sample cases, which use an "A." The rotation group code was not reassigned in order to maintain the structure of the sample. In states targeted for the general sample increase, sample was added in all existing CPS sample areas.

The general sample increase differs from the other parts of the SCHIP expansion in two ways:

- It affects CPS every month of the year; and

- It is implemented differently for different states. More sample was added in same states, while other states received no general sample increase.

For the other two parts of the SCHIP expansion, each state's sample size was increased proportional to the amount of CPS sample already there.

Allocating the Sample to States (Round 1)

The estimates to be improved by the SCHIP sample expansion are 3-year averages of the percentage of low-income children without health insurance. Low-income children are defined, for SCHIP purposes, as those in households with incomes at or below 200 percent of the poverty level. To determine the states where added sample would be of the most benefit, the detectable difference of two consecutive 3-year averages was calculated (example: comparing the 1998-2000 3-year average to the 1999-2001 3-year average). The detectable difference is the smallest percentage difference between the two estimates that is statistically significant at the 90-percent confidence level.

To reassign only portions of retired CPS sample, the natural units to use were reduction groups. The assignment of reduction group codes is explained in Appendix C. Using reduction group codes, the sample was divided into eight approximately equal parts called expansion groups. This was done to simplify the iteration process performed to reassign sample. Each expansion group contains 12 to 13 reduction groups. In most states, only some of the eight expansion groups were reassigned sample designations. Reassigning all eight expansion groups is roughly equivalent to doubling the sample size, meaning that all expired sample from the old CPS sample designations and rotation groups is brought back to be interviewed again.

---

[1]"Retired sample" consists of households that finished their series of eight attempted CPS interviews.

The uninsured estimates used for allocating the sample can be found at
http://www.census.gov/hhes/hlthins/liuc98.html.

To find the current detectable difference for each state, a variation of a formula found in the Source and Accuracy statement (http://www.bls.census.gov/cps/ads/1998/ssrcacc.htm) for the March Supplement was used:

$$se_{x,p} = \sqrt{\frac{bf^2}{x} p(100 - p)} \qquad (1)$$

where: $se_{x,p}$ = standard error of the proportion
    x = the total number of children meeting the low-income criteria
    p = the percentage of low-income children without health insurance
    b = a general variance function (GVF) parameter
    f^2 = state factor to adjust for differences in the CPS sample design between states

The same proportion (p) was used for all of the states (the 3-year national average from 1996 to 1998, 24.53 percent); this is consistent with CPS procedures to not favor one state over another, based upon rates that may change. Because the estimates are 3-year averages, some manipulation was necessary to find the standard error of the difference from one year to the next. To do so, the following formula was used:

$$se(p_{234} - p_{123}) = \frac{1}{3} \sqrt{se(p_4)^2 + se(p_1)^2} \qquad (2)$$

where p_{123} is the average estimate for years 1, 2, and 3, and p_{123} is the average for years 2, 3, and 4. Since the single year standard error estimates were not available, the standard errors calculated using formula (1) were used for both standard errors beneath the radical. The detectable difference was calculated by finding the coefficient of variation (by dividing by 24.53 percent), then multiplying by 1.645 to create a 90-percent confidence interval. The resulting detectable difference is proportional to the standard errors calculated.

Of the total annual funds allocated for this project, $7 million was budgeted for the general sample increase. This translated into approximately 12,000 additional assigned housing units each month. The 12,000 additional housing units were allocated among states by applying the formula below over several iterations, with the goal of lowering each state's detectable difference on the health insurance coverage rate for poor children:

$$SS' = SS * \left(\frac{DD}{DD'}\right)^2 \qquad (3)$$

where: SS' = new state sample size
    SS = old state sample size
    DD = old detectable difference for the state

    DD' = new detectable difference for the state (The number being decreased with each iteration.)

An increase of 12,000 housing units was reached when the state detectable differences were lowered to a maximum of 10.01 percent. There were some states that did not achieve this threshold, even after the sample size was doubled. For example, Connecticut started with a detectable difference of 14.51 percent. When lowering the desired detectable difference by iteration, the necessary sample size there doubled when the detectable difference reached 10.25 percent. No more sample was added there because of the decision to no more than double the existing sample in any state. The iterative process continued to lower the detectable difference (and raise the sample size) for the remainder of the states, however.

Alabama started with a detectable difference of 10.81 percent. After adding one expansion group, the detectable difference was 10.19 percent. After adding two, the detectable difference was 9.67 percent. Since the limit was 10.01 percent, only two expansion groups were added for Alabama. Arizona started with a detectable difference of 8.34 percent. Since the limit was 10.01 percent, no sample was added there.

Round 2 of State Allocation

The proposed allocation of sample to the states was then sent to the Census Bureau's Field Division for approval. Based upon their comments, the proposed sample increases were scaled back in Maryland, Virginia, Delaware, and the District of Columbia. After setting limits on the amount of sample to be added to these areas, the iteration process was continued to get back to roughly 12,000 housing units added to sample. All states where four or more expansion groups had already been added were also removed from consideration. These states were prohibited from getting further sample increases because of the proportionally large increases they had already received. After these final iterations, an extra expansion group was assigned in Iowa, North Dakota, Nevada, Oklahoma, Oregon, Utah, West Virginia, and Wyoming. These were states where three or fewer expansion groups had been added in Round 1. Table J-1 shows the final allocation of the general sample increase to states.

Phase-in of the General Sample Increase

Successful implementation of the other two parts of the SCHIP expansion required that the general sample increase be fully implemented by November 2000. To provide enough time for listing and keying the new sample assignments, SCHIP interviews could not begin until September 2000 (for more details on these processes, see Chapter 4). It was advantageous to begin adding sample as soon as possible to make the additions as gradual as possible for the field staff.

Table J-1. **Size of SCHIP General Sample Increase and Effect on Detectable Difference by State**

State	Original detectable difference (percent)	After first iteration		After final iteration	
		Expansion groups added	Detectable difference (percent)	Expansion groups added	Detectable difference (percent)
Alabama	10.81	2	9.67	2	9.67
Alaska	11.54	3	9.84	3	9.84
Arizona	8.34	0	8.34	0	8.34
Arkansas	8.87	0	8.87	0	8.87
California	4.00	0	4.00	0	4.00
Colorado	12.68	5	9.95	5	9.95
Connecticut	14.51	8	10.26	8	10.26
Delaware	12.59	5	9.88	3	10.74
District of Columbia	11.31	3	9.65	2	10.12
Florida	5.93	0	5.93	0	5.93
Georgia	8.88	0	8.88	0	8.88
Hawaii	12.11	4	9.89	4	9.89
Idaho	9.30	0	9.30	0	9.30
Illinois	6.57	0	6.57	0	6.57
Indiana	12.76	5	10.01	5	10.01
Iowa	11.38	3	9.71	4	9.29
Kansas	11.75	4	9.60	4	9.60
Kentucky	10.79	2	9.65	2	9.65
Louisiana	9.25	0	9.25	0	9.25
Maine	13.64	7	9.96	7	9.96
Maryland	14.65	8	10.36	5	11.50
Massachusetts	9.40	0	9.40	0	9.40
Michigan	7.23	0	7.23	0	7.23
Minnesota	11.82	4	9.65	4	9.65
Mississippi	8.92	0	8.92	0	8.92
Missouri	11.84	4	9.67	4	9.67
Montana	9.23	0	9.23	0	9.23
Nebraska	11.31	3	9.64	3	9.64
Nevada	11.71	3	9.98	4	9.56
New Hampshire	14.75	8	10.43	8	10.43
New Jersey	8.53	0	8.53	0	8.53
New Mexico	7.96	0	7.96	0	7.96
New York	4.71	0	4.71	0	4.71
North Carolina	8.17	0	8.17	0	8.17
North Dakota	10.93	2	9.78	3	9.32
Ohio	7.01	0	7.01	0	7.01
Oklahoma	9.97	0	9.97	1	9.40
Oregon	11.35	3	9.68	4	9.27
Pennsylvania	6.93	0	6.93	0	6.93
Rhode Island	14.93	8	10.56	8	10.56
South Carolina	11.19	3	9.55	3	9.55
South Dakota	11.28	3	9.62	3	9.62
Tennessee	10.01	1	9.44	1	9.44
Texas	4.78	0	4.78	0	4.78
Utah	9.78	0	9.78	1	9.22
Vermont	13.40	7	9.79	7	9.79
Virginia	12.02	4	9.81	2	10.75
Washington	12.28	5	9.63	5	9.63
West Virginia	10.62	1	10.01	2	9.50
Wisconsin	12.59	5	9.88	5	9.88
Wyoming	10.44	1	9.84	2	9.33

Table J-2. **Phase-in Table (First 11 Months of Sample Expansion)**

New first month in sample	Old sample designation	New sample designation	Rotation group (both old and new)	Last month in sample previously	MIS upon return to sample
Sep. 2000	A63	B71	6	Aug. 1995	5
	A65	B73	2	Aug. 1996	1
Oct. 2000	A63	B71	7	Sep. 1995	5
	A65	*B73*	*1*	*July 1996*	*3*
	A65	B73	3	Sep. 1996	1
Nov. 2000	*A63*	*B71*	*5*	*July 1995*	*8*
	A63	B71	8	Oct. 1995	5
	A65	B73	4	Oct. 1996	1
Dec. 2000	A64	B72	1	Nov. 1995	5
	A65	B73	5	Nov. 1996	1
Jan. 2001	A64	B72	2	Dec. 1995	5
	A65	B73	6	Dec. 1996	1
Feb. 2001	A64	B72	3	Jan. 1996	5
	A65	B73	7	Jan. 1997	1
Mar. 2001	A64	B72	4	Feb. 1996	5
	A65	B73	8	Feb. 1997	1
Apr. 2001	A64	B72	5	Mar. 1996	5
	A66	B74	1	Mar. 1997	1
May 2001	A64	B72	6	Apr. 1996	5
	A66	B74	2	Apr. 1997	1
Jun. 2001	A64	B72	7	May 1996	5
	A66	B74	3	May 1997	1
Jul. 2001	A64	B72	8	June 1996	5
	A66	B74	4	June 1997	1
Aug. 2001	A66	B74	5	July 1997	1

Table J-2 shows the phase-in of the expanded sample. The sample designation/rotation groups in italics did not enter sample MIS 1 or 5.[2] B73(1) entered at MIS 3, and B71(5) entered at MIS 8. Once introduced, these cases followed the regular CPS rotation pattern. By November 2000, the phase-in of the general sample increase was completed. However, because of the 4-8-4 rotating design of the CPS sample, it was necessary to continue adding new SCHIP cases to MIS 1 and 5 assignments for the next 8 months. In other words, for the next 8 months, SCHIP sample was being interviewed for the first time in both MIS 1 and MIS 5, rather than just MIS 1. New SCHIP cases were added only to MIS 1 assignments after July 2001.

A rotation chart for SCHIP follows (Figure J-1), for those who are more familiar with this method of visualizing survey sample. An explanation of rotation charts is found in Chapter 3.

Use of General Sample Increase in Official Labor Force Estimates

Although the original purpose of the SCHIP general sample increase focused on improving the reliability of estimates of low-income uninsured children, the Bureau of Labor Statistics (BLS) saw the increase in the monthly CPS sample size as an opportunity to improve the reliability of labor force estimates, particularly at the state level. However, the added SCHIP sample initially was **not** used in the official CPS labor force estimates, for a couple of reasons:

- To give Field Division the time to train the new field representatives, and

- To analyze the labor force estimates both with and without the new sample, ensuring that no bias was introduced into the official time series.

In order to gauge the effect of the additional sample on estimates, CPS data were weighted both with and without the expanded sample beginning in November 2000. The analysis mentioned compared the estimates as calculated using both sets of weights. CPS weights for all SCHIP sample are 0, while combined CPS/SCHIP weights are the same for SCHIP sample and CPS sample cases within each Primary Sampling Unit (PSU).

Weighting the CPS sample with the SCHIP sample included was done using this formula:

$$BW_{CPS/SCHIP} = BW_{CPS} * \left(\frac{RG_{CPS}}{RG_{CPS} + RG_{SCHIP}} \right) \quad (4)$$

---

[2]To maintain the known properties of the sample rotation pattern, it is desirable for CPS sample to begin interviewing in MIS 1. If sample cannot be ready (for example, permit sample that does not finish processing) by MIS 1, then it is deferred to MIS 5, thus getting four interviews in consecutive months. However, in this case there was no choice but to add sample at other points in the CPS interview cycle, because of the short time frame available for phase-in.

Figure J-1. **SCHIP Rotation Chart: September 2000 to December 2002**

CPS-294B
(8-17-2000)

U.S. DEPARTMENT OF COMMERCE
Economics and Statistics Administration
U.S. CENSUS BUREAU

SCHIP ROTATION CHART
September 2000 - December 2002
Sample and Rotation

Year/Month		B71	B72	B73	B74	B75	B76
2000	SEP	. . . . . 6 . .	. . . .	. 2			
	OCT	. . . . . 6 7 .	. . . .	1 2 3			
	NOV	. . . . 5 6 7 8	. . . .	1 2 3 4			
	DEC	. . . . . 6 7 8	1	. 2 3 4 5			
2001	JAN	. . . . . . 7 8	1 2	. . 3 4 5 6			
	FEB	 8	1 2 3	. . . 4 5 6 7			
	MAR	. . . .	1 2 3 4	. . . . 5 6 7 8			
	APR	. . . .	. 2 3 4 5 . . .	. . . . . 6 7 8	1		
	MAY	. . . .	. . 3 4 5 6 . .	. . . . . . 7 8	1 2		
	JUN	. . . .	. . . 4 5 6 7 .	 8	1 2 3		
	JUL	. . . .	. . . 5 6 7 8	. . . .	1 2 3 4		
	AUG	. . . .	. . . . 6 7 8	1	. 2 3 4 5		
	SEP	. . . .	. . . . . . 7 8	1 2	. . 3 4 5 6		
	OCT	. . . .	 8	1 2 3	. . . 4 5 6 7		
	NOV	. . . .	. . . .	1 2 3 4	. . . . 5 6 7 8		
	DEC	. . . .	. . . .	. 2 3 4 5 . . .	. . . . . 6 7 8	1	
2002	JAN	. . . .	. . . .	. . 3 4 5 6 . .	. . . . . . 7 8	1 2	
	FEB	. . . .	. . . .	. . . 4 5 6 7 .	 8	1 2 3	
	MAR	. . . .	. . . .	. . . 5 6 7 8	. . . .	1 2 3 4	
	APR	. . . .	. . . .	. . . . 6 7 8	1	. 2 3 4 5	
	MAY	. . . .	. . . .	. . . . . . 7 8	1 2	. . 3 4 5 6	
	JUN	. . . .	. . . .	 8	1 2 3	. . . 4 5 6 7	
	JUL	. . . .	. . . .	. . . .	1 2 3 4	. . . . 5 6 7 8	
	AUG	. . . .	. . . .	. . . .	. 2 3 4 5 . . .	. . . . . 6 7 8	1
	SEP	. . . .	. . . .	. . . .	. . 3 4 5 6 . .	. . . . . . 7 8	1 2
	OCT	. . . .	. . . .	. . . .	. . . 4 5 6 7 .	 8	1 2 3
	NOV	. . . .	. . . .	. . . .	. . . . 5 6 7 8	. . . .	1 2 3 4
	DEC	. . . .	. . . .	. . . .	. . . . . 6 7 8	1	. 2 3 4 5

where:
$BW_{CPS/SCHIP}$ = baseweight of combined CPS and SCHIP sample

BW_{CPS} = baseweight of CPS sample alone

RG_{SCHIP} = reduction groups added in SCHIP expansion (as part of appropriate expansion groups)

RG_{CPS} = reduction groups currently in CPS sample (usually 101)

The updated baseweights (or sampling intervals) can be found in Table J-4, along with updated coefficients of variation. For more details on weighting, see Chapter 10.

After analyzing several months of data, with and without the added SCHIP sample, BLS decided to include the SCHIP data in the official monthly CPS estimates, beginning with the July 2001 data release. The national unemployment rates were virtually identical when estimated from the sample with and without the additional SCHIP cases. The SCHIP expansion did not result in systematically higher or lower labor force estimates across states, although there were a few statistically significant differences in some months in a few states. Results for a typical month are shown in Table J-3.

The additional sample will slightly improve the reliability of national labor force estimates, and will more substantially improve the state estimates for states in which SCHIP sample was added (see U.S. Department of Labor, Employment & Earnings, August 2001, for more details). Table J-4 shows the effect of the general sample increase on state sampling intervals and expected coefficients of variation (CVs) for the annual average estimated levels of unemployed assuming a 6 percent unemployment rate. The CVs will be improved (made smaller) by the additional sample in some states. The first row of the table shows the effect on the monthly national CV on unemployment level, which is expected to drop from about 1.9 percent to 1.8 percent.

For purposes of comparison, the state sampling intervals and CVs are included, as estimated following a sample maintenance reduction which began in December 1999. (Details on how maintenance reductions are accomplished are found in Appendix C.) All type A rates, design effects, estimated state populations, and civilian labor force levels used to estimate the new CVs are the same as those used for the maintenance reduction, so the differences in CVs shown are resulting only from the changes in the sampling intervals due to the SCHIP general sample increase.

Summary of CPS/SCHIP Sample Design and Reliability Requirements

The current sample design, introduced in July 2001, includes about 72,000 assigned housing units from 754 sample areas (see Table J-5). Sufficient sample is allocated to maintain, at most, a 1.9 percent CV on national monthly estimates of unemployment level, assuming a 6-percent unemployment rate. This translates into a change of 0.2 percentage point in the unemployment rate being significant at a 90-percent confidence level. For each of the 50 states and for the District of Columbia, the design maintains a CV of at most 8 percent on the annual average estimate of unemployment level, assuming a 6-percent unemployment rate. About 60,000 assigned housing units are required in order to meet the national and state reliability criteria. Due to the national reliability criterion, estimates for several large states are substantially more reliable than the state design criterion requires. Annual average unemployment estimates for California, Florida, New York, and Texas, for example, carry a CV of less than 4 percent.

Table J-3. **Effect of SCHIP on National Labor Force Estimates: June 2001**

Estimate	Population segment	CPS only	CPS/SCHIP combined	Difference
Labor force participation rate	Total, 16 years and over	67.4	67.4	0.0
	16 to 19 years	58.1	58.5	−0.4
	Men, 20 years and over	76.5	76.6	−0.1
	Women, 20 years and over	60.5	60.4	0.1
	White	67.6	67.7	−0.1
	Black	66.2	66.1	0.1
	Hispanic origin	67.9	67.8	0.1
Employment-population ratio	Total, 16 years and over	64.2	64.2	0.0
	16 to 19 years	48.5	48.6	−0.1
	Men, 20 years and over	73.6	73.7	−0.1
	Women, 20 years and over	58.0	58.0	0.0
	White	64.8	64.9	−0.1
	Black	60.4	60.5	−0.1
	Hispanic origin	63.4	63.4	0.0
Unemployment rate	Total, 16 years and over	4.7	4.7	0.0
	16 to 19 years	16.6	16.8	−0.2
	Men, 20 years and over	3.8	3.8	0
	Women, 20 years and over	4.0	4.0	0
	White	4.1	4.1	0
	Black	8.7	8.6	0.1
	Hispanic origin	6.6	6.6	0

Table J-4. **Effect of SCHIP on State Sampling Intervals and Coefficients of Variation**

State	Sampling intervals			Coefficients of variation (CVs) of annual average level of unemployed (percent)	
	Following 1996 reduction (Appendix H)	Following 1999 maintenance reduction	With SCHIP sample included	Following 1999 maintenance reduction	With SCHIP sample included
Total	2,255	2,366	2,128	1.9	1.8
Alabama	2,298	2,443	1,955	7.2	6.5
Alaska	336	336	244	7.4	6.5
Arizona	2,238	2,238	2,238	7.0	7.0
Arkansas	1,316	1,316	1,316	7.2	7.2
California	2,745	3,018	3,018	2.9	2.9
Los Angeles	1,779	1,919	1,919	4.3	4.3
Balance	3,132	3,438	3,438	3.7	3.7
Colorado	1,992	2,163	1,331	7.0	5.8
Connecticut	2,307	2,307	1,154	7.7	5.4
Delaware	505	505	367	7.4	6.3
District of Columbia	356	356	285	8.0	7.1
Florida	2,176	2,338	2,338	3.7	3.7
Georgia	3,077	3,451	3,451	6.3	6.3
Hawaii	769	769	513	7.8	6.4
Idaho	590	608	608	6.8	6.8
Illinois	2,227	2,227	2,227	4.2	4.2
Indiana	3,132	3,132	1,927	7.0	5.7
Iowa	1,582	1,582	1,055	7.2	6.2
Kansas	1,423	1,452	968	7.3	6.2
Kentucky	2,089	2,153	1,722	7.2	6.5
Louisiana	2,143	2,186	2,186	7.0	7.0
Maine	838	838	447	7.6	5.7
Maryland	3,061	3,061	1,884	7.1	5.7
Massachusetts	1,853	1,927	1,927	5.0	5.0
Michigan	2,098	2,181	2,181	4.5	4.5
Minnesota	2,437	2,437	1,625	7.3	6.3
Mississippi	1,433	1,508	1,508	7.3	7.3
Missouri	3,132	3,132	2,088	7.1	5.9
Montana	431	463	463	7.2	7.2
Nebraska	910	910	662	7.2	6.3
Nevada	961	1,067	711	7.0	5.8
New Hampshire	857	892	446	7.5	5.3
New Jersey	1,841	1,898	1,898	4.4	4.4
New Mexico	867	922	922	7.2	7.2
New York	1,970	2,034	2,034	3.2	3.2
New York City	1,774	1,951	1,951	5.2	5.2
Balance	2,093	2,093	2,093	4.3	4.3
North Carolina	1,986	2,145	2,145	5.6	5.6
North Dakota	363	363	264	7.4	6.5
Ohio	2,343	2,376	2,376	4.4	4.4
Oklahoma	1,548	1,629	1,448	7.4	7.1
Oregon	1,904	2,046	1,364	7.3	6.1
Pennsylvania	2,149	2,149	2,149	4.1	4.1
Rhode Island	687	687	344	7.8	5.5
South Carolina	2,291	2,385	1,735	7.4	6.4
South Dakota	376	376	273	7.1	6.3
Tennessee	3,016	3,140	2,791	7.0	6.6
Texas	2,658	2,796	2,796	3.6	3.6
Utah	958	1,063	945	7.1	6.8
Vermont	410	436	232	7.3	5.4
Virginia	3,084	3,268	2,619	6.8	6.2
Washington	2,999	3,329	2,048	7.6	6.3
West Virginia	896	896	717	7.1	6.4
Wisconsin	2,638	2,638	1,623	7.4	6.3
Wyoming	253	253	202	7.3	6.7

Table J-5. **CPS Sample Housing Unit Counts With and Without SCHIP**

Housing unit categories	CPS only	CPS and SCHIP
Assigned housing units	60,000	72,000
Eligible (occupied)	50,000	60,000
Interviewed.	46,800	55,500
Type A noninterviews	3,200	4,500

In support of the SCHIP, about 12,000 additional housing units are allocated to the District of Columbia and 31 states. These are generally the states with the smallest samples after the 60,000 housing units are allocated to satisfy the national and state reliability criteria.

In the first stage of sampling, the 754 sample areas are chosen. In the second stage, ultimate sampling unit clusters composed of about four housing units each are selected.[3] Each month, about 72,000 housing units are assigned for data collection, of which about 60,000 are occupied and thus eligible for interview. The remainder are units found to be destroyed, vacant, converted to nonresidential use, to contain persons whose usual place of residence is elsewhere, or to be ineligible for other reasons. Of the 60,000 eligible housing units, about 7.5 percent are not interviewed in a given month due to temporary absence (such as a vacation), other failures to make contact after repeated attempts, inability of persons contacted to respond, unavailability for other reasons, and refusals to cooperate (about half of the noninterviews). Information is obtained each month for about 112,000 persons 16 years of age or older.

OTHER ASPECTS OF THE SCHIP EXPANSION

Although the general sample increase was the most difficult to implement of the three parts of the SCHIP expansion, the other two parts actually provide more sample for the March Supplement. The following paragraphs provide more details on these.

Split-Path Assignments

As part of the SCHIP expansion, some CPS sample from February and April will receive the March CPS Supplement. This will take the place of the regularly scheduled

---

[3]In states with the SCHIP general sample increase, the size of these ultimate clusters can be between four and eight housing units.

supplement for that month. Only households that would ordinarily not receive the March Supplement will receive it. This means that February households in MIS 4 and 8 will receive the March Supplement in February, and April MIS 1 and 5 households will receive it in April.

Households from MIS 1 and 5 in the previous November will be screened at that time based upon whether they have children (18 or younger) or non-White members. Those that do meet these criteria will receive the March CPS basic instrument and supplement in February, when they are MIS 4 and 8, respectively. Households with Hispanic residents will be excluded from this screening, since they will already be interviewed as part of the November Hispanic oversampling. The March supplement will take the place of the February supplement for these households, effectively lowering the February supplement's sample size.

In April, sample cases from MIS 1 and 5 will be screened in the regular CPS interview for presence of children or Hispanic or non-White inhabitants. Those who meet any of these criteria will be given the March supplement instead of the April supplement. See Chapter 11 for more details on CPS supplements.

Month-in-Sample 9 Assignments

CPS sample from MIS 6, 7, and 8 in November will be assigned for interview in February or April using the March CPS basic instrument and income supplement. Before interviewing, these households will be screened, using CPS data, so that each household with at least one child age 18 or younger or a non-White member will be in the sample. (Hispanics will already be interviewed as a result of the Hispanic oversampling of the March supplement; see Chapter 11.) To the extent possible, these cases will be assigned for interview through one of the telephone centers, but recycles will be interviewed in the field during the week of the appropriate month in which CPS interviews are conducted.

EFFECT OF SCHIP EXPANSION ON STANDARD ERRORS OF ESTIMATES OF UNINSURED LOW-INCOME CHILDREN

Table J-6 shows the effect of the SCHIP sample expansion on the standard errors of the state-level estimates of

Table J-6. **Standard Errors on Percentage of Uninsured, Low-Income Children Compared at Stages of the SCHIP Expansion[1]**

[In percent]

State	Before SCHIP expansion (percent)	After general sample increase (percent)	After SCHIP expansion (percent)[2]
Alabama	1.61	1.44	1.22
Alaska	1.72	1.47	1.30
Arizona	1.24	1.24	1.14
Arkansas	1.32	1.32	1.17
California	0.60	0.60	0.53
Colorado	1.89	1.48	1.32
Connecticut	2.16	1.53	1.34
Delaware	1.88	1.60	1.41
District of Columbia	1.69	1.51	1.29
Florida	0.88	0.88	0.80
Georgia	1.32	1.32	1.12
Hawaii	1.81	1.47	1.18
Idaho	1.39	1.39	1.24
Illinois	0.98	0.98	0.86
Indiana	1.90	1.49	1.33
Iowa	1.70	1.39	1.24
Kansas	1.75	1.43	1.26
Kentucky	1.61	1.44	1.28
Louisiana	1.38	1.38	1.17
Maine	2.03	1.49	1.36
Maryland	2.19	1.71	1.48
Massachusetts	1.40	1.40	1.26
Michigan	1.08	1.08	0.94
Minnesota	1.76	1.44	1.25
Mississippi	1.33	1.33	1.15
Missouri	1.77	1.44	1.29
Montana	1.38	1.38	1.23
Nebraska	1.69	1.44	1.24
Nevada	1.75	1.43	1.30
New Hampshire	2.20	1.56	1.37
New Jersey	1.27	1.27	1.12
New Mexico	1.19	1.19	1.09
New York	0.70	0.70	0.62
North Carolina	1.22	1.22	1.07
North Dakota	1.63	1.39	1.22
Ohio	1.05	1.05	0.91
Oklahoma	1.49	1.40	1.23
Oregon	1.69	1.38	1.24
Pennsylvania	1.03	1.03	0.92
Rhode Island	2.23	1.57	1.40
South Carolina	1.67	1.42	1.23
South Dakota	1.68	1.43	1.23
Tennessee	1.49	1.41	1.25
Texas	0.71	0.71	0.64
Utah	1.46	1.38	1.20
Vermont	2.00	1.46	1.34
Virginia	1.79	1.60	1.39
Washington	1.83	1.44	1.24
West Virginia	1.58	1.42	1.28
Wisconsin	1.88	1.47	1.32
Wyoming	1.56	1.39	1.24

[1]Standard error on difference of two consecutive 3-year averages.
[2]Includes MIS 9 and split-path cases.

the proportions of low-income children without health insurance. These standard errors:

- Refer to the difference between two consecutive 3-year averages,

- Are calculated after 4 years of the SCHIP expansion; i.e. all 3 years of both estimates are expanded sample,

- Use the same rate of low-income children without health insurance for all states.

- Are based on expected, not actual, sample sizes.

The values in the middle column (After General Sample Increase) are proportional to the final detectable differences found in Table J-1. Included in the last column in the table is an estimated standard error for each state after the Split-path and MIS 9 cases are taken into account.

REFERENCES

Cahoon, L. (2000), "CPS 1990 Design Reassignment #1 - September 2000 (SCHIP90-2, SAMADJ90-20)," Memorandum for J. Mackey Lewis, April 11th, Demographic Statistical Methods Division, U.S. Census Bureau.

Rigby, B. (2000), "SCHIP CPS Sample Expansion used to Quantify Sample Expansion Among the States (SCHIP-1)," Memorandum for Documentation, March 29th, Demographic Statistical Methods Division, U.S. Census Bureau.

Shoemaker, H. (2000), "Effect of SCHIP/CPS General Sample Expansion on State Sampling Intervals and Coefficients of Variation (SAMADJ90-22, SCHIP90-5)," Memorandum for Distribution List, Demographic Statistical Methods Division, U.S. Census Bureau.

U.S. Department of Labor, Bureau of Labor Statistics, **Employment and Earnings,** Vol.48, No.8, Washington, DC: Government Printing Office.

U.S. Public Law 106-113.

Acronyms

ADS	Annual Demographic Supplement		MLR	Major Labor Force Recode
API	Asian and Pacific Islander		MOS	Measure of Size
ARMA	Autoregressing Moving Average		MSA	Metropolitan Statistical Area
BLS	Bureau of Labor Statistics		NCEUS	National Commission on Employment and Unemployment Statistics
BPC	Basic PSU Components			
CAPE	Committee on Adjustment of Postcensal Estimates		NCHS	National Center for Health Statistics
			NHIS	National Health Interview Survey
CAPI	Computer Assisted Personal Interviewing		NPC	National Processing Center
CATI	Computer Assisted Telephone Interviewing		NRAF	Net Recruits to the Armed Forces
CCM	Civilian U.S. Citizens Migration		NRO	Net Recruits to the Armed Forces from Overseas
CCO	CATI and CAPI overlap			
CES	Current Employment Statistics		NSR	Nonself-Representing
CESEM	Current Employment Statistics Survey Employment		OB	Number of Births
			OD	Number of Deaths
CNP	Civilian Noninstitutional Population		OMB	Office of Management and Budget
CPS	Current Population Survey		ORR	Office of Refugee Resettlement
CPSEP	CPS Employment-to-Population Ratio		PAL	Permit Address List
CPSRT	CPS Unemployment Rate		PES	Post-Enumeration Survey
CV	Coefficient of Variation		PIP	Performance Improvement Period
DA	Demographic Analysis		POP	Performance Opportunity Period
DAFO	Deaths of the Armed Forces Overseas		PSU	Primary Sampling Unit
DEFF	Design Effect		QC	Quality Control
DOD	Department of Defense		RDD	Random Digit Dialing
DRAF	Deaths of the Resident Armed Forces		RE	Response Error
FOSDIC	Film Optical Sensing Device for Input to the Computer		RO	Regional Office
			SCHIP	State Children's Health Insurance Program
FR	Field Representative		SECU	Standard Error Computation Unit
FSC	First- and Second-Stage Combined		SFR	Senior Field Representative
FSCPE	Federal State Cooperative Program for Population Estimates		SMSA	Standard Metropolitan Statistical Area
			SOC	Survey of Construction
GQ	Group Quarters		SR	Self-Representing
GVF	Generalized Variance Function		STARS	State Time Series Analysis and Review System
HVS	Housing Vacancy Survey			
I&O	Industry and Occupation		SW	Start-With
IM	International Migration		TE	Take-Every
INS	Immigration and Naturalization Service		UE	Number of People Unemployed
IRCA	Immigration Reform and Control Act		UI	Unemployment Insurance
LAUS	Local Area Unemployment Statistics		URE	Usual Residence Elsewhere
MARS	Modified Age, Race, and Sex		USU	Ultimate Sampling Unit
MIS	Month-in-Sample		WPA	Work Projects Administration

Index